Submitted to Statistical Science

The Matérn Model: A Journey through Statistics, Numerical Analysis and Machine Learning

Emilio Porcu[1], Moreno Bevilacqua, Robert Schaback and Chris J. Oates

Abstract. The Matérn model has been a cornerstone of spatial statistics for more than half a century. More recently, the Matérn model has been exploited in disciplines as diverse as numerical analysis, approximation theory, computational statistics, machine learning, and probability theory. In this article we take a Matérn-based journey across these disciplines. First, we reflect on the importance of the Matérn model for estimation and prediction in spatial statistics, establishing also connections to other disciplines in which the Matérn model has been influential. Then, we position the Matérn model within the literature on big data and scalable computation: the SPDE approach, the Vecchia likelihood approximation, and recent applications in Bayesian computation are all discussed. Finally, we review recent devlopments, including flexible alternatives to the Matérn model, whose performance we compare in terms of estimation, prediction, screening effect, computation, and Sobolev regularity properties.

Keywords: Approximation Theory, Compact Support, Covariance, Kernel, Kriging, Machine Learning, Maximum Likelihood, Reproducing Kernel Hilbert Spaces, Spatial Statistics, Sobolev Spaces.

1. INTRODUCTION

This paper serves two purposes: On the one hand, we provide a panoramic view, across several disciplines, of the Matérn model. On the other hand, the paper illustrates the role of the Matérn model in several disciplines, while discussing alternative or more general models and their relevance to many aspects of statistical modeling, estimation, prediction, computational statistics, numerical analysis, and machine learning.

A historical account of the Matérn model is provided by Guttorp and Gneiting [69]. The Matérn model – also called the Matérn *covariance function*, or the Matérn *kernel*, depending on context – is commonly attributed to Matérn [109], but can be found under alternative names in different branches of the scientific literature. The use of the Matérn model is widespread, and it is impossible to cover all its diverse applications here; our review focuses on a selection of applications that are of especial interest and significance. Specifically, we aim to cover

1. estimation and prediction using the Matérn model in statistics, with emphasis on maximum likelihood estimation, Kriging prediction, and the associated screening effect;
2. applications of the Matérn model in
   a) computational statistics, including the stochastic differential equation (SDE) and stochastic partial differential equation (SPDE) approaches, likelihood approximation, inference of partial differential equations (PDEs) and Charles Stein's method;
   b) statistical modeling, including non-standard scenarios, for instance when isotropy and sta-

Emilio Porcu is Professor, Department of Mathematics, Khalifa University, and Research Fellow at ADIA Lab, both institutions located in Abu Dhabi, The United Arab Emirates (e-mail: emilio.porcu@bku.ac.ae). Moreno Bevilacqua is Professor, Department of Statistics, Universidad Adolfo Ibanez (e-mail: moreno.bevilacqua@uai.cl). Robert Schaback is Professor, Department of Mathematics, University of Göttingen, Germany (e-mail: schaback@math.uni-goettingen.de). Chris J. Oates is Professor, School of Mathematics, Statistics & Physics, Newcastle University, UK (e-mail: chris.oates@ncl.ac.uk).

[1]Corresponding Author.

tionarity cannot be assumed, or to model directions and curves;
  c) approximation theory and numerical analysis, where the Matérn model is used to construct kernel-based interpolants;
  d) machine learning, where the Matérn model is central to the literature on Gaussian processes modelling; and
  e) probability theory, where the Matérn model has inspired several contributions based on properties of the sample paths of associated stochastic processes, in concert with the solution of certain classes of stochastic differential equations;
3. comparison with recent flexible alternatives to the Matérn model, with a focus on
  a) enhanced models with interesting features, such as compact support or polynomial decay;
  b) asymptotic estimation accuracy, misspecified prediction, and screening effects;
  c) the implications of using certain classes of compactly supported kernels within approximation theory, computational statistics, and machine learning.

This article is novel, in being the first to take a broad view of the scientific literature through the lens of the Matérn model. In particular, we do *not* attempt a review of covariance functions in general. Recent reviews provide a quite exhaustive panorama of covariance models, from space to space-time [128], to multivariate covariance functions [58], and covariance-based modeling on spheres and manifolds [123]. In addition, while there are many fascinating applications of the Matérn model across the scientific landscape, we cannot hope to do justice to them all. Our emphasis is therefore limited to methodological and theoretical issues which we hope are of relevance across a wide range of disciplines in which the Matérn model is used.

1.1 Setting and Notation

Throughout, bold letters refer to vectors and matrices, and the transpose operator is denoted $\top$. Let $d \in \mathbb{N}$ and let $Z = \{Z(\boldsymbol{x}),\ \boldsymbol{x} \in \mathbb{R}^d\}$ be a real-valued Gaussian random field, having zero mean and and *covariance function* $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined via $K(\boldsymbol{x}, \boldsymbol{y}) := \mathrm{Cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y}))$. Covariance functions are symmetric and positive definite, where in this paper the term *positive definite* is understood as

$$(1) \qquad \sum_{i=1}^{n}\sum_{j=1}^{n} c_i K(\boldsymbol{x}_i, \boldsymbol{x}_j) c_j \geq 0$$

for all $c_i \in \mathbb{R}$, all $n \in \mathbb{N}$ and all $\boldsymbol{x}_i \in \mathbb{R}^d$. If the inequality above is strict, then K will be called strictly positive definite.

Each symmetric positive definite function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defines *translate* functions $K(\boldsymbol{x}, \cdot)$ on $\mathbb{R}^d$, for all $\boldsymbol{x} \in \mathbb{R}^d$. In addition, one can define an inner product on two translates by

$$(2) \qquad \langle K(\boldsymbol{x}, \cdot), K(\boldsymbol{y}, \cdot)\rangle_{\mathcal{H}(\mathcal{K})} := K(\boldsymbol{x}, \boldsymbol{y}),\ \ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d,$$

in terms of K itself. This extends to all linear combinations of translates and *generates*, by completion, a Hilbert space $\mathcal{H}(\mathcal{K})$ of functions on $\mathbb{R}^d$. This space is called the *native* space for K. Notice that the Hilbert space allows for continuous point evaluations $\delta_{\boldsymbol{x}} : f \mapsto f(\boldsymbol{x})$ via a *reproduction formula*

$$(3) \qquad f(\boldsymbol{x}) = \langle f, K(\boldsymbol{x}, \cdot)\rangle_{\mathcal{H}(K)},\ \boldsymbol{x} \in \mathbb{R}^d,\ \ f \in \mathcal{H}(K)$$

that follows from (2). Then $\mathcal{H}(K)$ is called a *reproducing kernel Hilbert space (RKHS)* with *kernel* K. In particular, the translates $K(\boldsymbol{x}, \cdot)$ lie in $\mathcal{H}(K)$, forming its completion and being the Riesz representers of delta functionals $\delta_{\boldsymbol{x}}$. They are central to machine learning, numerical analysis and approximation theory, since (2) allows inner products in the abstract space $\mathcal{H}(K)$ to be explicitly computable using the kernel - the so-called *kernel trick*. See Section 6.1 and [167] for more detail. For a positive definite and stationary kernel K, its Fourier transform $\hat{K}$ can be used to recast the inner product (2) on the Hilbert space $\mathcal{H}(K)$ by

$$(4) \qquad \langle f, g\rangle_{\mathcal{H}(K)} = \int_{\mathbb{R}^d} \frac{\hat{f}(\boldsymbol{\omega})\overline{\hat{g}(\boldsymbol{\omega})}}{\hat{K}(\boldsymbol{\omega})} \mathrm{d}\boldsymbol{\omega},\ \ f, g \in \mathcal{H}(K),$$

up to a constant factor. Here, $\overline{g}$ denotes the complex conjugate of a function g, and $\hat{g}$ its Fourier transform. Note how the spectrum of K penalizes the spectrum of the functions in $\mathcal{H}(K)$. Roughly, the Hilbert space $\mathcal{H}(K)$ consists of functions f for which $\hat{f}/\sqrt{\hat{K}}$ is square integrable over $\mathbb{R}^d$. The subtle connections of the Hilbert space $\mathcal{H}(K)$ to sample paths of Gaussian processes with covariance function K will come up at many places in this paper, e.g. in Sections 2, 4.4, 6.3, and 7.1. In this sense, kernels are important links between deterministic and probabilistic models.

A strictly positive definite kernel K is called *stationary* if $K(\boldsymbol{x}, \boldsymbol{y}) \equiv K(\boldsymbol{x} - \boldsymbol{y})$. According to Bochner's theorem [27], K is the Fourier transform of a positive and bounded measure F, that is

$$K(\boldsymbol{x} - \boldsymbol{y}) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}(\boldsymbol{x}-\boldsymbol{y}, \boldsymbol{\omega})} F(\mathrm{d}\boldsymbol{\omega}), \qquad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d.$$

Here, $(\cdot, \cdot)$ is the inner product in $\mathbb{R}^d$ and i is the unit complex number. Fourier inversion is possible when K is absolutely integrable, in which case we call the Fourier transform $\widehat{K}$ its *spectral density*. We note that $\widehat{K}$ is nonnegative and integrable. Furthermore, most of the paper assumes stationarity and isotropy for the covariance function, K, so that

$$\text{(5)}\quad \operatorname{Cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y})) = K(\boldsymbol{x}-\boldsymbol{y}) = \sigma^2\varphi(\|\boldsymbol{x}-\boldsymbol{y}\|),$$

for $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d$ and $\|\cdot\|$ denoting the Euclidean distance. Here, we assume φ to be continuous with $\varphi(0) = 1$. Throughout, we shall equivalently call φ a *function* or a *correlation function*, the last as a shortcut to $\varphi(\|\cdot\|)$. Hence, the parameter $\sigma^2 > 0$ is the variance of $Z(\boldsymbol{x})$, for all $\boldsymbol{x} \in \mathbb{R}^d$. Let Φ_d denote the class of such functions φ inducing a covariance function K through the identity (5) *i.e.* Φ_d is the class of continuous isotropic correlation functions defined on $\mathbb{R}^d$. Such functions have a precise integral representation according to Schoenberg [143], given by

$$\text{(6)}\qquad \varphi(x) = \int_0^\infty \Omega_d(rx) F_d(\mathrm{d}r), \qquad x \geq 0,$$

with F_d being a probability measure and

$$\text{(7)}\quad \Omega_d(x) = \Gamma(d/2)\left(\frac{2}{x}\right)^{d/2-1} J_{d/2-1}(x), \qquad x \geq 0,$$

with $\Gamma(\cdot)$ the gamma function and J_ν the Bessel function of the first kind of order $\nu > 0$ [119, formula 10.2.2]. For a member φ of the class Φ_d, we can use that its d-variate Fourier transform of $\varphi(\|\boldsymbol{x}-\boldsymbol{y}\|)$ is isotropic again, and therefore reducible to a scalar integral formula

(8)
$$\widehat{\varphi}(z) = \frac{z^{1-d/2}}{(2\pi)^{d/2}} \int_0^\infty u^{d/2} J_{d/2-1}(uz)\varphi(u)\mathrm{d}u,\ z \geq 0,$$

defining its d-variate *isotropic spectral density*, and we assume this integral to exist. If the denominator $(2\pi)^{d/2}$ is omitted, the same formula holds for the inverse radial Fourier transform. Throughout, we write Φ_∞ for $\bigcap_{d\geq 1}\Phi_d$, the class of functions φ inducing positive definite radial functions on every d-dimensional Euclidean space. Hence, $\varphi \in \Phi_d$ if and only if $\varphi(\|\cdot\|)$ is a correlation function in $\mathbb{R}^d$.

2. THE MATÉRN MODEL

The *Matérn model*, $\mathcal{M}_{\nu,\alpha}$, is defined as [149]

$$\text{(9)}\qquad \mathcal{M}_{\nu,\alpha}(x) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{x}{\alpha}\right)^\nu \mathcal{K}_\nu\left(\frac{x}{\alpha}\right), \qquad x \geq 0,$$

with $\alpha > 0$ the *scale* parameter, $\nu > 0$ the *smoothness* parameters, and $\mathcal{K}_\nu$ a modified Bessel function of the second kind of order ν [2, 9.6.22]. It can be verified that $\mathcal{M}_{\nu,\alpha}(0) = 1$, so that (9) is a correlation function. Arguments in Stein [149, p48] show that $\mathcal{M}_{\nu,\alpha}$ belongs to the class Φ_∞. The function $\sigma^2\mathcal{M}_{\nu,\alpha}$ will be termed *Matérn covariance function*, and $\sigma^2 > 0$ will denote the variance of the associated Gaussian random field.

The importance of the Matérn class stems from the parameter ν that controls the differentiability of the sample paths of the associated Gaussian field. Specifically, for any positive integer k, the sample paths of a Gaussian field Z on $\mathbb{R}^d$ with Matérn correlation function are k-times mean square differentiable (in any direction) if and only if $\nu > k$. Also, a rescaled version of the Matérn correlation function converges to the Gaussian or squared exponential kernel as $\nu \to \infty$, that is

$$\text{(10)}\quad \mathcal{M}_{\nu,\alpha/(2\sqrt{\nu})}(x) \xrightarrow[\nu\to\infty]{} \exp(-x^2/\alpha^2), \qquad x \geq 0,$$

with convergence being uniform on any compact set of $\mathbb{R}^d$. For this reason, the parametrisation $\mathcal{M}_{\nu,\alpha/(2\sqrt{\nu})}$ is sometimes also adopted [170].

When $\nu = k + 1/2$, for k a nonnegative integer, the Matérn correlation function simplifies into the product of a negative exponential correlation function with a polynomial of order k. For instance, $\mathcal{M}_{1/2,1}(x) = \exp(-x)$ and $\mathcal{M}_{3/2,1}(x) = \exp(-x)(1+x)$. In general,

(11)
$$\mathcal{M}_{k+1/2,1}(x) = \exp(-x)\sum_{i=0}^{k} \frac{(k+i)!}{2k!}\binom{k}{i}(2x)^{k-i}$$

for $k \in \mathbb{N}_0$. This simple algebraic form for the Matérn correlation functions has undoubtedly contributed to the widespread popularity of the Matérn model.

Now we are in a position to explore in detail the many faces of the Matérn model. Section 3 discusses maximum likelihood estimation, Kriging prediction, and the screening effect, while Section 4 explores an SPDE characterisation of the Matérn model. Section 5 discusses the Matérn model as a building block to more sophisticated models, while Section 6 views the scientific landscape through the lens of the Matérn model, with special emphasis on numerical analysis, probability theory and machine learning. Section 7 introduces some recently developed alternatives and generalisations of the Matérn model, while Section 8 compares these alternative models in terms of estimation, prediction, and the screening effect.

3. ESTIMATION AND PREDICTION WITH THE MATÉRN MODEL

Let $D \subset \mathbb{R}^d$ be a subset of $\mathbb{R}^d$. Consider a set $X_n = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\}$ of (distinct) locations in D, at which values $\boldsymbol{Z}_n = (Z(\boldsymbol{x}_1),\ldots,Z(\boldsymbol{x}_n))^\top$ of the Gaussian random field Z, defined in Section 1.1, are observed. An important problem concerns the *prediction* of values $Z(\boldsymbol{x}_0)$ at an unobserved location $\boldsymbol{x}_0 \in D \setminus X_n$. Then an especially natural predictor for $Z(\boldsymbol{x}_0)$ is

$$\text{(12)}\qquad \widehat{Z}_n = \boldsymbol{c}_n^\top \boldsymbol{R}_n^{-1}\boldsymbol{Z}_n$$

with the vector $[\boldsymbol{c}_n]_i = K(\boldsymbol{x}_0,\boldsymbol{x}_i)$ and the *kernel matrix* $[\boldsymbol{R}_n]_{i,j} = K(\boldsymbol{x}_i,\boldsymbol{x}_j)$. The predictor (12) can be motivated from multiple directions. Classically, (12) is motivated as the best linear unbiased predictor (BLUP) for $Z(\boldsymbol{x}_0)$, and is often referred to as the *simple Kriging* predictor of $Z(\boldsymbol{x}_0)$ [42]. From a modern perspective, where the role of unbiased estimation is increasingly questioned, we can motivate this choice using alternative optimality properties, including:

1. it is the expectation of $Z(\boldsymbol{x}_0)$ conditionally on the realisation $\boldsymbol{Z}_n$;
2. it is the optimal estimate (i.e. the Bayes act) for $Z(\boldsymbol{x}_0)$ based on the data-set $\boldsymbol{Z}_n$, under squared error loss [117, Section 13.3];
3. it yields the minimal RKHS norm interpolant of the data evaluated at $\boldsymbol{x}_0$, by Section 6.1;
4. it is the algorithm for approximating $Z(\boldsymbol{x}_0)$ from $\boldsymbol{Z}_n$ that minimises the worst case error in the sense of information-based complexity [117, Section 10.2] and approximation theory (see Section 6.1),

to name but a few. The Matérn model provides a natural setting to study the performance of (12) if we suppose Z to have a stationary isotropic covariance function $\sigma^2\mathcal{M}_{\nu,\alpha}$. The crucial question of how to select suitable values for the parameters σ, α, ν will be considered first, in Section 3.1, and then the performance of (12) will be studied in Section 3.2. The possibility of a direct extension of the Matérn model to more general domains, such as manifolds and graphs, is discussed in Section 3.3.

3.1 Estimation Using Maximum Likelihood

Maximum likelihood (ML) and similar estimation methods are popular in this setting due to the availability of practical (inc. gradient-based) numerical methods for computation and the classical theory that underpins ML. On the other hand, implicit in the use of ML is that the statistical model is well-specified, and this judgement must be made on a case-by-case basis. To limit scope, we focus on ML estimation in the sequel. Our aim is to understand when the parameters of the Matérn model can be consistently estimated from data, and to understand the asymptotic distribution of the ML estimator. To this end, recall that the Gaussian log-likelihood function is

$$\mathcal{L}_n(\boldsymbol{\theta}) = -\frac{1}{2}\left(\log(|\sigma^2\boldsymbol{R}_n)|) + \frac{1}{\sigma^2}\boldsymbol{Z}_n^\top\boldsymbol{R}_n^{-1}\boldsymbol{Z}_n\right), \tag{13}$$

up to an additive constant, with $\boldsymbol{\theta} = (\nu, \alpha, \sigma^2)$. The ML estimator is defined as

$$\widehat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\mathbb{R}_+^3}{\operatorname{argmax}}\, \mathcal{L}_n(\boldsymbol{\theta}). \tag{14}$$

The ML estimate for the variance parameter can be computed in closed-form as $\widehat{\sigma}_n^2 = \boldsymbol{Z}_n^\top\boldsymbol{R}_n^{-1}\boldsymbol{Z}_n/n$; plugging this expression into (13) reduces the numerical problem to optimisation of a so-called *concentrated likelihood* over $\mathbb{R}_+^2$. However, maximizing the log-(concentrated) likelihood requires a nonlinear optimisation problem to be solved, for which numerical methods must be used; see Section 4.3.

The performance of ML estimation has been studied principally in two different asymptotic limits. Under *fixed domain asymptotics*, the sampling domain D is bounded and the set of sampled locations X_n becomes increasingly dense in D. Under *increasing domain asymptotics*, the domain D grows with the number n of observed data, and the distance between any two sampled locations is bounded away from zero. Zhang and Zimmerman [181] note that the peformance of the ML estimator can be quite different under these two frameworks, as will now be discussed.

3.1.1 Increasing Domain Asymptotics. Mardia and Marshall [108] make use of increasing domain asymptotics to establish, under mild regularity conditions, that the ML estimator is *strongly consistent*, meaning that $\widehat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$ for the *true* parameter $\boldsymbol{\psi}_0$. Furthermore, they establish that the ML estimator is *asymptotically normal*, meaning that

$$\boldsymbol{F}^{1/2}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \tag{15}$$

where $\boldsymbol{F}(\boldsymbol{\theta}) = -E[\mathcal{L}_n''(\boldsymbol{\theta})]$ is the Fisher information matrix, whose entries are

$$F(\boldsymbol{\theta})_{i,j} = \frac{1}{2}\mathrm{tr}\left(\frac{\mathrm{d}\boldsymbol{\Sigma}_n}{\mathrm{d}\boldsymbol{\theta}_i}\boldsymbol{\Sigma}_n^{-1}\frac{\mathrm{d}\boldsymbol{\Sigma}_n}{\mathrm{d}\boldsymbol{\theta}_j}\boldsymbol{\Sigma}_n^{-1}\right),$$

and $\boldsymbol{\Sigma}_n = \sigma^2\boldsymbol{R}_n$. Although our focus is on the Matérn model, we note that these kind of asymptotic results hold for any parametric correlation function obeying particular regularity conditions that are stated in terms of eigenvalue conditions on the correlation matrix and its derivatives [108], thought these may not be easy to verify in general (see for instance Shaby and Ruppert [145], for the exponential case). Generally speaking, as long as the spatial extent of the sampling region is large compared with the range of dependence of the random field, increasing-domain asymptotics provide a very accurate description of the behavior of the ML estimate [181, 145, 83].

3.1.2 Fixed Domain Asymptotics. Zhang [180] considered ML estimation for the Matérn model under fixed domain asymptotics, proving that when the smoothness parameter ν is known and fixed, none of the parameters σ^2 and α can be estimated consistently when $d = 1, 2, 3$. Instead, only the parameter

$$\mathrm{micro}_{\mathcal{M}} = \sigma^2/\alpha^{2\nu}, \tag{16}$$

sometimes called *microergodic* parameter [181, 149], can be consistently estimated. This is a consequence of the equivalence of the two corresponding Gaussian measures, that we denote with $P(\sigma_i^2\mathcal{M}_{\nu,\alpha_i})$, with $i=0,1$. In particular, for any bounded infinite set $D\subset\mathbb{R}^d$, $d=1,2,3$, $P(\sigma_0^2\mathcal{M}_{\nu,\alpha_0})$ is equivalent to $P(\sigma_1^2\mathcal{M}_{\nu,\alpha_1})$ on the paths of $Z(\boldsymbol{x}), \boldsymbol{x}\in D$, if and only if

$$\sigma_0^2/\alpha_0^{2\nu}=\sigma_1^2/\alpha_1^{2\nu}. \tag{17}$$

In contrast, for $d\geq 5$, Anderes [7] proved the orthogonality of two Gaussian measures with different Matérn covariance functions and hence, in this case, all the parameters can be consistently estimated under fixed-domain asymptotics. The case $d=4$ has been recently studied in Bolin and Kirchner [30].

Asymptotic results associated with ML estimation of the microergodic parameter, again for a fixed known smoothness parameter ν, can be found in Zhang [180], and later on in Kaufman and Shaby [83]. In particular, for a zero mean Gaussian field defined on a bounded infinite set $D\subset\mathbb{R}^d$, $d=1,2,3$, with a Matérn covariance function $\sigma_0^2\mathcal{M}_{\nu,\alpha_0}$ the ML estimator $\hat\sigma_n^2/\hat\alpha_n^{2\nu}$ of the microergodic parameter is strongly consistent, $i.e.$,

$$\hat\sigma_n^2/\hat\alpha_n^{2\nu}\xrightarrow{a.s.}\sigma_0^2/\alpha_0^{2\nu},$$

and its asymptotic distribution is given by

$$\sqrt{n}(\hat\sigma_n^2/\hat\alpha_n^{2\nu}-\sigma_0^2/\alpha_0^{2\nu})\xrightarrow{d}\mathcal{N}(0,2(\sigma_0^2/\alpha_0^{2\nu})^2).$$

Generally speaking, when the range of dependence of the random field is large with respect to the spatial extent of the sampling region, fixed domain asymptotics provide a very accurate description of the behavior of the ML estimate of the microergodic parameter [83]. Extensions of these results to the case where Z is observed with Gaussian errors can be found in Tang et al. [157], while results for a space-time version of the Matérn model can be found in Ip and Li [76] and Faouzi et al. [53]. Finally we highlight that the efficient estimation of the microergodic parameter assuming the smoothness parameter unknown is still an open problem; some promising results in this direction can be found in Loh et al. [106].

A recent article [105] relaxes the conditions imposed by [157] where the latter assumes that ν is known, in concert with some technical assumptions. Recent contributions deal with Bayesian fixed domain asymptotics for Matérn Gaussian random fields, and we mention [95] and more recently [96].

3.2 Prediction and the Screening Effect

The equivalence of Gaussian measures within the Matérn class has consequences for prediction of $Z(\boldsymbol{x}_0)$ at an unobserved location $\boldsymbol{x}_0\in D\setminus X_n$; these consequences will now be discussed. In what follows, ν is supposed known and fixed, and we consider the setting where σ and α are *misspecified*. That is, we suppose Z is a Gaussian field with Matérn covariance $\sigma_0^2\mathcal{M}_{\nu,\alpha_0}$, and we consider the performance of the predictor (12) when a Matérn model $\sigma_1^2\mathcal{M}_{\nu,\alpha_1}$ is used. This situation is typical, since the true parameters σ_0 and α_0 of the data-generating process will be unknown in general. Our theoretical setting will be fixed domain asymptotics.

Note, first, that (12) does not depend on the value of σ_1, but does depend on the value of the parameter α_1 (and the parameter ν, but this parameter is fixed). This dependence will be emphasised using the notation $\boldsymbol{c}_n(\alpha_1)$ and $\boldsymbol{R}_n(\alpha_1)$. Under the Gaussian measure $P(\sigma_0^2\mathcal{M}_{\nu,\alpha_0})$ associated with the *true* model $\sigma_0^2\mathcal{M}_{\nu,\alpha_0}$, the mean squared error of the predictor $\widehat{Z}_n(\alpha_1)$ is given by

$$\begin{aligned}&\mathrm{VAR}_{\alpha_0,\sigma_0^2}\left[\widehat{Z}_n(\alpha_1)-Z(\boldsymbol{x}_0)\right]\\&=\sigma_0^2\Big(1-2\boldsymbol{c}_n(\alpha_1)^\top\boldsymbol{R}_n(\alpha_1)^{-1}\boldsymbol{c}_n(\alpha_0)\\&\qquad+\boldsymbol{c}_n(\alpha_1)^\top\boldsymbol{R}_n(\alpha_1)^{-1}\boldsymbol{R}_n(\alpha_0)\boldsymbol{R}_n(\alpha_1)^{-1}\boldsymbol{c}_n(\alpha_1)\Big),\end{aligned}$$

while if there is no misspecification then the previous expression reduces to

$$\begin{aligned}&\mathrm{VAR}_{\alpha_0,\sigma_0^2}\big[\widehat{Z}_n(\alpha_0)-Z(\boldsymbol{x}_0)\big]\\&\qquad=\sigma_0^2\big(1-\boldsymbol{c}_n(\alpha_0)^\top\boldsymbol{R}_n^{-1}(\alpha_0)\boldsymbol{c}_n(\alpha_0)\big).\end{aligned} \tag{18}$$

Under regularity conditions, and for fixed domain asymptotics, Stein [147] shows that both asymptotically efficient prediction and asymptotically correct estimation of prediction variance hold when the two Gaussian measures $P(\sigma_i^2\mathcal{M}_{\nu,\alpha_i})$, $i=0,1$ are equivalent, *i.e.* (17). Specifically,

$$\frac{\mathrm{VAR}_{\sigma_0^2,\alpha_0}\big[\widehat{Z}_n(\alpha_1)-Z(\boldsymbol{x}_0)\big]}{\mathrm{VAR}_{\sigma_0^2,\alpha_0}\big[\widehat{Z}_n(\alpha_0)-Z(\boldsymbol{x}_0)\big]}\xrightarrow{a.s.}1 \tag{19}$$

and

$$\frac{\mathrm{VAR}_{\sigma_1^2,\alpha_1}\big[\widehat{Z}_n(\alpha_1)-Z(\boldsymbol{x}_0)\big]}{\mathrm{VAR}_{\sigma_0^2,\alpha_0}\big[\widehat{Z}_n(\alpha_1)-Z(\boldsymbol{x}_0)\big]}\xrightarrow{a.s.}1. \tag{20}$$

The implication of (19) is that, under the true model, if the correct value of ν is used, any value of α_1 will give asymptotic efficiency. The implication of (20) is stronger and guarantees that using the misspecified predictor under the correct and misspecified models is asymptotically equivalent from mean squared error point of view. Note that these kind of results does not consider the uncertainty associated with the covariance parameters of the misspecified model. Kaufman and Shaby [83] show that (20) still holds by considering the ML estimator of the variance $\hat\sigma_n^2=\boldsymbol{Z}_n^\top\boldsymbol{R}_n^{-1}(\alpha_1)\boldsymbol{Z}_n/n$ in place σ_1^2.

Conditions of equivalence of two Gaussian measures based on a space-time [76] and bivariate [13] version of the Matérn model have also been established. Next, we consider a practically important aspect of prediction; the co-called screening effect.

Screening Effect. The *screening effect* refers to the phenomenon where the predictor (12) depends almost exclusively on those observations that are located nearest to the predictand [150]. As such, the screening effect is an important tool that can be used to mitigate the computational burden of evaluating (12) in the presence of big datasets. This issue has traditionally been an important subject in geostatistics [110, 111, 112, 39]. Indeed, Matheron [110, 111], in the School of Geostatistics at the Ecole des Mines, developed a first formalisation of screening effect, referring to situations where the observations located far from the predictand receive a zero kriging weight. Matheron's definition has a direct connection with the Markov property on the real line, which happens when kriging is performed under the exponential model (indeed, $\mathcal{M}_{1/2,\alpha}$).

M. Stein [149, 150, 152, 153] adopts an alternative definition of the screening effect that will now be described. Let Z be a mean-square continuous, zero mean and weakly stationary Gaussian random field on $\mathbb{R}^d$. Let $e(X_n)$ be the error of the predictor (12) of $Z(\boldsymbol{x}_0)$ based on $\boldsymbol{Z}_n$. Two choices for the set X_n of observation locations will be considered, and to this end we let F_ϵ, N_ϵ be sets, indexed by $\epsilon > 0$, such that N_ϵ contains the nearest observations to the predictand, and F_ϵ the furthest observations. Then Stein [150] says that N_ϵ *asymptotically screens out* F_ϵ when

$$(21) \qquad \lim_{\epsilon \downarrow 0} \frac{\mathbb{E}\, e(N_\epsilon \cup F_\epsilon)^2}{\mathbb{E}\, e(N_\epsilon)^2} = 1.$$

A thorough discussion of the implications of this definition can be found in Porcu et al. [130], where nontrivial differences between fixed domain and increasing domain asymptotics are reported.

The spatial configuration of the sampling point X_n determines whether the screening effect will hold. Porcu et al. [130] refer to a *regular scheme* as one for which $F_\epsilon = \{\epsilon(\boldsymbol{x}_0 + j)\}$, for $j \in \mathbb{Z}^d$ and N_ϵ being the restriction of F_ϵ to some fixed region with $\boldsymbol{x}_0$ in its interior, assuming $\boldsymbol{x}_0 \notin \mathbb{Z}^d$. For regular schemes, Stein [150] established (21) whenever *the spectrum $\widehat{K}$ varies regularly at infinity [26] in every direction with a common index of variation* [quoted from 130]. However, this condition may not be useful for space-time processes, where differentiability properties in the space and time coordinates are not necessarily identical. To overcome such a problem, we instead consider an *irregular scheme*: for $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ being distinct nonzero elements of $\mathbb{R}^d$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ distinct elements of $\mathbb{R}^d$, $\boldsymbol{x}_0 = \boldsymbol{0} \in \mathbb{R}^d$ and $\boldsymbol{y}_0 \in \mathbb{R}^d$ being nonzero, we have $N_\epsilon = \{\epsilon \boldsymbol{x}_1, \ldots, \epsilon \boldsymbol{x}_n\}$ and $F_\epsilon = \{\boldsymbol{y}_0 + \epsilon \boldsymbol{y}_1, \ldots, \boldsymbol{y}_0 + \epsilon \boldsymbol{y}_N\}$. The *Stein hypothesis* [termed in 130]

$$(22) \qquad \forall R > 0, \qquad \lim_{\|\boldsymbol{\omega}\| \to \infty} \sup_{\|\boldsymbol{\tau}\| < R} \left| \frac{\widehat{K}(\boldsymbol{\omega} + \boldsymbol{\tau})}{\widehat{K}(\boldsymbol{\omega})} - 1 \right| = 0,$$

provides a sufficient condition for the screening effect in this setting (under some mild additional conditions on $\widehat{K}$ and N_ϵ), which can be verified in dimensions $d = 1$ and $d = 2$ for mean-square continuous but non-differentiable random fields, for some specific designs N_ϵ [152]. The Matérn model with $K = \mathcal{M}_{\alpha,\nu}$ admits a simple expression for its spectrum [2, 11.4.44]:

$$(23) \quad \widehat{\mathcal{M}}_{\nu,\alpha}(z) = \frac{\Gamma(\nu + d/2)}{\pi^{d/2}\Gamma(\nu)} \frac{\alpha^d}{(1 + \alpha^2 z^2)^{\nu + d/2}}, \quad z \geq 0,$$

from which (22) can be verified.

The screening effect can thus be established for the Matérn model, under both regular and irregular schemes, justifying the use of "local" approximations to the predictor (12).

3.3 Matérn on Manifolds and Graphs

Let M be a general manifold. A pragmatic question is whether the Matérn correlation function (9) can be composed with a suitable metric g, defined on the manifold, to preserve positive definiteness over M. For the case of the sphere, a natural metric is the geodesic distance; the length of the arc connecting any pair of points located over the spherical shell. For this metric, $(x, y) \mapsto \mathcal{M}_{\nu,\alpha}(g(x, y))$ is a correlation function only for $0 < \nu \leq 1/2$ [61]. This limitation is emphasised in Alegría et al. [3], who propose the $\mathcal{F}$ family, a model that is valid on the sphere, and having the same properties as the Matérn function in terms of mean-square differentiability and fractal dimension. The Matérn function on other general manifolds has been studied by Li et al. [98]. Guinness and Fuentes [68] propose a spectral expansion to define a covariance function that mimics the Matérn model, but this construction is criticised by Lindgren et al. [101] as being incorrect as the spectral expansion does not reproduce the same properties of the Matérn model.

Unfortunately, it seems that the limited applicability of the Matérn model on any space that is not a flat surface extends to more abstract settings as well. An elegant isometric embedding argument in Anderes et al. [8] proves that the restriction $0 < \nu \leq 1/2$ is required when the input space is a graph with Euclidean edges. A more general argument in Menegatto et al. [113] proves that the same restriction is inherited for a general quasi metric space endowed with a geodesic metric. The notable effort by Bolin and Kirchner [29] provides a model that is once differentiable over metric graphs. It is reasonable to conclude that some form of the SPDE approach, which we discuss next in Section 4.2, is needed in general to extend the Matérn model to a general manifold.

4. THE MATÉRN MODEL IN COMPUTATIONAL STATISTICS

This section explores the interaction of the Matérn model with computational statistics, starting with numerical methods for *implementation* of the Matérn model

(Sections 4.1, 4.2 and 4.3), and then turning to uses of the Matérn model to *facilitate* numerical computation itself (Section 4.4).

4.1 Implementation as an SDE

The Matérn model admits a *state space* representation as an SDE, which enables efficient computational techniques from the signal processing literature to be employed for simulation, estimation and prediction. Indeed, focusing on dimension $d = 1$, and letting

$$\mathbf{Z}(x) = (Z, \mathrm{d}Z/\mathrm{d}x, \ldots, \mathrm{d}^k Z/\mathrm{d}x^k),$$

the Matérn model $\mathcal{M}_{\nu,\alpha}$ with $\nu = k + 1/2$ admits the characterisation

$$\mathrm{d}\mathbf{Z} = \begin{pmatrix} 0 & 1 & & \\ & \vdots & \vdots & \\ & & 0 & 1 \\ -a_0 & -a_1 & \ldots & -a_{k-1} \end{pmatrix} \mathbf{Z}\,\mathrm{d}x + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \mathrm{d}\mathcal{W}$$

where $a_i = {}_{k+1}\mathrm{C}_i \cdot \alpha^{-k-1+i}$, the $._{}\mathrm{C}_{.}$ are binomial coefficients, and $\mathcal{W}(x)$ represents a zero-mean white noise process on $x \in \mathbb{R}$ [72]. The advantage of state space formulations is that both estimation and prediction can be performed in a *single pass* through the data, at linear $O(n)$ cost, using familiar Kalman updating equations as described in Sarkka et al. [136] and in further detail in Chapter 6 of Hennig et al. [74]. Similar characterisations for higher dimensions, including spatio-temporal versions of the Matérn model, can be found in Sarkka et al. [136], though we note these retain linear complexity only in the number of time steps; complexity is cubic in the size of the spatial grid. The SPDE approach can offer a solution in this respect, and we discuss this next.

4.2 Implementation as an SPDE

A major reason for the continued popularity of the Matérn model is the availability of efficient and scalable numerical methods for simulation, due in large part to Lindgren et al. [102]. These authors consider the SPDE

$$(\alpha^{-2} - \Delta)^{\gamma/2} Z(\boldsymbol{x}) = \mathcal{W}(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^d, \tag{24}$$

where $\alpha > 0$, Δ is the Laplacian, and $\mathcal{W}$ is a Gaussian white noise on $\mathbb{R}^d$, so that $\mathrm{Cov}\,(\mathcal{W}(A_1), \mathcal{W}(A_2)) = |A_1 \cap A_2|$, where A_i are subsets of $\mathbb{R}^d$, $i = 1, 2$, and where $|\cdot|$ is the volume integral. Whittle [168] and Whittle [169] proved that the solution to (24) is a Gaussian field with Matérn covariance $\sigma^2 \mathcal{M}_{\nu,\alpha}$ with parameters α (as before) and

$$\sigma^2 = \frac{\Gamma(\nu)\alpha^{2\nu}}{\Gamma(\nu + d/2)(4\pi)^{d/2}}, \qquad \nu = \gamma - d/2.$$

This perspective offers two insights; first, tools developed for the numerical approximation of SPDEs can be brought to bear on the Matérn model, and second, there is a clear path to generalise the definition of the Matérn model to any (planar or non planar) manifold on which the analogous SPDE may be defined. (For example, Jansson et al. [77] take this perspective to generalise the Matérn model to the sphere $\mathbb{S}^d$.)

To provide a computationally convenient approximation to (24), Lindgren et al. [102] considered the weak solution to (24) and approximation of the weak solution using basis functions with compact support over a compact domain $\Omega \subset \mathbb{R}^d$ (specifically, a *Galerkin* approximation using finite element basis functions was used). As a result, the authors establish a formal route to approximation of the random field Z with a *Gauss–Markov* random field having a *sparse* precision matrix. Sparse matrix algebra enables fast simulation of realisations from the Matérn random field, and fast evaluation of the likelihood (13) (albeit not fast evaluation of the gradient of the likelihood).

The choice of domain Ω introduces boundary effects which must be carefully mitigated. Khristenko et al. [86], Brown et al. [35] provide a solution for the case where γ is an integer; the non-integer case is considered in Bolin and Kirchner [29]. The extension of the Matérn field based on SPDEs to space-time is provided by Cameletti et al. [37] and subsequently by Bakka et al. [14], Clarotto et al. [40], while the multivariate Matérn case has been explored in Bolin and Wallin [32]. Alternative approximations based on Galerkin methods on manifolds have been provided by Lang and Pereira [90]. An interesting approach that allows working on manifolds with huge datasets is proposed by Pereira et al. [122]. The interest in this literature is dual. On the one hand, the technical aspects related to the finite dimensional representation of Gaussian random fields are extremely interesting *per se*. On the other hand, this group of authors is actually driven by providing tools for efficient computation. This is witnessed by the relevant existing R packages (R-INLA, inlabru, and rSPDE for instance) and we refer to the review of Lindgren et al. [101].

Sanz-Alonso and Yang [134] attempt to explain the trade-off between accuracy and scalability in numerical approximation of the Matérn model. Recall that, in the SPDE approach [102], Z in (24) is numerically approximated using a Gaussian process

$$Z_\delta(\boldsymbol{x}) = \sum_{k=1}^{n_\delta} \omega_k \epsilon_k(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega, \tag{25}$$

where ϵ_k are finite element basis functions and the vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{n_\delta})^\top$ is multivariate Gaussian with zero mean and with a sparse precision matrix. The accuracy of the approximation Z_δ is dependent on (a) the compact support of the finite elements basis functions, (b) boundary effects due to the domain Ω, and (c) by the mesh width δ that determines the cardinality n_δ in (25). Most of the earlier literature has considered (25) with n_δ proportional to the sample size n of the dataset being modelled. Sanz-Alonso and Yang [134] adopt a fixed domain asymptotic approach to explain when $n_\delta \ll n$ might be a legitimate strategy. To do so, they consider Gaussian process regression and work under the framework of Bayesian contraction rates. Their results provide justification for specific scalings of n_δ with $n_\delta = o(n)$, provided that the smoothness ν is sufficiently high.

A different path to SPDE and Gauss–Markov random fields was recently taken in Sanz-Alonso and Yang [135], who adopt graph-based discretisations of SPDEs. This approach can be well-suited to working with discrete and unstructured point clouds, such as in machine learning tasks where the data belong to an implicitly defined low-dimensional manifold. A second advantage of this approach is that an explicit triangulation of the domain is not required.

4.3 Approximate Likelihood and the Matérn Model

In estimating the parameters of the Matérn model using ML (14), numerical optimisation is required. Although generic optimisation routines can be used, an often better approach is to first construct a cheap approximation to the likelihood, which can then be more readily maximised. Indeed, approximate likelihoods are essential when dealing with large datasets, since the evaluation of (13) requires computing the inverse and the determinant of the correlation matrix, usually via the Cholesky decomposition at complexity $O(n^3)$ and storage cost $O(n^2)$.

Perhaps the most successful approximation is *Vecchia's method* [164], which has attracted a remarkable amount of attention in recent times [inc. 148, 46, 47, 66, 45]. The Vecchia approximation can be used with any correlation model and its basic idea is is to replace (13) with a product of Gaussian conditional distributions, in which each conditional distribution involves only a small subset of the data. This approximation requires that the data are *ordered* and the number m of 'previous' data on which to condition is to be specified. Generally, larger m entails more accurate and computationally expensive approximation, while the choice of ordering affects the accuracy of the approximation [66]. The Vecchia method provides a sparse approximation to the Cholesky factor of the precision matrix, such that the approximate likelihood can be computed in $O(nm^3)$ time and with $O(nm^2)$ storage cost. See the recent review of Katzfuss and Guinness [82] for further detail. The Vecchia likelihood can be viewed as a specific instance of a more general class of estimation methods called quasi- or composite likelihood [103, 163] that have been widely used for the estimation of Gaussian fields with the Matérn model [50, 24, 12].

An alternative method of mitigating the computational burden of ML estimation is *covariance tapering* [57]. The basic idea is to multiply the Matérn model with a compactly supported correlation function, resulting in a 'modified' Matérn model with compact support. This induces sparseness in the associated covariance matrix, so that algorithms for sparse matrices can be exploited for a computationally efficient evaluation of the Cholesky decomposition [57]. However, some authors [23, 21] suggest that tapering might be an obsolete approach in view of the fact that flexible compactly supported models that include the Matérn model as a special case have been recently proposed; see Section 8. A comprehensive review of the likelihood approximations is beyond the scopes of this paper, so we refer the reader to Sun et al. [155] and Heaton et al. [73] for further detail.

4.4 The Matérn Model *for* Bayesian Computation

In the last decade there has been increasing interest in the use of kernel methods for solving PDEs. Consider a system

$$\begin{aligned} \mathcal{A}u &= f \qquad \text{in } \Omega \\ \mathcal{B}u &= g \qquad \text{on } \partial\Omega \end{aligned}$$

specified by a differential equation involving $\mathcal{A}$ and f, and initial or boundary conditions specified by $\mathcal{B}$ and g. Dating back at least to Fasshauer [55] in the deterministic setting, and reinterpreted through a Bayesian lens by authors such as Cockayne et al. [41], one can seek an approximation to the strong solution $u : \Omega \to \mathbb{R}$ by modelling u as *a priori* a Gaussian random field and conditioning that field to satisfy the differential equation at locations $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\} \subset \Omega$ and satisfy the boundary conditions at locations $\{\boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n\} \subset \partial\Omega$. The conditional mean of this process coincides with the *symmetric collocation* method introduced by Fasshauer [55], which we return to in Section 6.1, while the conditional variance provides probabilistic uncertainty quantification for the solution, expressing the uncertainty that remains as a result of using only a finite computational budget. To implement these methods, one requires a Gaussian process whose sample paths possess sufficient regularity for the operation of conditioning on the derivative $\mathcal{A}u$ to be well-defined. On the other hand, assuming excessive smoothness could lead to over-confident uncertainty quantification. One therefore requires a kernel with customisable smoothness, which can be adapted to the differential equation at hand. The Matérn class satisfies this requirement, but is not alone in doing so; we continue discussion of this point in Section 7.

A specific PDE that has received considerable recent attention in the Bayesian statistical community is the *Stein equation*, for which $\mathcal{A}u = c + p^{-1}\nabla \cdot (p \nabla u)$, where p is the probability density function of a posterior distribution of interest, f is a function whose posterior expectation we seek to compute, and c is a constant. If the Stein equation has a solution, then c *must* be the value of the posterior expectation we seek. This has motivated several efforts to numerically solve the Stein equation, as a more direct alternative to first approximating p (for example using Markov chain Monte Carlo) and then using the approximation of p to approximate the expectation of interest. In this context kernel methods are typically used [118, 146] and in particular the kernel should have smoothness that is two orders higher than that of the function f whose expectation is of interest, since the Stein equation is a second-order PDE. The generalisation of the Stein equation to Riemannian manifolds was considered in [18], who advocated for the use of kernels with customisable smoothness that reproduce Sobolev spaces of functions on the manifold, such as the (manifold generalisation of the) Matérn model. The connection between the Matérn model and Sobolev spaces is set out in Section 6.1.

5. FLEXIBLE MODELLING WITH MATÉRN

One might object that the Matérn model is insufficiently flexible for many statistical applications, being limited to scalar-valued random fields that are stationary, isotropic and Gaussian. However, the Matérn model is also an important building block for many more sophisticated models, some of which will now be described. This is a rich literature, and our discussion is necessarily succinct; an extended version of this section can be found in [125].

5.1 Scalar Valued Random Fields

Let us start by discussing models for scalar-valued random fields that build on the Matérn model. Note that one can trivially introduce non-zero mean functions into the Matérn model, or combine (additively or multiplicatively) kernels to obtain a potentially more expressive kernel; we will not dwell on either point.

To relax the isotropy assumption of the Matérn model, [6] consider scale mixtures that take into account preferential directions in which spatial dependence develops. On the other hand, the case of space-time models requires special treatment, and non-separable versions of the Matérn kernel are described in Gneiting [60], Zastavnyi and Porcu [179].

The stationarity assumption was relaxed in a parametric manner in Paciorek and Schervish [120], and then in a nonparametric manner in Roininen et al. [133]. An attempt to strike a balance between the computational tractability of parametric models and the flexibility of nonparametric models was reported in Wilson et al. [171], who proposed *input warping* to transform the inputs to the Matérn model using a neural network.

The Gaussian assumption can be relaxed through *output warping*, meaning transformation of the form $\tilde{Z}(\boldsymbol{x}) = w(Z(\boldsymbol{x}))$ where $w(\cdot)$ is a nonlinear map from $\mathbb{R}^d$ to $\mathbb{R}^d$. The covariance function of $\tilde{Z}$ will not be Matérn in general, when the covariance function of Z is Matérn , but if w is sufficiently regular then the smoothness properties of Z transfer to $\tilde{Z}$. The question of whether there exist non-Gaussian processes whose covariance function is nevertheless of Matérn class was answered positively in Åberg and Podgórski [1]. Yan and Genton [175] have proposed *trans-Gaussian* random fields with Matérn covariance function. Bolin [28] and subsequently Wallin and Bolin [165] provided SPDE-based constructions for non-Gaussian Matérn fields. General classes of non-Gaussian fields with covariance $g(\mathcal{M}_{\nu,\alpha})$, for $g(\cdot)$ a suitable function that preserves the positive definiteness and smoothness properties of the Matérn model, have been provided for instance by Palacios and Steel [121], Xua and Genton [174], Bevilacqua et al. [22], Morales-Navarrete et al. [114].

An important extension of the Matérn model, which has received recent attention, is to random fields on spaces for which classical notions of smoothness are not well-defined. For example, Anderes et al. [8] consider graphs with Euclidean edges, equipped with either the geodesic distance over the graph, or the resistance metric. Menegatto et al. [113] provide a generalisation of this setting by considering quasi-metric spaces. Bolin et al. [31] adopt a different approach to build random fields with their covariance structure on metric graphs. Space-time version of the Matérn model, for graphs with Euclidean edges, have been considered by Tang and Zimmerman [156] and Porcu et al. [129]. These efforts considerably extend the applicability of the Matérn model.

The Matérn covariance function decays exponentially with distance, which can be inappropriate for modelling processes that involve long memory. Several approaches have been developed to modify the tails of the Matérn correlation function while preserving many of its desirable characteristics; we describe these in Section 7.

[67] considers Gaussian random fields defined for lattices $\mathbb{Z}^d$ with a covariance function that is the restriction of the Matérn covariance to $\mathbb{Z}^d$. The resulting spectrum is smoothed version of the spectral density associated with the Matérn covariance. For this specific situation, the SPDE approximation can overestimate the scale, α. Yet, it is not clear how this message extends to Gaussian fields that are continuously indexed in $\mathbb{R}^d$.

5.2 Vector-Valued Random Fields

There has been a plethora of approaches related to multivariate spatial modeling, and the reader is referred to Genton and Kleiber [58]. Here, the isotropic covariance function $\boldsymbol{K} : [0,\infty) \to \mathbb{R}^{p\times p}$ is matrix-valued. The elements on the diagonal, K_{ii}, are called *auto-covariance* functions, and the elements K_{ij}, $i \neq j$, are called *cross-covariance* functions. Gneiting et al. [62] proposed a multivariate Matérn model

$$K_{ij}(x) = \sigma_{ii}\sigma_{jj}\rho_{ij}\mathcal{M}_{\nu_{ij},\alpha_{ij}}(x), \qquad x \geq 0, \tag{26}$$

where σ_{ii}^2 is the variance of Z_i, the ith component of a multivariate random field in $\mathbb{R}^p$, and ρ_{ij} is the collocated correlation coefficient. There are restrictions on the parameters ν_{ij}, α_{ij} and ρ_{ij} required to ensure positive definiteness, and often the restrictions on the collocated correlations coefficients ρ_{ij} are rather strict. This last remark has motivated alternative approaches, and the reader is referred to Apanasovich et al. [10] and more recently to Emery et al. [52]. Extensions to multivariate space-time Matérn structures have been provided by Allard et al. [5] and through a technical approach by Porcu et al. [127]. Multivariate nonstationary Matérn functions have been proposed by Kleiber and Nychka [87]. Multivariate Matérn models with *dimple* effect have been studied by Alegría et al. [4]; a 'dimple' in a space-time covariance model refers to the case when $\mathrm{Cov}(Z(\boldsymbol{x},t), Z(\boldsymbol{x}',t'))$ is bigger than $\mathrm{Cov}((Z(\boldsymbol{x},t), Z(\boldsymbol{x}',t))$, which requires special mathematical treatment.

Multivariate Matérn modeling on graphs has been recently investigated in Dey et al. [49], who propose a class of multivariate graphical Gaussian processes through *stitching*, a construction that gets multivariate covariance functions from the graph, and ensures process-level conditional independence between variables. When coupled with the Matérn model, stitching yields a multivariate Gaussian process whose univariate components are Matérn Gaussian processes, and which agrees with process-level conditional independence as specified by the graphical model. Stitching can offer massive computational gains and parameter dimension reduction. An ingenious approach to Gaussian process construction involving the Matérn covariance function has been recently proposed by Li et al. [97], who considered a product space involving the d-dimensional Euclidean space cross an abstract set that allows to index group labels.

5.3 Directions, Shapes and Curves

The Matérn model has an important role in the study of directional processes, with Banerjee et al. [17] formalising the notions of directional *finite difference processes* and *directional derivative processes* with special emphasis on the Matérn model. The Matérn model also has a role in shape analysis, where Banerjee and Gelfand [15] introduced *Bayesian wombling* to measure *spatial* gradients related to curves through 'wombling' boundaries, and approach taken further in Halder et al. [70]. The smoothness properties of the Matérn model are ideally suited to such a framework. Modeling approaches to *temporal* gradients using the Matérn model have been proposed by Quick et al. [132]. Related to these approaches, the smoothness parameter ν of the Matérn model plays a central role in the recent paper by Halder et al. [70], who analyse random surfaces in order to explain latent dependence within a response variable of interest.

This represents a short tour of *statistical* applications of the Matérn model, but its reach goes well beyond statistics, and we explore the importance of the Matérn model to related fields next.

6. THE MATÉRN MODEL OUTSIDE STATISTICS

This section explores the impact of the Matérn model on numerical analysis and approximation theory (Section 6.1), machine learning (Section 6.2), and probability theory (Section 6.3).

6.1 Numerical Analysis and Approximation Theory

The problem considered here is to *reconstruct* a real-valued function f defined on a domain $D \subset \mathbb{R}^d$ from given *data values* $y_i = f(\boldsymbol{x}_i)$ available at a set $X_n = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ of distinct *data locations*. In contrast to the statistical exposition in Section 3.1, from a numerical analysis standpoint these data are not assumed to be random in any way. Nevertheless, many of the mathematical expressions that we previously motivated from a statistical perspective appear also in the solution of this numerical task. The data vector $\boldsymbol{Z}_n$ is reinterpreted as $\boldsymbol{Z}_n = (f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_n))^\top$ and the task is to approximate the value $f(\boldsymbol{x})$ of the unknown function f at an unsampled location $\boldsymbol{x} \in D \setminus X_n$. A natural solution is a minimal-norm interpolant

$$s_{f,X_n,K} = \underset{s \in \mathcal{H}(K)}{\arg\min} \|s\|_{\mathcal{H}(K)} \quad \text{s.t.} \quad \begin{array}{r} s(\boldsymbol{x}_i) = f(\boldsymbol{x}_i), \\ i = 1, \dots, n, \end{array}$$

which we recall was the third optimality property referred in Section 3. Thus, using again the *kernel matrix* $\boldsymbol{R}_n = [K(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1}^n$, the system $\boldsymbol{R}_n \boldsymbol{b} = \boldsymbol{Z}_n$ is solved for a fixed coefficient vector $\boldsymbol{b}$ that determines a linear combination

$$s_{f,X_n,K}(\boldsymbol{x}) = \sum_{i=1}^n b_i K(\boldsymbol{x}_i, \boldsymbol{x}), \qquad \boldsymbol{x} \in D,$$

in the span of the *translates* $K(\boldsymbol{x}_i, \cdot)$. This follows easily from the reproduction formula (3) and (2). The above formula is identical to (12) when setting $\boldsymbol{x} = \boldsymbol{x}_0$, and the resulting value $s_{f,X_n,K}(\boldsymbol{x})$ is interpreted as a numerical approximation to $f(\boldsymbol{x})$. The log-likelihood function (13) can equivalently be viewed as penalising the norm of the interpolant, since $\|s_{f,X_n,K}\|^2_{\mathcal{H}(K)} = \boldsymbol{Z}_n^\top \boldsymbol{R}_n^{-1} \boldsymbol{Z}_n$.

The fourth optimality principle in Section 3 corresponds here to the fact that the norm of the error functional $\epsilon_{\boldsymbol{x}} : f \mapsto f(\boldsymbol{x}) - s_{f,X_n,K}(\boldsymbol{x})$ in the dual space $\mathcal{H}(K)^*$ of $\mathcal{H}(K)$ is minimal under all linear reconstruction algorithms in $\mathcal{H}(K)$ that use the same data $\boldsymbol{Z}_n$. The key tool is the *power function* P_{K,X_n}, defined for all $\boldsymbol{x} \in D$ by

$$P_{K,X_n}(\boldsymbol{x})$$
$$= \sup \left\{ f(\boldsymbol{x}) \,:\, f \in \mathcal{H}(K),\ f(X_n) = 0,\ \|f\|_{\mathcal{H}(K)} \le 1 \right\}$$

It has the property $P_{K,X_n}(\boldsymbol{x}) = \|\epsilon_{\boldsymbol{x}}\|_{\mathcal{H}^*(K)}$ and leads to optimal error bounds of the form

$$|f(\boldsymbol{x}) - s_{f,X_n,K}(\boldsymbol{x})| \le P_{K,X_n}(\boldsymbol{x})\|f\|_{\mathcal{H}_K}.$$

for all $\boldsymbol{x} \in D$ and $f \in \mathcal{H}(K)$. It can be numerically calculated using the kernel matrix based on $X_n \cup \{\boldsymbol{x}\}$, but we omit the detail. Strikingly, the power function coincides with the square root of the *kriging variance* [142], giving the variance of the kriging error at $\boldsymbol{x}$ for given data locations X_n and kernel K.

Analysis of the approximation error in this context thus reduces to analysis of the power function, and in turn analysis of the space $\mathcal{H}(K)$. From (4) and (23), the RKHS generated by the Matérn kernel $\mathcal{M}_{\nu,1}$ has the inner product

$$\langle f, g\rangle_{\mathcal{H}(\mathcal{M}_{\nu,1})} = \int_{\mathbb{R}^d} \frac{\hat{f}(\boldsymbol{\omega})\overline{\hat{g}(\boldsymbol{\omega})}}{(1+\|\boldsymbol{\omega}\|^2)^{\nu+d/2}}\mathrm{d}\boldsymbol{\omega} \tag{27}$$

up to constants, which we recognise as the inner product of the classical *Sobolev space* $W_2^{\nu+d/2}(\mathbb{R}^d)$. By the Sobolev embedding theorem, the elements of this space are well-defined continuous functions whenever $\nu > 0$. This space is a canonical setting for mathematical analysis of PDEs, a connection that we trailed in Section 4.4. Summarising, the use of Matérn kernels yields optimal recovery techniques for functions in Sobolev spaces from given sampled data. Generalised recoveries using derivative data produce *meshless* numerical methods for solving PDEs in Sobolev spaces, including the *symmetric collocation* method which uses derivative data for the PDE based on Wu [173], and shares similar Hilbert space optimality properties Schaback [138]. The use of the Matérn kernel is strongly motivated by the fact that PDE theory often implies that solutions lie in Sobolev spaces. On the other hand, there are also good arguments to replace Matérn kernels by polyharmonics [139, 48].

Plenty of other results on deterministic recovery problems using kernels can be found in Wendland [167], while applications are in Schaback and Wendland [140] and MATLAB programs combined with the essential theory are in Fasshauer and McCourt [56].

In numerical analysis and approximation theory, Matérn and other kernels are normally used for rather large values of their smoothness parameter, because they seek to solve an interpolation rather than a regression task. Narcowich et al. [115] proved that convergence rates then depend on the minimum of the smoothness of the function f providing the data and the kernel; a *misspecified* Matérn kernel, for which the smoothness parameter ν is taken to be too large relative to the smoothness of f, produces an error that converges at the same rate as we would have achieved had ν been correctly specified. On the other hand, Tuo and Wang [160] prove in the same setting that the prediction error becomes more sensitive to the space-filling property of the design points. In particular, optimal convergence rates require also that the *quasi-uniformity* of the experimental design is controlled.

Of course, the use of kernels in numerical analysis and approximation theory requires estimation of kernel parameters. The quantity σ does not arise in the correlation matrix $\boldsymbol{R}_n$, but the scale parameter α has a strong influence on the error of the interpolant. There is a vast literature on *scale estimation* that partially builds on statistical notions like ML (see references in Section 3). On the other hand, specific alternatives to the Matérn model, such as the polyharmonic kernels of Section 7.3, are able to bypass scale estimation due to the remarkable property that the interpolant is independent of the value of the scale parameter used. See Wendland [167] and Section 7.3.

6.2 Machine Learning

Kernel methods are a major strand of machine learning research, where kernels are routinely used to solve a variety of supervised and unsupervised learning tasks. Compared to the interpolatory setting of Section 6.1, data in machine learning are usually observed with noise, necessitating either a likelihood or a loss function to be specified.

The Matérn model is often convenient for the analysis of kernel methods; for example, Tuo et al. [161] provide sufficient conditions for the rates of convergence of the Matérn kernel ridge regression to exceed the standard minimax rates under both the L_2 norm and the norm of the RKHS. However, the presence of noise in the data can pose a substantial challenge to selection of smoothness parameters such as ν in the Matérn model. Karvonen [80] proves that the ML estimate of ν cannot asymptotically *undersmooth* the truth under fixed domain asymptotics; that is, if the true regression function has a Sobolev smoothness $\nu_0 + d/2$, then the smoothness parameter estimate cannot be asymptotically less than $\nu_0 + d/2$, but this in itself it not compelling motivation to use ML [81]. As a result of these additional challenges, standard practice is to keep the kernel general as far as possible when developing methodology, and as far as possible to learn a suitable form for the kernel using the data and model selection criteria. However, recent machine learning methodology for non-Euclidean data hinges on the SPDE approach, and as a consequence the Matérn and related models are explicitly being used.

As the types of data that researchers seek to analyse become more heterogeneous and structured, there has been a demand for flexible Gaussian process models defined on such non-Euclidean domains as manifolds and discrete, graph-based domains. Under the framework of Gaussian processes, Borovitskiy et al. [34] proposed to avoid numerical solution of the SPDE (24) and instead to work with a finite-rank approximation to the Gaussian process model. Specifically, they consider the SPDE in (24) appropriately adapted to a Riemannian manifold M, for which the corresponding Matérn model admits a series expansion of the type

$$\sum_{n=0}^{\infty}\left(\frac{2\nu}{\alpha^2}+\lambda_n\right)^{-\nu-d/2} f_n(\boldsymbol{x})f_n(\boldsymbol{x}'), \qquad \boldsymbol{x},\boldsymbol{x}'\in M$$

where $\{\lambda_n\}_{n=0}^{\infty}$ and $\{f_n\}_{n=0}^{\infty}$ are, respectively, the sequences of eigenvalues and eigenfunctions from the Laplace–Beltrami operator $-\Delta_M$. The authors propose to first solve numerically for the leading eigenfunctions $\{f_i\}_{i=0}^{n}$ of the Laplace–Beltrami operator, and then working with a finite-rank Gaussian process whose realisations are linear combinations of the $\{f_i\}_{i=0}^{n}$. Though solving the eigenproblem may be harder than numerically solving the SPDE, the authors argue that caching of the eigenfunctions can lead to a cost saving in settings where multiple tasks are to be solved on the same manifold. Such an approach is ingeniously extended to undirected graphs by Borovitskiy et al. [33], and has had a direct impact on Gaussian processes defined on neural networks [79], pathwise conditioning of Gaussian processes [172], simulation intelligence in AI [91] and extension to kernel methods withing graphs cross time [116]. Other applications include Thomson sampling in neural information systems [162], Bayesian optimisation in robotics [78], and Gaussian processes regression on metric spaces [89].

6.3 Probability Theory and Stochastic Processes

The Matérn model is well-studied from a probability theory and stochastic process viewpoint. From the perspective of regularity, Scheuerer [141] summarises the properties of Gaussian random fields with Matérn covariance functions; sample paths are k-times differentiable in the mean-square sense if and only if $\nu > k$. Under the same condition, the sample paths have (local) Sobolev space exponent being identically equal to k. Further, a Gaussian random field with Matérn covariance has fractal dimension that is identically equal to $\min(\nu, d)$, for d being the dimension of the Euclidean space on which the random field is defined. For non-Gaussian random fields with Matérn covariance, continuity properties are studied by Kent [85].

Several other properties of the Matérn model have been investigated. Kelbert et al. [84] study fractional random fields under the scenario of stochastic fractional heat equations under a Matérn model; see also Leonenko et al. [94]. Random fields defined on the unit ball embedded in $\mathbb{R}^d$, with a covariance function that is the restriction of the Matérn model to a finite range, were studied in Leonenko et al. [93]. Tensor-valued random fields with an equivalent class of Matérn covariance functions were studied in Leonenko and Malyarenko [92]. Terdik [158] considers angular spectra for non-Gaussian random fields with Matérn covariance function. A recent contribution [159] provides interesting connection between the Matérn model and certain Laplacian ARMA representations of a class of stochastic processes. Lilly et al. [99] show that the Matérn process is a damped version of fractional Brownian motion. Lim and Teo [100] study random fields with a generalised Matérn covariance obtained as the solution to the fractional stochastic differential equation with two fractional orders, enabling the authors to deduce the sample path properties of the associated random field. Space-time extensions of Matérn random fields through stochastic Helmholtz equations are provided by Angulo et al. [9].

According to N. Leonenko[1], a major contributor to this literature, the importance of Matérn model is based on the Duality theorem [71, Theorem 1] which provides an explicit relation between certain classes of characteristic functions of symmetric random vectors and their density. Specifically, the spectral density associated with the Matérn model is by itself a covariance function, called the Cauchy or inverse multiquadric covariance function, that allows to parameterise the Hurst effect of the associated Gaussian random field.

This completes our tour across the scientific landscape through the lens of the Matérn model. Our attention turns now to the future, and promising enhancements that can be made to the Matérn model.

7. ENHANCEMENTS OF THE MATÉRN MODEL

This section described *enhancements* of the Matérn model; covariance functions that share (at least partially) the local properties of the Matérn model while providing additional features and functionality. Here we first introduces the models one at a time, with critical commentary on their features deferred to Section 8.

7.1 Models with Compact Support

Compactly supported covariance models have a long history that can be traced back to Askey [11], who proposed the kernel

$$\mathcal{A}_{\mu,\beta}(x) = \left(1-\frac{x}{\beta}\right)_+^{\mu}, \qquad x \ge 0, \tag{28}$$

with β and μ being strictly positive, and where $(x)_+ = \max(0,x)$ is the *truncated power*. It was shown in that work that $\mathcal{A}_{\mu,\beta}$ belongs to Φ_d for all $\beta > 0$ if and only if $\mu \ge (d+1)/2$. Clearly, the mapping $\boldsymbol{x} \mapsto \mathcal{A}_{\mu,\beta}(\|\boldsymbol{x}\|)$ is compactly supported over a ball with radius β embedded in $\mathbb{R}^d$. As a result, covariance matrices contain exact zero entries whenever the associated states $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ satisfy $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \ge \beta$; the computational advantages of this sparsity are discussed further in Section 8.5.

Matheron's *montée* and *descente* [111] approach was applied by Wendland [166] to the Askey functions, obtaining compactly supported covariance functions with

[1] Personal Communication, January 2023.

higher-order smoothness that are truncated polynomials as functions of $\|\boldsymbol{x}\|$. This strategy was unable to generate integer-order Sobolev spaces in even space dimensions, a problem that was resolved in Schaback [137] who identified the 'missing' Wendland functions. A unified view of Wendland functions was provided by Gneiting [60]. Zastavnyi [176] provided necessary and sufficient conditions for a general class encompassing both ordinary and missing Wendland functions. Buhmann [36] provided a generalisation of Wendland functions, with sufficient parametric conditions that allow the new class to belong to Φ_d for a given d. Those functions, termed *Buhmann* functions, were then studied by Zastavnyi [177] and subsequently by Zastavnyi and Porcu [178], Porcu et al. [131] and Faouzi et al. [54]. Alternative representations and properties of the Wendland functions have been studied by Hubbert [75] and Chernih and Hubbert [38]. Extensions of the Wendland functions to multivariate [126, 44], spatio-temporal [124] and non-stationary processes [88] have also been developed.

A more technical discussion follows, in which we introduce two further classes of correlation functions with compact support, each of which will be the subject of discussion in Section 8.

1. The generalized Wendland ($\mathcal{GW}$) family [59, 177] contains correlation functions with compact support that, as in the Matérn model, admit a continuous parameterisation of smoothness of the underlying Gaussian random field. The $\mathcal{GW}_{\kappa,\mu,\beta}$ model depends on parameters $\kappa \geq 0$ and $\mu, \beta > 0$ through the identity

$$(29)\quad \mathcal{GW}_{\kappa,\mu,\beta}(x) = \frac{\Gamma(\kappa)\Gamma(2\kappa+\mu+1)}{\Gamma(2\kappa)\Gamma(\kappa+\mu+1)2^{\mu+1}}\mathcal{A}_{\kappa+\mu,\beta^2}\left(x^2\right) \times {}_2F_1\left(\frac{\mu}{2}, \frac{\mu+1}{2}; \kappa+\mu+1; \mathcal{A}_{1,\beta^2}\left(x^2\right)\right),$$

where $\mu \geq (d+1)/2+\kappa$ is needed for $\mathcal{GW}_{\kappa,\mu,\beta}$ to belong to the class Φ_d and ${}_2F_1(a,b,c,\cdot)$ is the Gaussian hypergeometric function [2]. Sample paths of the $\mathcal{GW}_{\kappa,\mu,\beta}$ model are k times mean-square differentiable, in any direction, if and only if $\kappa > k - 1/2$ [59], so that κ plays the role of the smoothness parameter in this model. When $\kappa = k \in \mathbb{N}$, $\mathcal{GW}_{k,\mu,\beta}$ factors into the product of the Askey function $\mathcal{A}_{\mu+k,\beta}$ with a polynomial of degree k. This model includes the Wendland functions ($\kappa = k$, a positive integer), as well as the missing Wendland functions ($\kappa = k+1/2$). Theorem 1(3) in Bevilacqua et al. [23] implies that the RKHS induced by $\mathcal{GW}_{\kappa/2-(d+1)/4,\mu,\beta}$, with $\kappa \geq (d+1/2)$, is norm-equivalent to the Sobolev space $W_2^{\kappa}(\mathbb{R}^d)$.

2. The Gauss hypergeometric ($\mathcal{GH}$) family [51] is defined as

$$(30)\quad \mathcal{GH}_{\kappa,\delta,\gamma,\beta}(x) = \frac{\Gamma(\delta-d/2)\Gamma(\gamma-d/2)}{\Gamma(\delta-\kappa+\gamma-d/2)\Gamma(\kappa-d/2)} \times \mathcal{A}_{\delta-\kappa+\gamma-d/2+1,\beta^2}\left(x^2\right) \times {}_2F_1\left(\delta-\kappa;\gamma-\kappa;\delta-\kappa+\gamma-d/2;\mathcal{A}_{1,\beta^2}(x^2)\right).$$

This model has four parameters and it belongs to the class Φ_d for every positive β provided $\kappa > d/2$ with

$$2(\delta-\kappa)(\gamma-\kappa) \geq \kappa, \quad \text{and} \quad 2(\delta+\gamma) \geq 6\kappa+1.$$

Sample paths of the $\mathcal{GH}_{\kappa,\delta,\gamma,\beta}$ model are $\lceil k/2 \rceil$ times mean-square differentiable, in any direction, if and only if $\kappa > (k+d)/2$. The parameter κ thus also controls the smoothness of samples from this model.

The importance of the $\mathcal{GW}$ and $\mathcal{GH}$ models is discussed in Section 8.

7.2 Models with Polynomial Decay

Correlation models with polynomial decay such as the generalized Cauchy [63] or the Dagum models [20] can be useful when modelling data with long-range dependence. However, in using these correlation models one loses control over the differentiability of the the sample paths, a key property of the Matérn model. Ma and Bhadra [107] recently proposed a modification of the Matérn class that allows for polynomial decay, while maintaining the local properties of the conventional Matérn model. The correlation function associated to this model is given by

$$(31)\quad \mathcal{CH}_{\nu,\eta,\beta}(x) = \frac{\Gamma(\nu+\eta)}{\Gamma(\nu)}\mathcal{U}\left(\eta, 1-\nu, \nu\left(\frac{x}{\beta}\right)^2\right),\ x \geq 0,$$

where $\mathcal{U}$ is the confluent hypergeometric function of the second kind [2]. Here $\nu > 0$ controls mean-square differentiability near the origin, as in the Matérn case, while $\eta > 0$ controls the heaviness of the tail. The construction (31) is based on a scale mixture of (a reparameterised version of) the Matérn model involving the inverse-gamma distribution. Ma and Bhadra [107] have shown that this class is particularly useful for extrapolation problems where large distances are predominant.

7.3 Polyharmonic Kernels

Our catalogue of enhancements of the Matérn model finishes with *polyharmonic kernels*, defined as

$$(32)\qquad H_{\nu,d}(x) := \begin{Bmatrix} x^{2\nu-d}\log x & \text{for } 2\nu-d \in 2\mathbb{Z} \\ x^{2\nu-d} & \text{else} \end{Bmatrix}$$

up to the sign $(-1)^{\lfloor \nu-d/2 \rfloor+1}$. As a function of $x = \|\boldsymbol{x}\|$, $\boldsymbol{x} \in \mathbb{R}^d$, the Matérn kernel $\mathcal{M}_{\nu-d/2,1}$ starts with even powers of x followed by $H_{\nu,d}$, and in this sense the two models are related. Up to a constant factor, the generalised Fourier transform of $H_{\nu,d}(\|x\|)$ on $\mathbb{R}^d$ is $\|\omega\|^{-2\nu}$, and then a scale parameter is just another constant factor. This makes kernel-based interpolation by polyharmonics scale-independent. Compare with (23) to see the connection to $\mathcal{M}_{\nu-d/2,\alpha}$ in Fourier space. Stein [151] provides a formal connection between polyharmonic kernels, for which the name *power law* covariance functions is also used, and the Matérn model. Polyharmonic kernels are *conditionally positive definite* of order $\lfloor \nu - d/2 \rfloor + 1$; for a technical definition see Wendland [167]. Instead of Hilbert Spaces, polyharmonic kernels generate *Beppo–Levi spaces*, which share similarities to Sobolev spaces modulo that an additional polynomial space has to be added to enable prediction (Section 3) and interpolation (Section 6.1); see Wendland [167]. In general, polyharmonic kernels arise as covariances in *fractional* Gaussian fields, including forms of Brownian motion [104, Theorem 3.3].

Next our attention turns to a critical discussion of whether such enhancements to the Matérn model are needed.

8. OTHER MODELS

This final section provides critical commentary on the Matérn model and the enhanced versions of the model introduced in Section 7.

8.1 Rigorous Generalisation of the Matérn Model

The Matérn model does not allow for compact support, hole effects (oscillations between positive and negative values) at large distances, or slowly decaying tails suitable for modeling long-range dependence. Most of the enhancements in Section 7 aim to resolve these kind of issues; here we describe how the $\mathcal{GW}$, $\mathcal{GH}$ and $\mathcal{CH}$ models can be viewed as rigorous generalisations of the Matérn model.

Bevilacqua et al. [21] have shown that the Matérn model is a limit case of a rescaled version of the $\mathcal{GW}$ model. In particular they have considered the model $\widetilde{\mathcal{GW}}$ defined as

$$\widetilde{\mathcal{GW}}_{\kappa,\mu,\beta}(x) = \mathcal{GW}_{\kappa,\mu,\beta\left(\frac{\Gamma(\mu+2\kappa+1)}{\Gamma(\mu)}\right)^{\frac{1}{1+2\kappa}}}(x), \qquad x \geq 0,$$

and proved that

$$\lim_{\mu\to\infty} \widetilde{\mathcal{GW}}_{\kappa,\mu,\beta}(x) = \mathcal{M}_{\kappa+1/2,\beta}(x), \quad \kappa \geq 0,$$

with uniform convergence over the set $x \in (0,\infty)$. Figure 1 (first row) depicts the convergence result for $\nu = 0, 1, 2$ (from left to right).

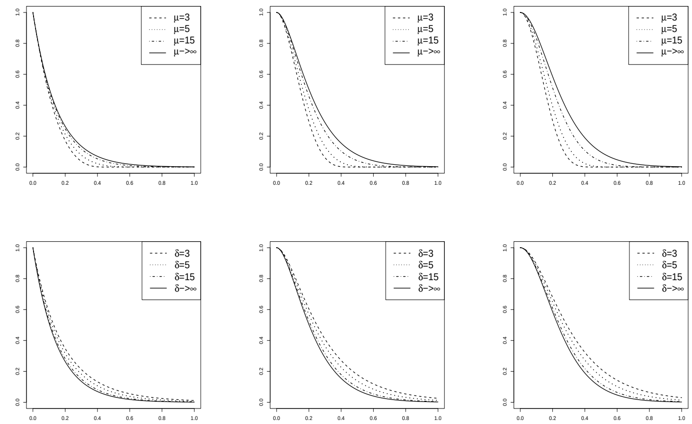

FIG 1. *First row: the $\widetilde{\mathcal{GW}}_{\kappa,\mu,\beta}(x)$ model with $\mu = 3, 5, 15$ and $\mu \to \infty$ (the Matérn model $\mathcal{M}_{\kappa+1/2,\beta}(x)$) with $\beta = 0.15, 0.12, 0.1$ and $\kappa = 0, 1, 2$ (from left to right) respectively. Second row: the $\mathcal{CH}_{\nu,\eta,2\sqrt{\nu(\eta+1)}\beta}(x)$ model with $\eta = 3, 5, 15$ and $\eta \to \infty$ (the Matérn model $\mathcal{M}_{\nu,\beta}(x)$) with $\beta = 0.15, 0.12, 0.1$ and $\nu = 0.5, 1.5, 2.5$ (from left to right) respectively.*

The parameter μ thus allows for switching from compactly supported to globally supported models, and can either be fixed to ensure sparse correlation matrices, or can be estimated based on the dataset. However, this equivalence applies only to smoothness parameters greater than or equal to $1/2$ in the Matérn model, so the full range of the smoothness parameter is not covered. This is unfortunate, since the fractal dimension [a widely used measure of roughness of the sample paths for time series and spatial data; 64] is fully parameterised using the Matérn model when the smoothness parameter lies between 0 and 1. As a consequence, the $\mathcal{GW}$ (or $\widetilde{\mathcal{GW}}$) model cannot fully parameterise the fractal dimension of the random field. This kind of issue can be solved with the $\mathcal{GH}$ model, which includes the $\mathcal{GW}$ model as a special case [51]:

$$\mathcal{GH}_{\frac{d+1}{2}+\nu,\frac{d+\mu+1}{2}+\nu,\frac{d+\mu}{2}+1+\nu,\beta}(x) = \mathcal{GW}_{\nu,\mu,\beta}(x)$$

Letting β, δ and γ tend to infinity in such a way that $\beta/\sqrt{4\delta\gamma}$ tends to $\alpha > 0$, the $\mathcal{GH}$ model (30) converges uniformly to the Matérn model $\mathcal{M}_{\kappa-d/2,\alpha}(x)$, and in this case the *full range* of the smoothness parameter of the Matérn model is covered.

The Matérn model also arises as a special limit case of the $\mathcal{CH}$ model. Specifically, Ma and Bhadra [107] show that

$$\lim_{\eta\to\infty} \mathcal{CH}_{\nu,\eta,2\sqrt{\nu(\eta+1)}\beta}(x) = \mathcal{M}_{\nu,\beta}(x),$$

with convergence being uniform on any compact set. Figure 1 (second row) depicts the convergence result for $\nu = 0.5, 1.5, 2.5$ (from left to right).

The *turning band* operator of Matheron [110] can be applied to a correlation function to create hole effects while retaining positive definiteness of the kernel. An argument in Schoenberg proves that, for an isotropic correlation in $\mathbb{R}^d$, the correlation values cannot be smaller than $-1/d$ [144]. Since the Matérn model is a valid model for all d, this implies that the application of turning bands to the Matérn model will not provide any hole effect. On the other hand, the $\mathcal{GW}$ and $\mathcal{GH}$ models allow for such an effect.

8.2 Estimation of Enhanced Models

ML estimation for the Matérn model are well-understood; here we discuss the extent to which similar results can be obtained for enhancements of the Matérn model.

In the context of increasing domain asymptotics, parameters of the $\mathcal{GW}$ and $\mathcal{CH}$ models can be estimated consistently using ML and the associated asymptotic distribution is known; see Section 3.1.1.

In the context of fixed domain asymptotics, similar to the classical Matérn model, the parameters of the these enhanced models cannot be consistently estimated. For instance, Bevilacqua et al. [23] show that the microergodic parameter of the covariance model $\sigma^2\mathcal{GW}_{\kappa,\mu,\beta}$, assuming κ and μ known, is given by $\mathrm{micro}_{\mathcal{GW}} = \sigma^2/\beta^{2\kappa+1}$. In addition they prove that for a zero mean Gaussian field defined on a bounded infinite set $D \subset \mathbb{R}^d$ ($d = 1, 2, 3$), with covariance model $\sigma_0^2\mathcal{GW}_{\kappa,\mu,\beta_0}$, the ML estimator $\hat\sigma_n^2/\hat\beta_n^{2\kappa+1}$ of the microergodic parameter is strongly consistent, *i.e.*,

$$\hat\sigma_n^2/\hat\beta_n^{2\kappa+1} \xrightarrow{a.s.} \sigma_0^2/\beta_0^{2\kappa+1}.$$

Additionally, for $\mu > (d+1)/2 + \kappa + 3$, its asymptotic distribution is given by

$$\sqrt{n}(\hat\sigma_n^2/\hat\beta_n^{2\kappa+1} - \sigma_0^2/\beta_0^{2\kappa+1}) \xrightarrow{d} \mathcal{N}(0, 2(\sigma_0^2/\beta_0^{2\kappa+1})^2).$$

Analogous for the $\mathcal{GH}$ model proposed are not available at present.

Similarly, Ma and Bhadra [107] show that the microergodic parameter of the covariance model $\sigma^2\mathcal{CH}_{\nu,\eta,\beta}$, assuming ν known, is given by

$$\mathrm{micro}_{\mathcal{CH}} = (\sigma^2\Gamma(\nu+\eta))/(\beta^{2\nu}\Gamma(\eta)).$$

In addition they prove that for a zero mean Gaussian field defined on a bounded infinite set $D \subset \mathbb{R}^d$ ($d = 1, 2, 3$), with covariance model $\sigma_0^2\mathcal{CH}_{\nu,\eta_0,\beta_0}$, the ML estimator $(\hat\sigma_n^2/\hat\beta_n^{2\nu})(\Gamma(\nu+\hat\eta_n)/\Gamma(\hat\eta_n))$ of the microergodic parameter is strongly consistent, *i.e.*,

$$\frac{\hat\sigma_n^2(\Gamma(\nu+\hat\eta_n)}{\hat\beta_n^{2\nu}\Gamma(\hat\eta_n)} \xrightarrow{a.s.} \frac{\sigma_0^2\Gamma(\nu+\eta_0)}{\beta_0^{2\nu}\Gamma(\eta_0)}$$

and, if $\eta_0 > d/2$, its asymptotic distribution is given by

$$\frac{\hat\sigma_n^2(\Gamma(\nu+\hat\eta_n)}{\hat\beta_n^{2\nu}\Gamma(\hat\eta_n)} - \frac{\sigma_0^2\Gamma(\nu+\eta_0)}{\beta_0^{2\nu}\Gamma(\eta_0)} \xrightarrow{d} \mathcal{N}\left(0, 2\left(\frac{\sigma_0^2\Gamma(\nu+\eta_0)}{\beta_0^{2\nu}\Gamma(\eta_0)}\right)^2\right).$$

These results broadly support the use of ML plug-in estimates for these enhanced versions of the Matérn model; the issue of predictive performance is discussed next.

8.3 Prediction with Enhanced Models

If two Gaussian measures are equivalent then the associated predictions and mean squared errors are asymptotically identical (c.f. Section 3.2). To this end, recent results have sought to establish equivalence between Gaussian measures for the Matérn model and enhancements of the Matérn model. Bevilacqua et al. [23] consider the $\sigma_1^2\mathcal{GW}_{\kappa,\mu,\beta}$ model and show that for given $\sigma_1 \geq 0$, $\nu \geq 1/2$, and $\kappa \geq 0$, if $\nu = \kappa + 1/2$, $\mu > d + \kappa + 1/2$ and

$$\sigma_0^2\alpha^{-2\nu} = \left(\frac{\Gamma(2\kappa+\mu+1)}{\Gamma(\mu)}\right)\sigma_1^2\beta^{-(1+2\kappa)}, \tag{33}$$

then $P(\sigma_0^2\mathcal{M}_{\nu,\alpha})$ is equivalent to $P(\sigma_1^2\mathcal{GW}_{\kappa,\mu,\beta})$, for $d = 1, 2, 3$, on the paths of $Z(\boldsymbol{x})$ for $\boldsymbol{x} \in D \subset \mathbb{R}^d$. Thus predictions made using the $\mathcal{GW}$ model with compact support are asymptotically identical to those made using the Matérn model. Likewise, Ma and Bhadra [107] show that for a given $\eta \geq d/2$ and $\nu \geq 0$, if

$$\sigma_0^2\alpha^{-2\nu} = \left(\frac{\Gamma(\nu+\eta)}{\Gamma(\eta)}\right)\sigma_1^2\left(\frac{\beta^2}{2}\right)^{-\nu}, \tag{34}$$

then $P(\sigma_0^2\mathcal{M}_{\nu,\alpha})$ is equivalent to $P(\sigma_1^2\mathcal{CH}_{\nu,\eta,\beta})$, for $d = 1, 2, 3$, on the paths of $Z(\boldsymbol{x})$ for $\boldsymbol{x} \in D \subset \mathbb{R}^d$. Thus predictions made using the $\mathcal{GW}$ model with polynomial tail decay are asymptotically identical to those made using the Matérn model.

If interest is in the predictor (12), but not the predictive uncertainty resulting from the associated Gaussian random field, then it is interesting to note that the stationarity assumption of the Matérn model may not be needed. Stein et al. [154] showed that, under suitable parametric conditions, one can consider $\alpha = 0$ in the Matérn model, and this is equivalent to prediction using the polyharmonic kernels $H_{\nu,d}$ in (32). Theorem 1 in that work shows that if $d \leq 3$ and the parameter ν satisfies condition (2) therein (or $d = 1$), then *it is impossible to distinguish* $\alpha > 0$ *from* $\alpha = 0$ *on a bounded domain.* The above observation reflects the fact that prediction using polyharmonic kernels, like in Section 6.1, is scale-independent. This follows from homogeneity of the Fourier transform and eliminates the need for scale estimation in this context.

8.4 Screening with Enhanced Models

The screening effect extends also to enhanced versions of the Matérn model. For regular schemes, Theorem 1 in Porcu et al. [130] shows that the $\mathcal{GW}$ model allows for an asymptotic screening effect when $\mu > (d+1)/2+\kappa$. This condition is not restrictive, since $\mu \geq (d+1)/2+\nu$ is already required for $\mathcal{GW}_{\kappa,\mu,\beta}$ to belong to the class Φ_d. For irregular schemes the situations is more complicated. For example, for non-differentiable fields in $d=1$, Theorem 1 in Stein [152] in concert with Theorem 1 in Porcu et al. [130] explains that the Askey model $\mathcal{GW}_{0,\mu,\beta}$ allows for a screening effect provided $\mu > 1$. For $d=2$, Theorem 2 in Stein [152] implies that the Askey model allows for screening provided that $\mu > 3/2$. The $\mathcal{GW}$ model satisfies Stein's condition in (1.3) of Porcu et al. [130], which in turn allows the Stein hypothesis (22) to be verified.

The numerical experiments in Porcu et al. [130] suggest that the screening effect is even stronger under enhanced models with compact support, compared to the standard Matérn model. This can deliver computational advantages, which we discuss next.

8.5 Scalable Computation

Scalable computation generally refers to the computational complexity associated with the optimal predictor (12) and/or of the likelihood function (13) when increasing n the number of data.

The eternal fight between statistical accuracy and computational scalability has produced methods that attempt to deal with this notorious trade-off. The discussion that follows focuses specifically on this trade-off in the context of the Matérn model or its enhancements, in particular compactly supported models. General approaches, such as those based on predictive processes [16] and those based on fixed-rank kriging [43], will not be discussed; the interested reader is referred to the review of Sun et al. [155].

The computational complexity associated with the Matérn model is broadly governed by the number of data (n) and partially by the input space dimension (d), the dimension, p of the (scalar or vector) random field. In the case of scalar-valued random fields, we have $p=1$.

These challenges will be considered in turn.

The flexibility of some enhanced models is lost in the case of large domains; the condition $\mu \geq (d+1)/2+\kappa$ in the $\widetilde{\mathcal{GW}}_{\kappa,\mu,\beta}$ model forces the parameter μ to go to infinity with d, which in turn forces $\widetilde{\mathcal{GW}}_{\kappa,\mu,\beta}$ to approach $\mathcal{M}_{\nu,\alpha}$. From this point of view the class $\mathcal{GH}_{\kappa,\delta,\gamma,\beta}$ seems more promising to use for large d. An additional remark is that, for $d \geq 5$, all Gaussian measures with Matérn covariance functions are orthogonal [7]. This has philosophical consequences for Gaussian process regression when the Matérn model is viewed as a *prior* distribution encoding *a priori* belief, since a small change to the kernel parameters results in the entire support of the prior being changed.

In the case of a large number of variables p in a multivariate the model, a large number of parameters needs to be estimated.

The multivariate Matérn model suffers from the fact, not only does the number of parameters increase polynomially with p, but the conditions for validity of the model imply severe restrictions on the collocated correlation coefficient ρ_{ij} in (26). Emery et al. [52] show that such restrictions become extremely severe already with $p=3$. Similar comments apply to other multivariate covariance functions, including the multivariate $\mathcal{GW}$ model in Daley et al. [44].

Finally we consider the case where the number n of data is large, entailing a $O(n^3)$ computational and $O(n^2)$ storage cost associated with the predictor (12) or the likelihood function(13). Several approaches have been proposed to reduce these costs, many of which take advantage of the (approximate) sparsity of the covariance ($\boldsymbol{\Sigma}_n$) or precision ($\boldsymbol{\Sigma}_n^{-1}$), or its Cholesky factor ($\mathrm{ch}(\boldsymbol{\Sigma}_n^{-1})$):

- Sparsity in the covariance matrix $\boldsymbol{\Sigma}_n$ can be directly exploited by using compactly supported models such as the $\widetilde{\mathcal{GW}}_{\kappa,\mu,\beta}$ or the $\mathcal{GH}_{\kappa,\delta,\gamma,\beta}$ families. Such approaches can be useful when the (estimated) compact support is relatively small with respect to the spatial extent of the sampling region, so that approximations are extremely sparse; see below for an empirical investigation of this point.
- The precision matrix $\boldsymbol{\Sigma}_n^{-1}$ associated with the Matérn model is in general non-sparse (except for the case $d=1$ and $\nu=0.5$) but it turns out that the matrix values are in general relatively close to 0, i.e. $\boldsymbol{\Sigma}_n^{-1}$ is *quasi-sparse*. As a consequence, approximating $\boldsymbol{\Sigma}_n^{-1}$ with a sparse matrix can be a good strategy. A notable instance of this approach is the SPDE approach from Section 4.2. This approach can be also motivated from results in numerical linear algebra, which demonstrate that if the elements of a matrix show a property of decay, then the elements of its inverse also show a similar (and faster) behavior [19].
- Vecchia's approximation [164] and its extensions [e.g. 46, 66, 82, 45] imply a sparse approximation of $\mathrm{ch}(\boldsymbol{\Sigma}_n^{-1})$ and are often applied to the Matérn model, although they can be applied to any covariance model. One potential limitation of these method is that they depend on an ordering of the variables and the choice of conditioning sets which determines the Cholesky sparsity pattern [see 66].

It is instructive to numerically investigate the sparseness of matrices associated with enhancements of the Matérn model, and for this we focus on the $\widetilde{\mathcal{GW}}_{\kappa,\mu,\beta}$ model,

	$\kappa=0$								$\kappa=1$								$\kappa=2$						
μ	C	$\mathbf{\Sigma}_n$		$\mathbf{\Sigma}_n^{-1}$		ch$(\mathbf{\Sigma}_n^{-1})$		μ	C	$\mathbf{\Sigma}_n$		$\mathbf{\Sigma}_n^{-1}$		ch$(\mathbf{\Sigma}_n^{-1})$		μ	C	$\mathbf{\Sigma}_n$		$\mathbf{\Sigma}_n^{-1}$		ch$(\mathbf{\Sigma}_n^{-1})$	
		1156	4489	1156	4489	1156	4489			1156	4489	1156	4489	1156	4489			1156	4489	1156	4489	1156	4489
1.5	0.07	98.4	98.3	32.3	1.46	35.7	2.17	2.5	0.11	97.1	96.7	3.80	1.67	6.39	4.44	3.5	0.13	94.7	95.0	2.90	1.31	5.63	6.12
4	0.20	90.1	89.6	45.9	56.0	45.0	54.9	4	0.16	93.4	93.3	34.2	47.1	36.6	48.7	4	0.15	94.8	94.2	10.4	11.9	15.0	22.6
8	0.40	66.4	65.7	56.6	64.2	52.0	60.0	8	0.28	80.9	80.7	69.7	57.0	69.3	59.2	8	0.25	84.7	84.4	42.8	61.0	47.3	64.8
16	0.80	16.4	15.2	59.8	71.5	53.8	66.6	16	0.54	48.1	47.1	64.2	80.1	64.0	76.0	16	0.45	59.1	58.3	51.7	73.0	54.6	74.2
32	1.60	0	0	61.2	79.0	55.8	72.8	32	1.04	1.91	1.61	66.4	83.8	66.5	82.2	32	0.86	11.0	10.1	52.3	75.7	54.8	77.2
120	6.01	0	0	65.2	76.6	58.9	68.2	120	3.82	0	0	66.2	84.6	65.9	81.2	120	3.09	0	0	52.4	75.7	54.1	76.5
∞	∞	0	0	66.2	77.5	58.9	70.0	∞	∞	0	0	66.2	84.9	65.1	80.7	∞	∞	0	0	52.4	75.4	53.6	74.6

TABLE 1

Sparsity (percentage of zero values in the upper triangular part) of the covariance matrix $\mathbf{\Sigma}_n$, and quasi-sparsity (defined in the main text) in the precision matrix ($\mathbf{\Sigma}_n^{-1}$) and its Cholesky factor (ch($\mathbf{\Sigma}_n^{-1}$)) for the $\widetilde{\mathcal{GW}}_{\kappa,\mu,\beta}$ model. The case $\widetilde{\mathcal{GW}}_{\kappa,\infty,\beta}$ corresponds to the Matérn model $\mathcal{M}_{\nu+1/2,\beta}$. The β parameters are chosen so that the practical range of the Matérn model is equal to 0.15.

	$\kappa=0$								$\kappa=1$								$\kappa=2$						
μ	C	$\mathbf{\Sigma}_n$		$\mathbf{\Sigma}_n^{-1}$		ch$(\mathbf{\Sigma}_n^{-1})$		μ	C	$\mathbf{\Sigma}_n$		$\mathbf{\Sigma}_n^{-1}$		ch$(\mathbf{\Sigma}_n^{-1})$		μ	C	$\mathbf{\Sigma}_n$		$\mathbf{\Sigma}_n^{-1}$		ch$(\mathbf{\Sigma}_n^{-1})$	
		1156	4489	1156	4489	1156	4489			1156	4489	1156	4489	1156	4489			1156	4489	1156	4489	1156	4489
1.5	0.20	90.0	89.1	0	0	0	0	2.5	0.28	80.5	80.2	0	0	0	0	3.5	0.35	72.6	71.4	0	0	0	0
4	0.53	48,7	47.3	1.03	9.62	0.58	2.77	4	0.42	65.0	63.7	0	1.03	1.12	0.81	4	0.39	67.2	66.4	0	0	0	0
8	1.07	1.43	1.21	4.14	25.6	1.84	8.15	8	0.75	20.7	19.7	5.91	14.5	4.49	12.9	8	0.67	30.7	29.6	0.71	2.73	2.13	6.90
16	2.13	0	0	15.0	43.5	4.98	21.3	16	1.43	0	0	21.2	45.2	17.1	30.8	16	1.21	1.72	1.29	13.7	25.8	14.8	27.9
32	4.27	0	0	22.8	44.4	12.7	24.3	32	2.78	0	0	37.2	55.0	29.0	41.0	32	2.29	0	0	25.5	23.5	24.8	42.7
120	16.0	0	0	21.0	46.2	8.33	21.9	120	10.2	0	0	40.8	59.8	32.2	46.2	120	8.24	0	0	26.0	10.9	25.9	46.4
∞	∞	0	0	23.4	47.9	9.85	24.0	∞	∞	0	0	42.6	61.3	34.3	48.0	∞	∞	0	0	26.0	18.2	25.9	46.6

TABLE 2

As in Table 1, but with β chosen such that the practical range of the Matérn model is equal to 0.4.

which allows us to switch from a model with compact support of radius

$$C=\beta\left(\frac{\Gamma(\mu+2\kappa+1)}{\Gamma(\mu)}\right)^{\frac{1}{1+2\kappa}}$$

to the Matérn model by increasing the μ parameter. In our experiment, the sparseness of $\mathbf{\Sigma}_n$ and the *quasi-sparseness* of $\mathbf{\Sigma}_n^{-1}$ and $\mathrm{ch}(\mathbf{\Sigma}_n^{-1})$ are reported, the latter being defined as the percentage of values in the upper triangular matrix with absolute value lower than an arbitrary small constant ϵ, and in our example we set $\epsilon = 1.e-8$.

The empirical assessment considers $n = 1,156$ and $n = 4,489$ location sites over $[0,1]^2$, where the points are equally spaced by 0.03 and 0.015 respectively in a regular grid. For $\nu = 0,1,2$, we set β such that the practical range of the Matérn model is equal to 0.15 ($\beta = 0.050, 0.0316, 0.0253$ respectively), and consider increasing $\mu = 1.5+\kappa, 4, 8, 16, 32, 120, \infty$ (with $\widetilde{\mathcal{GW}}_{\kappa,\infty,\beta}$ being the Matérn model $\mathcal{M}_{\kappa+1/2,\beta}$).

The results are reported in Table 1. For the low values $\mu = 1.5, 2.5, 3.5$ and $\nu = 0,1,2$, the covariance matrix is highly sparse, while the sparseness decreases when increasing μ, as expected. There is a clear trade-off between the sparseness of $\mathbf{\Sigma}_n$ and quasi-sparseness of $\mathbf{\Sigma}_n^{-1}$ and $\mathrm{ch}(\mathbf{\Sigma}_n^{-1})$ for each $\nu = 0,1,2$. However, when increasing μ, that is when $\mathbf{\Sigma}_n$ approaches the Matérn covariance matrix, then $\mathbf{\Sigma}_n^{-1}$ or $\mathrm{ch}(\mathbf{\Sigma}_n^{-1})$ tends to be highly quasi-sparse.

We replicate the same experiment but with a practical range of the Matérn model equal to 0.4. This leads to $\beta = 0.133, 0.084, 0.067$ for $\nu = 0,1,2$ respectively. The results are reported in Table 2. The conclusions are the same of the previous setting but in this case, we have lower levels of sparseness for $\mathbf{\Sigma}_n$ and of quasi-sparseness for $\mathbf{\Sigma}_n^{-1}$ and $\mathrm{ch}(\mathbf{\Sigma}_n^{-1})$.

These numerical experiments highlight a clear trade-off between the (quasi-)sparseness of $\mathbf{\Sigma}_n^{-1}$ (or $\mathrm{ch}(\mathbf{\Sigma}_n^{-1})$) and $\mathbf{\Sigma}_n$ when increasing μ for fixed β and ν i.e. when switching from a compactly supported to a globally supported Matérn model. In particular, when $\mu \to \infty$ (the Matérn model), then $\mathbf{\Sigma}_n^{-1}$ is highly quasi-sparse and $\mathbf{\Sigma}_n$ is dense. In contrast, when μ is small then $\mathbf{\Sigma}_n^{-1}$ is not quasi-sparse yet $\mathbf{\Sigma}_n$ is highly sparse. This seems to suggest that sparse precision matrix approximation should work reasonably well for the Matérn model, but could be problematic when handling data exhibiting short compactly supported dependence. In this case a better approach should be to exploit the sparsity of $\mathbf{\Sigma}_n$, as enabled by enhanced versions of the Matérn model.

As a final comment, the evaluation of the Matérn model and its enhancements requires the computation of some special functions such as the modified Bessel function of the second kind, the Gaussian hypergeometric function and the confluent hypergeometric function of the second kind that can be found in different libraries such as the GNU scientific library [65] and the most important statistical softwares including R, MATLAB and Python. For instance, the R package GeoModels [25] implements the computation of the Matérn model and its enhancements for $d = 1, 2$.

9. CONCLUSION

The impact of the Matérn model since its conception has been substantial, and the model continues to be widely used, across a broad range of scientific disciplines and beyond. While the original motivation for the Matérn model came from its flexibility in context of spatial interpolation, there is now also a rich literature of alternative and enhanced versions of the model. In particular, the SPDE and related approaches enable one to define analogues of the Matérn model on quite general domains, admitting sparse approximations to precision matrices, while recent advances in enhanced models with compact support can facilitate scalable computation through sparse approximation of covariance matrices, and are well-suited to processes with short-scale dependence. The theoretical and empirical properties of these enhanced models have been recently and actively studied. On the other hand, there remain open theoretical issues of practical importance, such as parameter estimation at finite sample sizes, and the impact of parameter estimation on the performance of the associated predictions.

Our current understanding of the Matérn model has emerged as the result of engagement between scientists and practitioners from different disciplines, and our hope is that this multi-disciplinarity perspective will shine further light onto the Matérn model.

ACKNOWLEDGMENTS

We thank the Associate Editor and three anonymous Reviewers for their thorough reading and criticisms that allowed for an improved version of the manuscript. We are very grateful to Toni Karvonen for pointing out an important technicality about Sobolev spaces associated with the Matérn kernel. Moreno Bevilacqua acknowledges financial support from grant FONDECYT 1240308 and ANID/PIA/ANILLOS ACT210096 and ANID project Data Observatory Foundation DO210001 from the Chilean government and project MATH-AMSUD 22-MATH-06 (AMSUD220041).

REFERENCES

[1] Åberg, S. and Podgórski, K. (2011). A class of non-Gaussian second order random fields. *Extremes*, 14(2):187–222.

[2] Abramowitz, M. and Stegun, I. A., editors (1970). *Handbook of Mathematical Functions*. Dover, New York.

[3] Alegría, A., Cuevas-Pacheco, F., Diggle, P., and Porcu, E. (2021a). The F-family of covariance functions: A Matérn analogue for modeling random fields on spheres. *Spatial Statistics*, 43:100512.

[4] Alegría, A., Emery, X., and Porcu, E. (2021b). Bivariate Matérn covariances with cross-dimple for modeling coregionalized variables. *Spatial Statistics*, 41:100491.

[5] Allard, D., Clarotto, L., and Emery, X. (2022). Fully nonseparable Gneiting covariance functions for multivariate space-time data. *Spatial Statistics*, 52:100706.

[6] Allard, D., Senoussi, R., and Porcu, E. (2016). Anisotropy models for spatial data. *Mathematical Geosciences*, 48(3):305–328.

[7] Anderes, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *The Annals of Statistics*, 38(2):870–893.

[8] Anderes, E., Møller, J., and Rasmussen, J. G. (2020). Isotropic covariance functions on graphs and their edges. *The Annals of Statistics*, 48(4):2478–2503.

[9] Angulo, J., Kelbert, M. Y., Leonenko, N., and Ruiz-Medina, M. D. (2008). Spatiotemporal random fields associated with stochastic fractional Helmholtz and heat equations. *Stochastic Environmental Research and Risk Assessment*, 22(1):3–13.

[10] Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193.

[11] Askey, R. (1973). Radial characteristic functions. *Technical report, Research Center, University of Wisconsin*.

[12] Bachoc, F., Bevilacqua, M., and Velandia, D. (2019). Composite likelihood estimation for a Gaussian process under fixed domain asymptotics. *Journal of Multivariate Analysis*, 174:104534.

[13] Bachoc, F., Porcu, E., Bevilacqua, M., Furrer, R., and Faouzi, T. (2022). Asymptotically equivalent prediction in multivariate geostatistics. *Bernoulli*, 28(4):2518 – 2545.

[14] Bakka, H., Krainski, E., Bolin, D., Rue, H., and Lindgren, F. (2020). The diffusion-based extension of the Matérn field to space-time. *arXiv preprint arXiv:2006.04917*.

[15] Banerjee, S. and Gelfand, A. E. (2006). Bayesian wombling: Curvilinear gradient assessment under spatial process models. *Journal of the American Statistical Association*, 101(476):1487–1501.

[16] Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.

[17] Banerjee, S., Gelfand, A. E., and Sirmans, C. F. (2003). Directional rates of change under spatial process models. *Journal of the American Statistical Association*, 98(464):946–954.

[18] Barp, A., Oates, C. J., Porcu, E., and Girolami, M. (2022). A Riemann–Stein kernel method. *Bernoulli*, 28(4):2181–2208.

[19] Benzi, M. (2016). Localization in matrix computations: Theory and applications. In Benzi, M. and Simoncini, V., editors, *Exploiting Hidden Structure in Matrix Computations: Algorithms and Applications : Cetraro, Italy 2015*, pages 211–317. Springer International Publishing.

[20] Berg, C., Mateu, J., and Porcu, E. (2008). The Dagum family of isotropic correlation functions. *Bernoulli*, 14(4):1134–1149.

[21] Bevilacqua, M., Caamaño-Carrillo, C., and Porcu, E. (2022). Unifying compactly supported and Matérn covariance functions in spatial statistics. *Journal of Multivariate Analysis*, 189:104949.

[22] Bevilacqua, M., Caamaño-Carrillo, C., Arellano-Valle, R. B., and Morales-Oñate, V. (2021). Non-Gaussian geostatistical modeling using (skew) t processes. *Scandinavian Journal of Statistics*, 48(1):212–245.

[23] Bevilacqua, M., Faouzi, T., Furrer, R., and Porcu, E. (2019). Estimation and prediction using generalized Wendland functions under fixed domain asymptotics. *Annals of Statistics*, 47(2):828–856.

[24] 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.

[25] Bevilacqua, M., Morales-Oñate, V., and Caamaño-Carrillo, C. (2023). *GeoModels: Procedures for Gaussian and Non Gaussian Geostatistical (Large) Data Analysis*. R package version 1.1.5.

[26] Bingham, N. H., Goldie, C. M., and Teugels, J. (1987). Regular variation. *Encyclopedia of Mathematics and its Applications*, 27:145–156.

[27] Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. California Monographs in mathematical sciences. University of California Press.

[28] Bolin, D. (2014). Spatial Matérn fields driven by non-Gaussian noise. *Scandinavian Journal of Statistics*, 41(3):557–579.

[29] Bolin, D. and Kirchner, K. (2020). The rational SPDE approach for Gaussian random fields with general smoothness. *Journal of Computational and Graphical Statistics*, 29(2):274–285.

[30] Bolin, D. and Kirchner, K. (2022). Equivalence of measures and asymptotically optimal linear prediction for Gaussian random fields with fractional-order covariance operators. *arXiv preprint arXiv:2101.07860v2*.

[31] Bolin, D., Simas, A. B., and Wallin, J. (2022). Gaussian Whittle-Matérn fields on metric graphs.

[32] Bolin, D. and Wallin, J. (2016). Multivariate type G Matérn stochastic partial differential equation random fields. *arXiv preprint arXiv:1606.08298*.

[33] Borovitskiy, V., Azangulov, I., Terenin, A., Mostowsky, P., Deisenroth, M., and Durrande, N. (2021). Matérn Gaussian processes on graphs. In Banerjee, A. and Fukumizu, K., editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 2593–2601. PMLR.

[34] Borovitskiy, V., Terenin, A., Mostowsky, P., et al. (2020). Matérn Gaussian processes on Riemannian manifolds. *Advances in Neural Information Processing Systems*, 33:12426–12437.

[35] Brown, R. D., Bardsley, J. M., and Cui, T. (2020). Semivariogram methods for modeling Whittle–Matérn priors in Bayesian inverse problems. *Inverse Problems*, 36(5):055006.

[36] Buhmann, M. (2001). A new class of radial basis functions with compact support. *Mathematics of Computation*, 70(233):307–318.

[37] Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the spde approach. *AStA Advances in Statistical Analysis*, 97(2):109–131.

[38] Chernih, A. and Hubbert, S. (2014). Closed form representations and properties of the generalised Wendland functions. *Journal of Approximation Theory*, 177:17–33.

[39] Chilès, J. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York.

[40] Clarotto, L., Allard, D., Romary, T., and Desassis, N. (2022). The spde approach for spatio-temporal datasets with advection and diffusion: A matrix-free approach. In *15th International Conference of the ERCIM WG on Com-putational and Methodological Statistics (CMStatistics 2022)*.

[41] Cockayne, J., Oates, C. J., Sullivan, T. J., and Girolami, M. (2019). Bayesian probabilistic numerical methods. *SIAM review*, 61(4):756–789.

[42] Cressie, N. (1990). The origins of Kriging. *Mathematical geology*, 22(3):239–252.

[43] Cressie, N. and Johannesson, G. (2008). Fixed rank Kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

[44] Daley, D. J., Porcu, E., and Bevilacqua, M. (2015). Classes of compactly supported covariance functions for multivariate random fields. *Stochastic Environmental Research and Risk Assessment*, 29(4):1249–1263.

[45] Datta, A. (2022). Nearest-neighbor sparse Cholesky matrices in spatial statistics. *WIREs Computational Statistics*, 14(5):e1574.

[46] Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016a). Hierarchical nearest-neighbor Gaussian process models for large

geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.

[47] Datta, A., Banerjee, S., Finley, A. O., Hamm, N. A., and Schaap, M. (2016b). Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *The annals of applied statistics*, 10(3):1286.

[48] Davydov, O. and Schaback, R. (2019). Optimal stencils in Sobolev spaces. *IMA Journal of Numerical Analysis*, 39:398–422.

[49] Dey, D., Datta, A., and Banerjee, S. (2022). Graphical Gaussian process models for highly multivariate spatial data. *Biometrika*, 109(4):993–1014.

[50] Eidsvik, J., Shaby, B., Reich, B., Wheeler, M., and Niemi, J. (2013). Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, to appear.

[51] Emery, X. and Alegría, A. (2022). The Gauss hypergeometric covariance kernel for modeling second-order stationary random fields in euclidean spaces: its compact support, properties and spectral representation. *Stochastic Environmental Research and Risk Assessment*, 36:2819—-2834.

[52] Emery, X., Porcu, E., and White, P. (2022). New validity conditions for the multivariate Matérn Coregionalization model, with an application to exploration geochemistry. *Mathematical Geosciences*, pages 1–26.

[53] Faouzi, T., Porcu, E., and Bevilacqua, M. (2022). Space-time estimation and prediction under fixed-domain asymptotics with compactly supported covariance functions. *Statistica Sinica*, 32:1–17.

[54] Faouzi, T., Porcu, E., Bevilacqua, M., and Kondrashuk, I. (2020). Zastavnyi operators and positive definite radial functions. *Statistics & Probability Letters*, 157:108620.

[55] Fasshauer, G. (1997). Solving partial differential equations by collocation with radial basis functions. In LeMéhauté, A., Rabut, C., and Schumaker, L., editors, *Surface Fitting and Multiresolution Methods*, pages 131–138. Vanderbilt University Press, Nashville, TN.

[56] Fasshauer, G. and McCourt, M. (2015). *Kernel-based Approximation Methods using MATLAB*, volume 19 of *Interdisciplinary Mathematical Sciences*. World Scientific, Singapore.

[57] Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.

[58] Genton, M. G. and Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science*, 30(2):147–163.

[59] Gneiting, T. (2002a). Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83:493–508.

[60] Gneiting, T. (2002b). Nonseparable, Stationary Covariance Functions for Space-Time Data. *J. Am. Statist. Ass.*, 97:590–600.

[61] Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19(4):1327–1349.

[62] Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn Cross-Covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105:1167–1177.

[63] Gneiting, T. and Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review*, 46(2):269–282.

[64] Gneiting, T., Ševčíková, H., and Percival, D. B. (2012). Estimators of Fractal Dimension: Assessing the Roughness of Time Series and Spatial Data. *Statistical Science*, 27(2):247 – 277.

[65] Gough, B. (2009). *GNU scientific library reference manual*. Network Theory Ltd.

[66] Guinness, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics*, 60(4):415–429.

[67] Guinness, J. (2022). Inverses of Matérn covariances on grids. *Biometrika*, 109(2):535–541.

[68] Guinness, J. and Fuentes, M. (2016). Isotropic covariance functions on spheres: Some properties and modeling considerations. *Journal of Multivariate Analysis*, 143:143–152.

[69] Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika*, 93(4):989–995.

[70] Halder, A., Banerjee, S., and Dey, D. K. (2023). Bayesian modeling with spatial curvature processes. *Journal of the American Statistical Association*, ?(just-accepted):1–27.

[71] Harrar, S. W., Seneta, E., and Gupta, A. K. (2006). Duality between matrix variate t and matrix variate vg distributions. *Journal of multivariate analysis*, 97(6):1467–1475.

[72] Hartikainen, J. and Särkkä, S. (2010). Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In *2010 IEEE international workshop on machine learning for signal processing*, pages 379–384. IEEE.

[73] Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., and Zammit-Mangion, A. (2019). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Environ. Stat.*, 24:398–425.

[74] Hennig, P., Osborne, M. A., and Kersting, H. P. (2022). *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press.

[75] Hubbert, S. (2012). Closed form representations for a class of compactly supported radial basis functions. *Advances in Computational Mathematics*, 36(1):115–136.

[76] Ip, H. L. and Li, W. K. (2017). On some Matérn covariance functions for spatio-temporal random fields. *Statistica Sinica*, 27:805–822.

[77] Jansson, E., Kovács, M., and Lang, A. (2022). Surface finite element approximation of spherical Whittle–Matérn Gaussian random fields. *SIAM Journal on Scientific Computing*, 44(2):A825–A842.

[78] Jaquier, N., Borovitskiy, V., Smolensky, A., Terenin, A., Asfour, T., and Rozo, L. (2022). Geometry-aware Bayesian optimization in robotics using Riemannian Matérn kernels. In *Conference on Robot Learning*, pages 794–805. PMLR.

[79] Jensen, K., Kao, T.-C., Tripodi, M., and Hennequin, G. (2020). Manifold gplvms for discovering non-euclidean latent structure in neural data. *Advances in Neural Information Processing Systems*, 33:22580–22592.

[80] Karvonen, T. (2022). Asymptotic bounds for smoothness parameter estimates in Gaussian process regression. *arXiv preprint arXiv:2203.05400*.

[81] Karvonen, T. and Oates, C. J. (2022). Maximum likelihood estimation in Gaussian process regression is ill-posed. *Journal of Machine Learning Research*. To appear.

[82] Katzfuss, M. and Guinness, J. (2021). A General Framework for Vecchia Approximations of Gaussian Processes. *Statistical Science*, 36(1):124 – 141.

[83] Kaufman, C. G. and Shaby, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*, 100:473–484.

[84] Kelbert, M. Y., Leonenko, N. N., and Ruiz-Medina, M. (2005). Fractional random fields associated with stochastic fractional heat equations. *Advances in Applied Probability*, 37(1):108–133.

[85] Kent, J. T. (1989). Continuity properties for random fields. *The Annals of Probability*, pages 1432–1440.

[86] Khristenko, U., Scarabosio, L., Swierczynski, P., Ullmann, E., and Wohlmuth, B. (2019). Analysis of boundary effects on PDE-based sampling of Whittle-Matérn random fields. *SIAM/ASA Journal on Uncertainty Quantification*, 7(3):948–974.

[87] Kleiber, W. and Nychka, D. (2012). Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis*, 112:76–91.

[88] Kleiber, W. and Porcu, E. (2015). Nonstationary matrix covariances: compact support, long range dependence and quasi-arithmetic constructions. *Stochastic environmental research and risk assessment*, 29(1):193–204.

[89] Koepernik, P. and Pfaff, F. (2021). Consistency of Gaussian process regression in metric spaces. *J. Mach. Learn. Res.*, 22:244–1.

[90] Lang, A. and Pereira, M. (2021). Galerkin–Chebyshev approximation of Gaussian random fields on compact Riemannian manifolds. *arXiv preprint arXiv:2107.02667*.

[91] Lavin, A., Zenil, H., Paige, B., Krakauer, D., Gottschlich, J., Mattson, T., Anandkumar, A., Choudry, S., Rocki, K., Baydin, A. G., et al. (2021). Simulation intelligence: Towards a new generation of scientific methods. *arXiv preprint arXiv:2112.03235*.

[92] Leonenko, N. and Malyarenko, A. (2017). Matérn class tensor-valued random fields and beyond. *Journal of Statistical Physics*, 168(6):1276–1301.

[93] Leonenko, N., Malyarenko, A., and Olenko, A. (2022). On spectral theory of random fields in the ball. *Theory of Probability and Mathematical Statistics*, 107:61–76.

[94] Leonenko, N., Ruiz-Medina, M. D., and Taqqu, M. S. (2011). Fractional elliptic, hyperbolic and parabolic random fields. *Electronic Journal of Probability*, 16:1134–1172.

[95] Li, C. (2022). Bayesian fixed-domain asymptotics for covariance parameters in a gaussian process model. *The Annals of Statistics*, 50(6):3334–3363.

[96] Li, C., Sun, S., and Zhu, Y. (2023). Fixed-domain posterior contraction rates for spatial gaussian process model with nugget. *Journal of the American Statistical Association*, pages 1–25.

[97] Li, D., Jones, A., Banerjee, S., and Engelhardt, B. E. (2021a). Multi-group Gaussian processes. *arXiv preprint arXiv:2110.08411*.

[98] Li, D., Tang, W., and Banerjee, S. (2021b). Inference for Gaussian processes with Matérn covariogram on compact Riemannian manifolds. *arXiv preprint arXiv:2104.03529*.

[99] Lilly, J. M., Sykulski, A. M., Early, J. J., and Olhede, S. C. (2017). Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion. *Nonlinear Processes in Geophysics*, 24(3):481–514.

[100] Lim, S. and Teo, L. P. (2009). Generalized Whittle-Matérn random field as a model of correlated fluctuations. *Journal of Physics A: Mathematical and Theoretical*, 42(10):105202.

[101] Lindgren, F., Bolin, D., and Rue, H. (2022). The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running. *Spatial Statistics*, page 100599.

[102] Lindgren, F., Rue, H., and Lindstroem, J. (2011). An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: the Stochastic Partial Differential Equation Approach. *J. R. Statist. Soc. B*, 73:423–498.

[103] Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80:221–239.

[104] Lodhia, A., Sheffield, S., Sun, X., and Watson, S. S. (2016). Fractional Gaussian fields: A survey. *Probability Surveys*, 13(none):1 – 56.

[105] Loh, W.-L. and Sun, S. (2023). Estimating the parameters of some common gaussian random fields with nugget under fixed-domain asymptotics. *Bernoulli*, 29(3):2519–2543.

[106] Loh, W.-L., Sun, S., and Wen, J. (2021). On fixed-domain asymptotics, parameter estimation and isotropic Gaussian random fields with Matérn covariance functions. *The Annals of Statistics*, 49(6):3127 – 3152.

[107] Ma, P. and Bhadra, A. (2022). Beyond Matérn: on a class of interpretable confluent hypergeometric covariance functions. *Journal of the American Statistical Association*, pages 1–14.

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

[109] Matérn, B. (1986). *Spatial Variation: Stochastic Models and their Applications to Some Problems in Forest Surveys and Other Sampling Investigations*. Springer, Heidelberg, 2nd edition.

[110] Matheron, G. (1963). *Traité de Géostatistique Appliquée. Tome II: le Krigeage*. Mémoires du Bureau de Recherches Géologiques et Minières, no. 24. Editions BRGM, Paris.

[111] Matheron, G. (1965). *Les Variables Régionalisées et leur Estimation*. Masson, Paris.

[112] Matheron, G. (1971). *The Theory of Regionalized Variables and its Applications*. Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, France.

[113] Menegatto, V., Oliveira, C., and Porcu, E. (2020). Gneiting class, semi-metric spaces and isometric embeddings. *Constructive Mathematical Analysis*, 3(2):85–95.

[114] Morales-Navarrete, D., Bevilacqua, M., Caamaño-Carrillo, C., and Castro, L. (2022). Modelling point referenced spatial count data: a Poisson process approach. *Journal of the American Statistical Association*, 0(0):1–25.

[115] Narcowich, F. J., Ward, J. D., and Wendland, H. (2006). Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constructive Approximation*, 24(2):175–186.

[116] Nikitin, A. V., John, S., Solin, A., and Kaski, S. (2022). Non-separable spatio-temporal graph kernels via SPDEs. In *International Conference on Artificial Intelligence and Statistics*, pages 10640–10660. PMLR.

[117] Novak, E. and Woźniakowski, H. (2008). *Tractability of Multivariate Problems: Standard information for functionals*, volume 2. European Mathematical Society.

[118] Oates, C. J., Girolami, M., and Chopin, N. (2017). Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):695–718.

[119] Olver, F. W., Lozier, D. M., Boisvert, R. F., and Clark, C. W. (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press, Cambridge.

[120] Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics: The official journal of the International Environmetrics Society*, 17(5):483–506.

[121] Palacios, M. B. and Steel, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, 101:604–618.

[122] Pereira, M., Desassis, N., and Allard, D. (2022). Geostatistics for large datasets on Riemannian manifolds: a matrix-free approach. *arXiv preprint arXiv:2208.12501*.

[123] Porcu, E., Alegría, A., and Furrer, R. (2018). Modeling spatially global and temporally evolving data. *International Statistical Review*, 86:344–377.

[124] Porcu, E., Bevilacqua, M., and Genton, M. G. (2020a). Nonseparable, space-time covariance functions with dynamical compact supports. *Statistica Sinica*, 30:719–739.

[125] Porcu, E., Bevilacqua, M., Schaback, R., and Oates, C. (2024). Supplement to "the Matérn model: a journey through statistics, numerical analysis and machine learning". *Statistical Science*.

[126] Porcu, E., Daley, D. J., Buhmann, M., and Bevilacqua, M. (2013). Radial basis functions with compact support for multivariate geostatistics. *Stochastic environmental research and risk assessment*, 27(4):909–922.

[127] Porcu, E., Emery, X., and Mery, N. (2022a). Criteria and characterizations for spatially isotropic and temporally symmetric matrix-valued covariance functions. *Comp. Appl. Math.*, 41(223).

[128] Porcu, E., Furrer, R., and Nychka, D. (2021). 30 years of space–time covariance functions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(2):e1512.

[129] Porcu, E., White, P. A., and Genton, M. G. (2022b). Nonseparable space-time stationary covariance functions on networks cross time. *arXiv preprint arXiv:2208.03359*.

[130] Porcu, E., Zastavnyi, V., Bevilacqua, M., and Emery, X. (2020b). Stein hypothesis and screening effect for covariances with compact support. *Electronic Journal of Statistics*, 14(2):2510–2528.

[131] Porcu, E., Zastavyi, P., and Bevilacqua, M. (2016). Buhmann covariance functions, their compact supports, and their smoothness. *arXiv preprint arXiv:1606.09527*.

[132] Quick, H., Banerjee, S., and Carlin, B. P. (2013). Modeling temporal gradients in regionally aggregated California asthma hospitalization data. *The Annals of Applied Statistics*, 7(1):154.

[133] Roininen, L., Girolami, M., Lasanen, S., and Markkanen, M. (2019). Hyperpriors for Matérn fields with applications in Bayesian inversion. *Inverse Problems & Imaging*, 13(1).

[134] Sanz-Alonso, D. and Yang, R. (2022a). Finite element representations of Gaussian processes: Balancing numerical and statistical accuracy. *SIAM/ASA Journal on Uncertainty Quantification*, 10(4):1323–1349.

[135] Sanz-Alonso, D. and Yang, R. (2022b). The SPDE approach to Matérn fields: Graph representations. *Statistical Science*, 37(4):519–540.

[136] Sarkka, S., Solin, A., and Hartikainen, J. (2013). Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through kalman filtering. *IEEE Signal Processing Magazine*, 30(4):51–61.

[137] Schaback, R. (2011). The missing Wendland functions. *Advances in Computational Mathematics*, 34(1):67–81.

[138] Schaback, R. (2015). A computational tool for comparing all linear PDE solvers. *Advances of Computational Mathematics*, 41:333–355.

[139] Schaback, R. (2017). Error analysis of nodal meshless methods. In Griebel, M. and Schweitzer, M., editors, *Meshfree Methods for Partial Differential Equations VIII*, volume 115 of *Lecture Notes in Computational Science and Engineering*, pages 117–143. Springer.

[140] Schaback, R. and Wendland, H. (2006). Kernel techniques: from Machine Learning to Meshless Methods. *Acta Numerica*, 15:543–639.

[141] Scheuerer, M. (2010). Regularity of the sample paths of a general second order random field. *Stochastic Processes and their Applications*, 120(10):1879–1897.

[142] Scheuerer, M., Schaback, R., and Schlather, M. (2013). Interpolation of spatial data–a stochastic or a deterministic problem? *European Journal of Applied Mathematics*, 24(4):601–629.

[143] Schoenberg, I.-J. (1938). Metric spaces and completely monotone functions. *Annals of Mathematics*, 39(4):811–841.

[144] Schoenberg, I. J. (1942). Positive Definite Functions on Spheres. *Duke Math. Journal*, 9:96–108.

[145] Shaby, B. and Ruppert, D. (2012). Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, 21:433–452.

[146] South, L. F., Karvonen, T., Nemeth, C., Girolami, M., and Oates, C. J. (2022). Semi-exact control functionals from Sard's method. *Biometrika*, 109(2):351–367.

[147] Stein, M. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *The Annals of Statistics*, 16:55–63.

[148] Stein, M. (2005). Statistical methods for regular monitoring data. *Journal of the Royal Statistical Society B*, 67:667–687.

[149] Stein, M. L. (1999). *Interpolation of Spatial Data. Some Theory of Kriging*. Springer, New York.

[150] Stein, M. L. (2002). The screening effect in Kriging. *The Annals of Statistics*, 30(1):298–323.

[151] Stein, M. L. (2004). Equivalence of Gaussian measures for some nonstationary random fields. *Journal of Statistical Planning and Inference*, 123(1):1–11.

[152] Stein, M. L. (2011). 2010 Rietz lecture: When does the screening effect hold? *The Annals of Statistics*, 39(6):2795–2819.

[153] Stein, M. L. (2015). When does the screening effect not hold? *Spatial Statistics*, 11:65–80.

[154] Stein, M. L., Chi, Z., and Welty, L. J. (2004). Approximating Likelihoods for Large Spatial Data Sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):275–296.

[155] Sun, Y., Li, B., and Genton, M. G. (2012). Geostatistics for large datasets. In *Advances and challenges in space-time modelling of natural events*, pages 55–77. Springer.

[156] Tang, J. and Zimmerman, D. (2020). Space-time covariance models on networks with an application on streams. *arXiv preprint arXiv:2009.14745*.

[157] Tang, W., Zhang, L., and Banerjee, S. (2021). On identifiability and consistency of the nugget in Gaussian spatial process models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(5):1044–1070.

[158] Terdik, G. (2015). Angular spectra for non-Gaussian isotropic fields. *Brazilian Journal of Probability and Statistics*, 29(4):833–865.

[159] Terdik, G. (2022). Spatiotemporal covariance functions for Laplacian ARMA fields in higher dimensions. *Theory of Probability and Mathematical Statistics*, 107:111–132.

[160] Tuo, R. and Wang, W. (2020). Kriging prediction with isotropic Matérn correlations: Robustness and experimental designs. *J. Mach. Learn. Res.*, 21:187–1.

[161] Tuo, R., Wang, Y., and Jeff Wu, C. (2020). On the improved rates of convergence for Matérn-type kernel ridge regression with application to calibration of computer models. *SIAM/ASA Journal on Uncertainty Quantification*, 8(4):1522–1547.

[162] Vakili, S., Moss, H., Artemev, A., Dutordoir, V., and Picheny, V. (2021). Scalable Thompson sampling using sparse Gaussian process models. *Advances in Neural Information Processing Systems*, 34:5631–5643.

[163] Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42.

[164] Vecchia, A. V. (1988). Estimation and Model Identification for Continuous Spatial Processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 297–312.

[165] Wallin, J. and Bolin, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scandinavian Journal of Statistics*, 42(3):872–890.

[166] Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396.

[167] Wendland, H. (2004). *Scattered data approximation*, volume 17. Cambridge university press.

[168] Whittle, P. (1953). The Analysis of Multiple Stationary Time Series. *J. R. Statist. Soc. B*, 15(1):125–139.

[169] Whittle, P. (1963). Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute*, 40(2):974–994.

[170] Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.

[171] Wilson, A. G., Hu, Z., Salakhutdinov, R., and Xing, E. P. (2016). Deep kernel learning. In *Artificial intelligence and statistics*, pages 370–378. PMLR.

[172] Wilson, J. T., Borovitskiy, V., Terenin, A., Mostowsky, P., and Deisenroth, M. P. (2021). Pathwise conditioning of Gaussian processes. *J. Mach. Learn. Res.*, 22:105–1.

[173] Wu, Z. (1992). Hermite–Birkhoff interpolation of scattered data by radial basis functions. *Approximation Theory and its Applications*, 8/2:1–10.

[174] Xua, G. and Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of the American Statistical Association*, 112:1236 –1249.

[175] Yan, Y. and Genton, M. G. (2018). Gaussian likelihood inference on data from trans-Gaussian random fields with Matérn covariance function. *Environmetrics*, 29(5-6):e2458.

[176] Zastavnyi, V. (2002). Positive-definite radial functions and splines. *Doklady Mathematics*, 66:213–216.

[177] Zastavnyi, V. (2006). On some properties of Buhmann functions. *Ukrainian Mathematical Journal*, 58(8).

[178] Zastavnyi, V. and Porcu, E. (2017). On positive definiteness of some radial functions. *Lobachevskii Journal of Mathematics*, 38(2):386–394.

[179] Zastavnyi, V. P. and Porcu, E. (2011). Characterization Theorems for the Gneiting Class of Space–Time Covariance Functions. *Bernoulli*, 17(1):456–465.

[180] Zhang, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261.

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