Bayesian Analysis (2024) **TBA**, Number TBA, pp. 1–25

Default Priors for the Smoothness Parameter in Gaussian Matérn Random Fields

Zifei Han* and Victor De Oliveira†

Abstract. The Matérn family of covariance functions plays a prominent role in the analysis of geostatistical data due to its ability to model different smoothness behaviors. Although in many applications the smoothness parameter is set at an arbitrary value, a more satisfactory approach requires data–based inference about smoothness, especially in the light of new findings showing that the information this type of data has about the smoothness can be considerable in some settings. This work proposes a new class of easy–to–compute default priors for the parameters of a class of Gaussian random fields with Matérn covariance functions with unknown smoothness parameter. This class of priors is obtained by approximating a reference prior using the spectral representation of stationary random fields. This approximate reference prior has several advantages over the exact reference prior. First, the computation of the former is more stable and considerably less burdensome than that of the latter. Second, both the marginal prior of the smoothness parameter and the joint posterior of all parameters are proper for the approximate reference prior, while the status of these for the exact reference prior is currently unknown. Third, Bayesian inferences about the covariance parameters based on the approximate reference prior have satisfactory frequentist properties that are superior than those based on maximum likelihood. It was also found that a previously proposed ad–hoc prior for the smoothness and the approximate reference prior display in most cases similar statistical performance, with the former being computationally simpler. The methodology is illustrated with an analysis of rainfall totals in Switzerland.

MSC2020 subject classifications: Primary 62F15, 62F10; secondary 62P12.

Keywords: Geostatistics, Reference prior, Spectral approximation.

1 Introduction

Random fields are useful probabilistic tools for modeling and analyzing geostatistical data that are routinely collected in the natural and earth sciences. In particular, Gaussian random fields play a prominent role due to their versatility for modeling spatially varying phenomena, and because they serve as building blocks for the construction of some non–Gaussian random fields (De Oliveira et al., 1997; Diggle and Ribeiro, 2007; Han and De Oliveira, 2016). The Matérn family of isotropic covariance functions (Matérn, 1986; Stein, 1999) is currently the most commonly used in applications due to its ability to describe different smoothness behaviors.

*School of Statistics, University of International Business and Economics, China, zifeihan@uibe.edu.cn

†(Corresponding author) Department of Management Science and Statistics, The University of Texas at San Antonio, USA, victor.deoliveira@utsa.edu

 https://doi.org/10.1214/24-BA1431

It is common in practice to fix the smoothness of the random field at an arbitrary value, even though knowledge about this feature is seldom available. For instance, the smoothness parameter of the Matérn covariance function is often fixed at 0.5 or 1.5. This practice is due to two reasons. The first involves computational challenges in the estimation of all covariance parameters and the lack of a closed–form expression for the derivative of the covariance function w.r.t. the smoothness parameter. For this, De Oliveira and Han (2022) and Geoga et al. (2023) recently developed computational tools to evaluate this derivative. The second reason involves unqualified claims in the applied statistical literature stating that spatial data have little or no information about the smoothness of the random field. De Oliveira and Han (2022) argued that this claim is not true in general. It was shown that the information the data have about the smoothness parameter varies depending on the true model and sampling design, and this information can be substantial in some settings. These findings support the viability of data–based inference about the smoothness of Gaussian Matérn random fields.

For the estimation of smoothness in random fields, Wu et al. (2013), Wu and Lim (2016), Im et al. (2007) and Anderes and Stein (2008) proposed semiparametric spectral methods, and Loh (2015) and Loh et al. (2021) proposed parametric space methods. These methods are rarely used in practice and none of them is implemented in common software used to fit geostatistical models. On the other hand, maximum likelihood estimation (MLE) is used often and is implemented in several public software, e.g., in the `R` packages `geoR` and `georob` (Ribeiro and Diggle, 2024; Papritz, 2024). But the sampling properties of the MLE of the smoothness parameter remain largely unexplored. To investigate some of these, a simulation study is carried out in Section 3. It is found that the MLE of the smoothness parameter of the Matérn family has poor sampling properties, as it is severely upward biased in situations of practical relevance. It is conjectured that it might not even exist for some data sets.

When the main goal of the data analysis is spatial interpolation, the Bayesian approach offers advantages over the frequentist plug–in approach, since the former accounts for parameter uncertainty while the latter does not. A challenge to implementing the Bayesian approach is the specification of sensible prior distributions for covariance parameters. Subjective elicitation of these priors in spatial models is challenging due to the lack of subjective information and the difficulty in interpreting some parameters. Early works specified priors for variance and range parameters in an ad–hoc manner (Kitanidis, 1986; Handcock and Stein, 1993; De Oliveira et al., 1997), but these may yield unwanted results, including improper posteriors. Two ad–hoc priors for the smoothness parameter $\nu > 0$ of the Matérn family have been proposed, but seldom used. Handcock and Stein (1993) suggested using $\pi(\nu) = (1+\nu)^{-2}$, while others suggested $\nu \sim \text{unif}(0, L)$ for some L large. It will be shown that the first is a good option (see below), while the second seemingly non–informative prior can be very informative and misleading. The limit of the Matérn covariance function in Section 2 as $\nu \to \infty$ is the square exponential covariance function. The specification $\nu \sim \text{unif}(0, 100)$ (say) implies that the squared exponential covariance function has a prior probability of about 0.9, an inappropriate belief in most applications due to the extreme smoothness of this model. We show that both ad–hoc priors yield proper posterior distributions.

A sound alternative to subjective and ad–hoc prior specifications consists of using information–based default priors, and among these reference priors have been the most studied. By default priors we mean priors that, beyond the sampling model, require little or no input from the user. Berger et al. (2001) provided an extensive discussion of foundational issues involving the formulation of default priors and advocated for the use of reference priors in geostatistical models. They showed that these priors overcome drawbacks of previously proposed priors, e.g., reference posteriors were shown to be proper. Extensions of the methodology in Berger et al. (2001) were developed by De Oliveira (2007), Kazianka and Pilz (2012), Ren et al. (2012) and Kazianka (2013) for isotropic covariance models, and by Paulo (2005), Ren et al. (2013) and Gu et al. (2018) for separable covariance models. But all of these works assumed the smoothness of the random field is *known*.

In this work we derive a new class of easy–to–compute default priors for the smoothness parameter in Gaussian random fields with Matérn covariance functions. These priors approximate reference priors using the spectral approximation to stationary random fields. They depend on an auxiliary regular design that can be tuned to improve the approximation. Bayesian analyses based on these approximate reference priors are considerably simpler, and their computation less onerous and more stable than those based on exact reference priors. For models with a constant mean function, the simplifications are even more substantial as the resulting approximate reference prior has a matrix–free expression. In addition, the approximate marginal reference prior of the smoothness parameter is proper, and the joint approximate reference posterior of all parameters is proper as well. De Oliveira and Han (2023) used similar techniques to derive approximate reference priors for range parameters.

Numerical experiments in Section 7 show that the proposed approximation to the reference prior is satisfactory for a variety of designs and models, and inferences based on the exact and approximate reference priors are practicably equivalent. But inferences based on the latter are considerably less onerous than those based on the former. It is also shown that Bayesian inferences on the covariance parameters based on approximate references prior and the ad–hoc prior suggested by Handcock and Stein (1993) have similar satisfactory frequentist properties, while Bayesian inferences based on the uniform prior are inadequate. A data set of daily rainfall totals collected in Switzerland is used to illustrate the proposed default Bayesian analysis.

2 Data and Model

Spatial data of geostatistical type consist of triplets $\{(\mathbf{s}_i, \boldsymbol{f}(\mathbf{s}_i), z_i) : i = 1, \ldots, n\}$, where $\mathcal{S}_n = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ is a set of sampling locations in the region of interest $\mathcal{D}$, called the *sampling design*, $\boldsymbol{f}(\mathbf{s}_i) = (f_1(\mathbf{s}_i), \ldots, f_p(\mathbf{s}_i))^\top$ is a p–dimensional vector of covariates measured at $\mathbf{s}_i$ (usually $f_1(\mathbf{s}) \equiv 1$), and z_i is the measurement of the quantity of interest collected at $\mathbf{s}_i$. The stochastic approach to spatial interpolation/prediction relies on viewing the set of measurements $\{z_i\}_{i=1}^n$ as a partial realization of a random field $Z(\cdot)$.

Let $\{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ be a Gaussian random field with mean function $\mu(\mathbf{s})$ and covariance function $C(\mathbf{s}, \mathbf{u})$, with $\mathcal{D} \subset \mathbb{R}^d$ and $d \geq 1$. It would be assumed that

$\mu(\mathbf{s}) = \sum_{j=1}^{p} \beta_j f_j(\mathbf{s})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ are unknown regression parameters. Additionally, $C(\mathbf{s}, \mathbf{u})$ is assumed isotropic and belonging to a parametric family, $\left\{C_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{u}) = \sigma^2 K_{\boldsymbol{\vartheta}}(||\mathbf{s} - \mathbf{u}||) : \boldsymbol{\theta} = (\sigma^2, \boldsymbol{\vartheta}) \in (0, \infty) \times \Theta\right\}$, $\Theta \subset \mathbb{R}^q$, where $K_{\boldsymbol{\vartheta}}(\cdot)$ is an isotropic correlation function in $\mathbb{R}^d$ and $\|\cdot\|$ is the Euclidean norm. Among the many possible isotropic covariance families, we focus in this work on the Matérn family with the parametrization proposed by Handcock and Stein (1993)

$$\begin{aligned} C_{\boldsymbol{\theta}}(r) &= \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}}{\vartheta} r\right)^{\nu} \mathcal{K}_\nu \left(\frac{2\sqrt{\nu}}{\vartheta} r\right), \qquad r \geq 0 \\ &=: \sigma^2 K_{\boldsymbol{\vartheta}}(r), \end{aligned} \tag{2.1}$$

where $r = ||\mathbf{s} - \mathbf{u}||$ is Euclidean distance between two locations, $\sigma^2 > 0$, $\boldsymbol{\vartheta} = (\vartheta, \nu) \in (0, \infty)^2$ are correlation parameters, $\Gamma(\cdot)$ is the gamma function and $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of second kind and order ν (Gradshteyn and Ryzhik, 2000, 8.40). For this family, $\sigma^2 = \text{var}(Z(\mathbf{s}))$, ϑ (mostly) controls how fast $C_{\boldsymbol{\theta}}(r)$ goes to zero when r increases, and ν controls the degree of differentiability of $C_{\boldsymbol{\theta}}(r)$ at the origin. If $\nu > k$, then $C_{\boldsymbol{\theta}}(\cdot)$ is $2k$ times differentiable at $r = 0$. Because of these properties, σ^2 is called the *variance* parameter, ϑ the *range* parameter and ν the *smoothness* parameter.

In applications the measurements z_i often contain measurement error, in which case they are modeled as

$$z_i = Z(\mathbf{s}_i) + \epsilon_i, \qquad i = 1, \ldots, n, \tag{2.2}$$

where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with $\text{N}(0, \sigma^2\xi)$ distribution and independent of $Z(\cdot)$; $\xi \geq 0$ is the so–called *noise–to–signal variance ratio*. Then, the covariance structure of the data is indexed by $(\sigma^2, \xi, \vartheta, \nu)$. Most applications in the literature assume ν is known and fixed at an arbitrary value, but in this work we assume ν is unknown.

3 MLE of Smoothness: A Simulation Exploration

As mentioned in the Introduction, the sampling properties of the maximum likelihood estimator of the smoothness parameter of the Matérn family are unknown, so in this section we carry out a simulation study to explore them. For $\mathcal{D} = [0, 1]^2$, we consider two sampling designs of size $n = 225$, a regular design where the sampling locations form a 15×15 equally spaced grid, and an irregular design where the sampling locations are an i.i.d. sample from the $\text{unif}((0, 1)^2)$ distribution. We consider data following model (2.2), with $p = 1$, $\mu(\mathbf{s}) = 0$ and Matérn covariance function (2.1) with $\sigma^2 = 1$. We assume that ϑ is 0.1 or 0.5 and ν is 0.5 or 1.5. For each combination of the above sampling designs and models, we simulate 1000 independent data sets under the following three scenarios for the measurement error:

I: $\xi = 0$ known II: $\xi = 0.2$ known III: $\xi = 0.2$ unknown.

The parameters to be estimated are $(\mu, \sigma^2, \vartheta, \nu)$ for scenarios I and II, and $(\mu, \sigma^2, \xi, \vartheta, \nu)$ for scenario III. The computation of MLEs was carried out using the `optim` function in `R` with the `L-BFGS-B` algorithm. We required that $\nu \in (0, 50)$ in the search space of the optimization algorithm, while the other parameters were unrestricted. For all designs,

		$\nu = 0.5$		$\nu = 1.5$	
		$\vartheta = 0.1$	$\vartheta = 0.5$	$\vartheta = 0.1$	$\vartheta = 0.5$
Scenario I	Median	0.597	0.541	1.686	1.550
	Mean	1.752	0.561	2.914	1.562
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	1.6	0	1.1	0
Scenario II	Median	0.622	0.530	1.852	1.615
	Mean	3.111	0.972	6.208	6.508
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	4.1	0.6	6.8	8.8
Scenario III	Median	6.239	0.961	7.846	2.359
	Mean	23.210	7.352	23.220	12.065
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	42.8	11.6	40.5	18.5

Table 1: Sampling features of $\hat{\nu}_{\text{MLE}}$ for each combination of model and scenario, estimated from 1000 simulated data sets on the regular design.

		$\nu = 0.5$		$\nu = 1.5$	
		$\vartheta = 0.1$	$\vartheta = 0.5$	$\vartheta = 0.1$	$\vartheta = 0.5$
Scenario I	Median	0.528	0.524	1.572	1.527
	Mean	0.571	0.549	1.654	1.539
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	0	0	0	0
Scenario II	Median	0.555	0.505	1.653	1.727
	Mean	0.778	0.774	4.883	7.168
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	0.3	0.3	4.6	10.1
Scenario III	Median	0.749	0.704	1.939	2.476
	Mean	6.382	3.384	9.252	13.644
	% of $\hat{\nu}_{\text{MLE}} \geq 50$	10.2	4.6	12.8	21.9

Table 2: Sampling features of $\hat{\nu}_{\text{MLE}}$ for each combination of model and scenario, estimated from 1000 simulated data sets on the irregular design.

models and scenarios the sampling distribution of the MLE of ν, $\hat{\nu}_{\text{MLE}}$, was asymmetric with a heavy right tail, so the sampling features to be estimated are median($\hat{\nu}_{\text{MLE}}$), $E(\hat{\nu}_{\text{MLE}})$ and $P(\hat{\nu}_{\text{MLE}} \geq 50)$.

Table 1 reports the estimated sampling features of $\hat{\nu}_{\text{MLE}}$ from the data simulated on the regular design. The results show that for all models and scenarios $\hat{\nu}_{\text{MLE}}$ is both mean and median *biased*, as $\hat{\nu}_{\text{MLE}}$ tends to overestimate ν. The magnitudes of these biases are small when there is no measurement error (scenario I), but are large when the data contain measurement error (scenarios II and III). Moreover, when the parameter ξ is also estimated (scenario III), these biases are extremely large, to the point of making $\hat{\nu}_{\text{MLE}}$ quite unreliable, especially when the strength of correlation is weak ($\vartheta = 0.1$). It also holds for scenario III that $P(\hat{\nu}_{\text{MLE}} \geq 50)$ is large, suggesting that in this case the MLE of ν could not only be extremely large, but might not even exist for some data sets (see below).

Table 2 reports the estimated sampling features of $\hat{\nu}_{\text{MLE}}$ from the data simulated on the irregular design. These features follow similar patterns as those in Table 1, but are

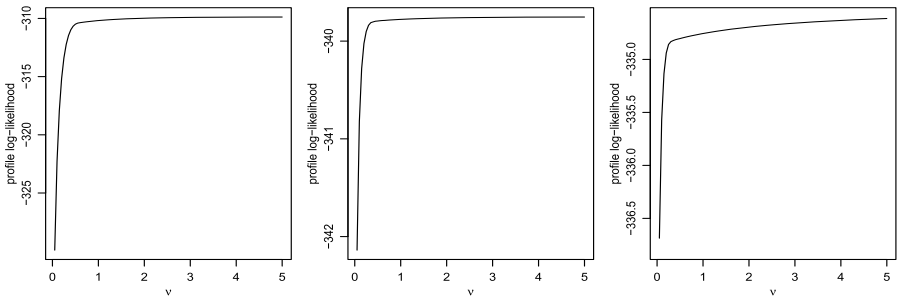

Figure 1: Profile log–likelihood functions of ν (up to an additive constant) for three simulated data sets with $(\mu, \sigma^2, \xi, \vartheta, \nu) = (0, 1, 0.2, 0.1, 0.5)$ and the regular design.

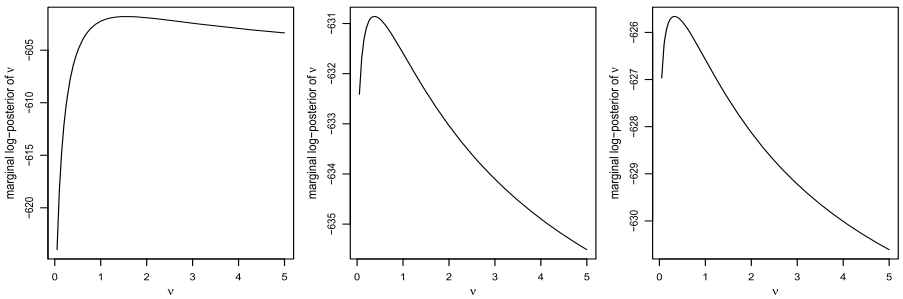

Figure 2: Log–marginal posterior densities of ν (up to an additive constant) based on the approximate reference prior for the three simulated data sets in Figure 1.

attenuated. Now $\hat{\nu}_{\text{MLE}}$ is close to being mean and median unbiased in scenario I, and $\hat{\nu}_{\text{MLE}}$ also tends to overestimate ν in scenarios II and III, although to a lesser extent. The biases and $P(\hat{\nu}_{\text{MLE}} \geq 50)$ are still substantial when ξ is estimated. The main take home messages from Tables 1 and 2 are that $\hat{\nu}_{\text{MLE}}$ tends to severely overestimate the smoothness when the data contain measurement error, especially when its variance is estimated, and this behavior is more severe for regular sampling designs.

To dwell more into the sampling behaviors of $\hat{\nu}_{\text{MLE}}$, Figure 1 displays plots of profile log–likelihoods of ν corresponding to three data sets simulated on the regular design with $\vartheta = 0.1$, $\nu = 0.5$ and scenario III, for which the iterative algorithm indicated that $\hat{\nu}_{\text{MLE}} \geq 50$. For each of these the algorithm stopped because the maximum number of allowed iterations (100) was reached, suggesting convergence did not occur. Inspection of these profile log–likelihoods for values of ν larger than 5 suggest that they are monotonically increasing, so $\hat{\nu}_{\text{MLE}}$ might not exist for these data sets.

A possible fix to the above undesirable behavior is using an estimation method that penalizes large values of ν, such as penalized maximum likelihood estimation or Bayesian estimation. Figure 2 displays the log–marginal posteriors of ν based on the approximate reference prior to be developed in this work, for the three data sets that were singled out in Figure 1. In stark contrast, the graphs in Figure 2 are unimodal

with well identified maxima, so the Bayesian maximum a posteriori estimates exist and provide sensible answers that are not far from the true value ($\nu = 0.5$).

4 Exact Reference Prior

In this section we describe the derivation of an exact reference prior for the parameters of a sub–class of Gaussian Matérn random fields described in Section 2. We assume the range ϑ and noise–to–signal variance ratio ξ are known, so from now on $q = 1$, $\boldsymbol{\vartheta} = \nu$, $\Theta = (0, \infty)$, and $\boldsymbol{\theta} = (\sigma^2, \nu)$ are the unknown covariance parameters.

For deriving reference priors of spatial models, the covariance parameters $\boldsymbol{\theta}$ are typically considered of primary interest, and the regression parameters $\boldsymbol{\beta}$ of secondary interest, so the reference prior is factored accordingly as $\pi^{\mathrm{R}}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \pi^{\mathrm{R}}(\boldsymbol{\beta} \,|\, \boldsymbol{\theta})\pi^{\mathrm{R}}(\boldsymbol{\theta})$. For the model under study, the conditional Jeffreys prior of the secondary parameters given the primary parameters is $\pi^{\mathrm{R}}(\boldsymbol{\beta} \mid \boldsymbol{\theta}) \propto 1$, while $\pi^{\mathrm{R}}(\boldsymbol{\theta})$ is computed using the Jeffreys rule based on the 'marginal model' defined via the integrated likelihood of $\boldsymbol{\theta}$

$$\begin{aligned} L^{\mathrm{I}}(\boldsymbol{\theta}; \boldsymbol{z}) &= \int_{\mathbb{R}^p} L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{z}) \pi^{\mathrm{R}}(\boldsymbol{\beta} \,|\, \boldsymbol{\theta}) d\boldsymbol{\beta} \\ &\propto (\sigma^2)^{-\frac{n-p}{2}} |\boldsymbol{\Psi}_\nu|^{-\frac{1}{2}} |\boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X}|^{-\frac{1}{2}} \exp\left\{ -\frac{S_\nu^2}{2\sigma^2} \right\}, \end{aligned} \tag{4.1}$$

where $L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{z})$ is the Gaussian likelihood of all model parameters based on the data $\boldsymbol{z} = (z_1, \ldots, z_n)^\top$, $S_\nu^2 = (\boldsymbol{z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_\nu)^\top \boldsymbol{\Psi}_\nu^{-1} (\boldsymbol{z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_\nu)$, $\hat{\boldsymbol{\beta}}_\nu = (\boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{z}$, $\boldsymbol{X}$ is the known $n \times p$ design matrix with entries $\boldsymbol{X}_{ij} = f_j(\mathbf{s}_i)$, and

$$\boldsymbol{\Psi}_\nu = \boldsymbol{\Sigma}_\nu + \xi \boldsymbol{I}_n,$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, $\boldsymbol{\Sigma}_\nu$ is the $n \times n$ matrix with entries $(\boldsymbol{\Sigma}_\nu)_{ij} = K_\nu(\|\mathbf{s}_i - \mathbf{s}_j\|)$ and $K_\nu(\cdot)$ is the Matérn correlation function defined in (2.1).

Proposition 4.1 (Reference Prior)**.** *The reference prior of* $(\boldsymbol{\beta}, \sigma^2, \nu)$ *is given by*

$$\pi^{\mathrm{R}}(\boldsymbol{\beta}, \sigma^2, \nu) \propto \frac{\pi^{\mathrm{R}}(\nu)}{\sigma^2}, \tag{4.2}$$

with

$$\pi^{\mathrm{R}}(\nu) \propto \left\{ \mathrm{tr}\left[\left\{ \left(\frac{\partial}{\partial \nu} \boldsymbol{\Sigma}_\nu \right) \boldsymbol{Q}_\nu \right\}^2 \right] - \frac{1}{n-p} \left[\mathrm{tr}\left\{ \left(\frac{\partial}{\partial \nu} \boldsymbol{\Sigma}_\nu \right) \boldsymbol{Q}_\nu \right\} \right]^2 \right\}^{\frac{1}{2}}, \tag{4.3}$$

where $\boldsymbol{Q}_\nu := \boldsymbol{\Psi}_\nu^{-1} - \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1}$ *and* $\frac{\partial}{\partial \nu} \boldsymbol{\Sigma}_\nu$ *is the entry–wise derivative of* $\boldsymbol{\Sigma}_\nu$ *w.r.t.* ν.

The expression (4.3) is similar to the one derived in Berger et al. (2001, Theorem 2) for range parameters. For the Matérn model, evaluation of (4.3) requires computing derivatives of the Bessel function $\mathcal{K}_\nu(x)$, both w.r.t. x and ν, where the latter lacks a

closed–form expression valid for any $\nu > 0$. De Oliveira and Han (2022) provide methods to compute these derivatives.

For propriety of the posterior distribution derived from the reference prior (4.2), it is required that the integral

$$\int_{\mathbb{R}^p \times (0,\infty)^2} L(\boldsymbol{\beta}, \sigma^2, \nu; \boldsymbol{z}) \frac{\pi^{\mathrm{R}}(\nu)}{\sigma^2} d\boldsymbol{\beta} d\sigma^2 d\nu \;=\; \int_{(0,\infty)} L^{\mathrm{I}}(\nu; \boldsymbol{z}) \pi^{\mathrm{R}}(\nu) d\nu, \tag{4.4}$$

is finite, where $L^{\mathrm{I}}(\nu; \boldsymbol{z})$ is the so–called *integrated likelihood* of ν obtained by integrating the product of the likelihood and $1/\sigma^2$ $(= \pi^{\mathrm{R}}(\sigma^2 \,|\, \nu))$ over $\boldsymbol{\beta}$ and σ^2.

Proposition 4.2 (Integrated Likelihood)**.** *The integrated likelihood of ν is given by*

$$L^{\mathrm{I}}(\nu; \boldsymbol{z}) \propto |\boldsymbol{\Psi}_\nu|^{-\frac{1}{2}} |\boldsymbol{X}^\top \boldsymbol{\Psi}_\nu^{-1} \boldsymbol{X}|^{-\frac{1}{2}} (S_\nu^2)^{-\frac{n-p}{2}}. \tag{4.5}$$

The above expression follows by direct calculation (Berger et al., 2001). The asymptotic behaviors of the integrated likelihood (4.5) follow from the following result.

Lemma 1. *Consider the Matérn family of correlation functions in (2.1). For any $\vartheta > 0$ and $r \geq 0$ fixed, it holds that*

(a) $\lim_{\nu \to 0^+} K_\nu(r) = \mathbf{1}\{r = 0\}$ *($\mathbf{1}\{A\}$ denotes the indicator function of set A).*

(b) $\lim_{\nu \to \infty} K_\nu(r) = \exp\big(-(r/\vartheta)^2\big)$.

Proof. The proof is given in Appendix A of the Supplementary Material (Han and De Oliveira, 2024).

The results in the above lemma imply that $L^{\mathrm{I}}(\nu; \boldsymbol{z})$ is bounded on $(0, \infty)$ for any $\boldsymbol{z}$ (see the proof of Theorem 6.2(b) in Section 6), and hence the finiteness of (4.4) depends entirely on the integrability of $\pi^{\mathrm{R}}(\nu)$ over $(0, \infty)$. Unfortunately, determining the asymptotic behavior of the reference prior of ν in (4.3) is quite challenging. There is no closed–form expression for $(\partial/\partial\nu)\boldsymbol{\Sigma}_\nu$ nor convenient approximations, and the asymptotic representation of correlations matrices in Berger et al. (2001) does not hold for smoothness parameters. As a result, the propriety status of $\pi^{\mathrm{R}}(\nu)$ and $\pi^{\mathrm{R}}(\boldsymbol{\beta}, \sigma^2, \nu \mid \boldsymbol{z})$ are currently unknown. Additionally, the computation of $\pi^{\mathrm{R}}(\nu)$ is quite demanding, even for data sets of moderate size. First, its evaluation involves the computation of the $n \times n$ matrix $\boldsymbol{\Psi}_\nu^{-1}$ which requires $O(n^3)$ operations. Second, for many families of correlation functions, including the Matérn, the matrix $\boldsymbol{\Psi}_\nu$ is often nearly singular when either ϑ or ν is large and ξ is small, so the computation of $\boldsymbol{\Psi}_\nu^{-1}$ will be unstable or infeasible in these situations. It also follows from the above lemma that the prior proposed by Handcock and Stein (1993) yields a proper posterior distribution.

In Section 6 we circumvent these theoretical and computational challenges by deriving an approximate reference prior that is more amenable for analysis and computation. This approximation depends neither on $K_\nu(\cdot)$ nor on the inverse of large and possibly numerically singular matrices, but instead on the spectral density function of the model.

5 Spectral Approximation to the Integrated Likelihood

The spectral approximation to stationary random fields has been used for different purposes by Royle and Wikle (2005), Paciorek (2007) and Bose et al. (2018), but unlike these works this device is employed here to approximate the random field over a set of locations that may not be the sampling design. Below we summarize the spectral approximation for random fields in the plane ($d = 2$); a detailed statement of the result and its proof are given by De Oliveira and Han (2023).

Let $\mu(\mathbf{s})$ and $\sigma^2 K_\nu(r)$ be, respectively, the mean and covariance functions of the random field $Z(\cdot)$, and $\sigma^2 f_\nu(\boldsymbol{\omega})$ be its spectral density function. For the Matérn family in (2.1) we have (Stein, 1999)

$$f_\nu(\boldsymbol{\omega}) = \frac{\Gamma(\nu+1)(4\nu)^\nu}{\pi\Gamma(\nu)\vartheta^{2\nu}}\left(\|\boldsymbol{\omega}\|^2 + \frac{4\nu}{\vartheta^2}\right)^{-(\nu+1)}, \quad \boldsymbol{\omega} = (\omega_1, \omega_2)^\top \in \mathbb{R}^2. \tag{5.1}$$

For M_1, M_2 positive even integers and $\Delta > 0$, let $\mathcal{U}_M = \{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \ldots, \mathbf{u}_{M_1,M_2}\} = \{\Delta, \ldots, \Delta M_1\} \times \{\Delta, \ldots, \Delta M_2\}$ be a set of spatial locations forming a regular rectangular grid in the plane, with $M := M_1 M_2$. The set $\mathcal{U}_M$, called the *auxiliary design*, does not need to be the sampling design $\mathcal{S}_n$, but is constructed in a way that contains the convex hull of the region of interest $\mathcal{D}$. Associated with $\mathcal{U}_M$ we define a corresponding set of M spatial frequencies, also forming a regular rectangular grid in the plane, as

$$\begin{aligned}\mathcal{W}_M &= \left\{\boldsymbol{\omega}_{-\frac{M_1}{2}+1,-\frac{M_2}{2}+1}, \ldots, \boldsymbol{\omega}_{0,0}, \ldots, \boldsymbol{\omega}_{\frac{M_1}{2},\frac{M_2}{2}}\right\} \\ &= \frac{2\pi}{\Delta M_1}\left\{-\frac{M_1}{2}+1, \ldots, 0, 1, \ldots, \frac{M_1}{2}\right\} \times \frac{2\pi}{\Delta M_2}\left\{-\frac{M_2}{2}+1, \ldots, 0, 1, \ldots, \frac{M_2}{2}\right\};\end{aligned}$$

$\mathcal{W}_M \subset [-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]^2$ is called the *spectral design*.

Now, let $Z_\Delta(\boldsymbol{k}) := Z(\Delta\boldsymbol{k})$, $\boldsymbol{k} = (k_1, k_2)^\top \in \mathbb{Z}^2$, be the discrete index random field defined by sampling the random field $Z(\cdot)$ at the rate Δ. This random field has mean function $\mu(\Delta\boldsymbol{k})$ and covariance function $\sigma^2 K_\nu(\Delta||\boldsymbol{k} - \boldsymbol{k}'||)$, for $\boldsymbol{k}, \boldsymbol{k}' \in \mathbb{Z}^2$ while, due to the so–called 'aliasing effect', its spectral density function is given by (Yaglom, 1987)

$$f_\nu^\Delta(\boldsymbol{\omega}) = \sum_{\boldsymbol{l}\in\mathbb{Z}^2} f_\nu\Big(\boldsymbol{\omega} + \frac{2\pi}{\Delta}\boldsymbol{l}\Big), \qquad \boldsymbol{\omega} \in \Big[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}\Big]^2. \tag{5.2}$$

Let $\tilde{\boldsymbol{z}} := \big(Z(\mathbf{u}_{i,j}) + W(\mathbf{u}_{i,j}) : \mathbf{u}_{i,j} \in \mathcal{U}_M\big)^\top$ be the vector of (potential) observations at the points of the auxiliary design, where $W(\cdot)$ is the Gaussian white noise in the plane with variance $\sigma^2\xi$. From the spectral representation of $Z_\Delta(\boldsymbol{k})$, De Oliveira and Han (2023) showed that when M_1 and M_2 are both large, it holds that

$$\tilde{\boldsymbol{z}} \overset{\text{approx}}{\sim} \mathrm{N}\big(\tilde{\boldsymbol{X}}\boldsymbol{\beta}, \sigma^2(\boldsymbol{H}_1\boldsymbol{G}_\nu\boldsymbol{H}_1^\top + \xi I_M)\big), \tag{5.3}$$

where $\tilde{\boldsymbol{X}}$ is the $M \times p$ matrix whose entries involve the covariates measured at the locations in $\mathcal{U}_M$, $\boldsymbol{H}_1 := (\mathbf{1}_M, \boldsymbol{H})$, with $\mathbf{1}_M$ the vector of ones and $\boldsymbol{H}$ the $M \times (M-1)$

matrix whose columns are formed by the multiples 1, 2 or -2 of either cosines or sines evaluated at the inner products of appropriate frequencies and auxiliary locations, and

$$\boldsymbol{G}_\nu = \frac{c_\Delta}{2M}\mathrm{diag}\Big(\big(2f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1,m_2) \in I_C\big)^\top,\ \big(f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1,m_2) \in I\big)^\top,\ \big(f_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1,m_2) \in I\big)^\top\Big),$$

where $c_\Delta \coloneqq (2\pi/\Delta)^2$ and I_C and I are indices that identify subsets of the spectral design $\mathcal{W}_M$; see De Oliveira and Han (2023) for details.

5.1 Approximate Integrated Likelihood

The distribution in (5.3) provides the basis to approximate the integrated likelihood of $\boldsymbol{\theta}$ based on a special linear combination of $\tilde{\boldsymbol{z}}$, somewhat similar to the use of restricted likelihood for inference about variance components. In all that follows, the aliased spectral density $f_\nu^\Delta(\boldsymbol{\omega})$ is approximated by truncating the series (5.2) so that only the terms for which $\max\{|l_1|,|l_2|\} \leq T$ are retained, for some $T \in \mathbb{N}$. This approximation, denoted as $\tilde{f}_\nu^\Delta(\boldsymbol{\omega})$, is not sensitive to T once this is large enough; see Appendix E of the Supplementary Material for details (Han and De Oliveira, 2024).

Let $\boldsymbol{L}_1 \coloneqq \boldsymbol{H}_1(\boldsymbol{H}_1^\top\boldsymbol{H}_1)^{-1/2}$, where $\boldsymbol{H}_1$ is the matrix defined in the previous section, and $\boldsymbol{V}_1 \coloneqq \boldsymbol{L}_1^\top\tilde{\boldsymbol{z}}$. It holds that (De Oliveira and Han, 2023)

$$\boldsymbol{V}_1 \overset{\text{approx}}{\sim} \mathrm{N}\big(\boldsymbol{X}_1\boldsymbol{\beta}, \sigma^2\tilde{\boldsymbol{\Lambda}}_\nu\big), \tag{5.4}$$

where $\boldsymbol{X}_1 \coloneqq \boldsymbol{L}_1^\top\tilde{\boldsymbol{X}}$ is an $M \times p$ matrix with full rank p, and $\tilde{\boldsymbol{\Lambda}}_\nu$ is the diagonal matrix

$$\tilde{\boldsymbol{\Lambda}}_\nu = c_\Delta\mathrm{diag}\Big(\big(\tilde{f}_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1,m_2) \in I_C\big)^\top,\ \big(\tilde{f}_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1,m_2) \in I\big)^\top,\ \big(\tilde{f}_\nu^\Delta(\boldsymbol{\omega}_{m_1,m_2}) : (m_1,m_2) \in I\big)^\top\Big) + \xi I_M. \tag{5.5}$$

Because $\boldsymbol{V}_1$ has a substantially simpler covariance structure than that of $\tilde{\boldsymbol{z}}$, the implementation of the reference prior algorithm described in Section 4 using the likelihood based on $\boldsymbol{V}_1$ results in substantial simplifications. In particular, when the mean is constant (so $p = 1$), direct calculation shows that $l^{\mathrm{I}}(\boldsymbol{\theta};\tilde{\boldsymbol{z}})$, the integrated log–likelihood function of $\boldsymbol{\theta} = (\sigma^2,\nu)$ based on $\tilde{\boldsymbol{z}}$, can be approximated by

$$l^{\mathrm{AI}}(\boldsymbol{\theta};\tilde{\boldsymbol{z}}) \ = \ -\frac{1}{2}\sum_{j=1}^{M-1}\left(\log\left(\sigma^2(c_\Delta\tilde{f}_\nu^\Delta(\boldsymbol{\omega}_j)+\xi)\right) + \frac{V_j^2}{\sigma^2(c_\Delta\tilde{f}_\nu^\Delta(\boldsymbol{\omega}_j)+\xi)}\right) + \tilde{c}, \tag{5.6}$$

where $\tilde{c}$ does not depend on $\boldsymbol{\theta}$, $\{\boldsymbol{\omega}_j\}$ are a re–indexing of the frequencies $\{\boldsymbol{\omega}_{m_1,m_2}\}$ appearing in (5.5), with $\boldsymbol{\omega}_{0,0}$ removed, and $V_1,\ldots,V_{M-1}$ are the last $M-1$ components of $\boldsymbol{V}_1$ (De Oliveira and Han, 2023). In this case, even more substantial simplifications accrue in the computation of approximate reference priors, since (5.6) is a matrix–free expression, the required expectations are simplified as the V_j s are independent gamma distributed variables, and differentiation with respect to the smoothness parameter is simplified as $l^{\mathrm{AI}}(\boldsymbol{\theta};\tilde{\boldsymbol{z}})$ is devoid of Bessel functions.

6 Approximate Reference Prior

The derivation of the approximate reference prior of ν, denoted $\pi^{\mathrm{AR}}(\nu)$, proceeds as follows. Rather than using the exact integrated likelihood (4.1) based on the data $\boldsymbol{z}$ measured at $\mathcal{S}_n$, we use the approximate integrated likelihood derived from the potential summary (5.4) measured at $\mathcal{U}_M$. This summary has a substantially simpler (diagonal) covariance matrix. This makes the computation and analysis of the resulting approximate reference prior much more manageable than that of the exact reference prior. In what follows it is assumed that the covariates, if any, are available everywhere in the region of interest.

Theorem 6.1 (Approximate Reference Prior)**.** *The approximate reference prior of* $(\boldsymbol{\beta}, \sigma^2, \nu)$ *derived from (5.4) is given by* $\pi^{\mathrm{AR}}(\boldsymbol{\beta}, \sigma^2, \nu) \propto \frac{\pi^{\mathrm{AR}}(\nu)}{\sigma^2}$*, with*

$$\pi^{\mathrm{AR}}(\nu) \propto \left\{ \mathrm{tr}\left[\left\{ \left(\frac{\partial}{\partial \nu} \tilde{\boldsymbol{\Lambda}}_\nu \right) \tilde{\boldsymbol{Q}}_\nu \right\}^2 \right] - \frac{1}{M-p} \left[\mathrm{tr}\left\{ \left(\frac{\partial}{\partial \nu} \tilde{\boldsymbol{\Lambda}}_\nu \right) \tilde{\boldsymbol{Q}}_\nu \right\} \right]^2 \right\}^{\frac{1}{2}}, \tag{6.1}$$

where $\tilde{\boldsymbol{\Lambda}}_\nu$ *was defined in (5.5) and* $\tilde{\boldsymbol{Q}}_\nu := \tilde{\boldsymbol{\Lambda}}_\nu^{-1} - \tilde{\boldsymbol{\Lambda}}_\nu^{-1} \boldsymbol{X}_1 (\boldsymbol{X}_1^\top \tilde{\boldsymbol{\Lambda}}_\nu^{-1} \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^\top \tilde{\boldsymbol{\Lambda}}_\nu^{-1}$.

Proof. The result follows from (4.3) by replacing $\boldsymbol{X}$ *with* $\boldsymbol{X}_1$ *and* $\boldsymbol{\Psi}_\nu$ *with* $\tilde{\boldsymbol{\Lambda}}_\nu$.

Note that $\tilde{\boldsymbol{\Lambda}}_\nu$ is a diagonal matrix, with diagonal elements having closed–form expressions devoid of Bessel functions, so computation of $(\partial/\partial\nu)\tilde{\boldsymbol{\Lambda}}_\nu$ is straightforward. The computation of $\tilde{\boldsymbol{Q}}_\nu$ only involves the inversion of the (small) $p \times p$ matrix $\boldsymbol{X}_1^\top \tilde{\boldsymbol{\Lambda}}_\nu^{-1} \boldsymbol{X}_1$, where the matrix $\boldsymbol{X}_1$ needs to be computed only once since $\boldsymbol{H}_1$ is fixed. The j^{th} diagonal element of the diagonal matrix $\left((\partial/\partial\nu)\tilde{\boldsymbol{\Lambda}}_\nu\right)\tilde{\boldsymbol{\Lambda}}_\nu^{-1}$ also has a closed–form expression which, after some algebra, is given by

$$\begin{aligned} \gamma_\nu(\boldsymbol{\omega}_j) &:= \frac{c_\Delta}{c_\Delta \tilde{f}_\nu^\Delta(\boldsymbol{\omega}_j) + \xi} \left(\frac{\partial}{\partial \nu} \tilde{f}_\nu^\Delta(\boldsymbol{\omega}_j) \right) \\ &= \frac{\sum_{\boldsymbol{l} \in \mathcal{T}_2} g_\nu(\boldsymbol{\omega}_{j(l_1,l_2)}) h(\nu, \boldsymbol{\omega}_{j(l_1,l_2)})}{q(\nu) + \sum_{\boldsymbol{l} \in \mathcal{T}_2} g_\nu(\boldsymbol{\omega}_{j(l_1,l_2)})}, \end{aligned} \tag{6.2}$$

where $\boldsymbol{\omega}_j$ is the frequency of the j^{th} diagonal element in (5.5), $\boldsymbol{l} = (l_1, l_2)$, $\mathcal{T}_2 := [-T, T]^2 \cap \mathbb{Z}^2$, $\boldsymbol{\omega}_{j(l_1,l_2)} := \boldsymbol{\omega}_j + \frac{2\pi}{\Delta}\boldsymbol{l}$, and

$$\begin{aligned} g_\nu(\boldsymbol{\omega}_{j(l_1,l_2)}) &:= \left(1 + \frac{\vartheta^2 \|\boldsymbol{\omega}_{j(l_1,l_2)}\|^2}{4\nu} \right)^{-(\nu+1)} \\ h(\nu, \boldsymbol{\omega}_{j(l_1,l_2)}) &:= \psi(\nu+1) - \psi(\nu) + \frac{\vartheta^2 \|\boldsymbol{\omega}_{j(l_1,l_2)}\|^2 - 4}{\vartheta^2 \|\boldsymbol{\omega}_{j(l_1,l_2)}\|^2 + 4\nu} - \log\left(1 + \frac{\vartheta^2 \|\boldsymbol{\omega}_{j(l_1,l_2)}\|^2}{4\nu} \right) \\ q(\nu) &:= \frac{\xi \Delta^2 \Gamma(\nu) \nu}{\pi \vartheta^2 \Gamma(\nu+1)}. \end{aligned}$$

As a result, the computation and analysis of the approximate reference prior $\pi^{\mathrm{AR}}(\nu)$ is substantially simpler than that of $\pi^{\mathrm{R}}(\nu)$. An important special case of Theorem 6.1

occurs when the mean function is constant, in which case the approximate reference prior of ν takes an even simpler matrix–free form.

Corollary 6.1 (Approximate Reference Prior–Constant Mean)**.** *Consider the model with constant mean function. In this case the approximate reference prior of* (β_1, σ^2, ν) *is* $\pi^{\mathrm{AR}}(\beta_1, \sigma^2, \nu) \propto \frac{\pi^{\mathrm{AR}}(\nu)}{\sigma^2}$*, with*

$$\pi^{\mathrm{AR}}(\nu) \propto \left\{ \sum_{j=1}^{M-1} \gamma_\nu^2(\boldsymbol{\omega}_j) - \frac{1}{M-1} \left(\sum_{j=1}^{M-1} \gamma_\nu(\boldsymbol{\omega}_j) \right)^2 \right\}^{\frac{1}{2}}, \tag{6.3}$$

where $\{\boldsymbol{\omega}_j\}$ *are the same frequencies used in (5.6) and* $\gamma_\nu(\boldsymbol{\omega}_j)$ *is given in (6.2).*

Proof. The proof is similar to the proof in De Oliveira and Han (2023, Corollary 1).

Theorem 6.2 (Propriety)**.** *Assume the mean function* $\mu(\mathbf{s})$ *has an intercept (so* $f_1(\mathbf{s}) \equiv 1$*) and the covariance function is the Matérn family in (2.1). Then*

(a) $\pi^{\mathrm{AR}}(\nu)$ *is a continuous function on* $(0, \infty)$ *that satisfies*

$$\pi^{\mathrm{AR}}(\nu) = \begin{cases} O(1), & \text{as } \nu \to 0^+ \\ O(\nu^{-2}), & \text{as } \nu \to \infty, \end{cases}$$

so $\pi^{\mathrm{AR}}(\nu)$ *is integrable on* $(0, \infty)$*.*

(b) The approximate reference posterior distribution based on the observed data, namely, $\pi^{\mathrm{AR}}(\boldsymbol{\beta}, \sigma^2, \nu \mid \boldsymbol{z}) \propto L(\boldsymbol{\beta}, \sigma^2, \nu; \boldsymbol{z}) \pi^{\mathrm{AR}}(\boldsymbol{\beta}, \sigma^2, \nu)$*, is proper.*

Proof. The proof is given in Appendix A of the Supplementary Material (Han and De Oliveira, 2024).

Interestingly, the ad–hoc default prior $\pi^{\mathrm{HS}}(\nu) = (1+\nu)^{-2}$ mentioned in the Introduction has the same tail behavior as $\pi^{\mathrm{AR}}(\nu)$. But unlike $\pi^{\mathrm{HS}}(\nu)$, the behavior of $\pi^{\mathrm{AR}}(\nu)$ near zero changes depending on the model structure (e.g., on the assumed mean function and range parameter). It will be shown in the next section and in Appendix C of the Supplementary Material (Han and De Oliveira, 2024) that, for most designs and models, Bayesian inferences based on both priors are similar and have satisfactory frequentist properties.

7 Numerical Studies

In this section, we conduct three numerical studies. We compare the marginal priors $\pi^{\mathrm{R}}(\nu)$ and $\pi^{\mathrm{AR}}(\nu)$ for various sampling designs and model features, compare the computational effort of these priors, and compare the frequentist properties of Bayesian predictive inferences based on default priors with those derived from several variations of plug–in prediction. Additionally, in Appendices C and D of the Supplementary Material (Han and De Oliveira, 2024) we investigate frequentist properties of Bayesian estimates based on several default priors, and in Appendix E of the Supplementary Material (Han and De Oliveira, 2024) we explore further the empirical rules for hyperparameter tuning.

7.1 Comparison of Exact and Approximate Reference Priors of ν

As sampling designs in $\mathcal{D} = [0,1]^2$ we use a 10×10 equally spaced regular design and three irregular designs of size $n = 100$, to be described below. For the mean function we use $\mu(\mathbf{s}) = 1$ and $\mu(\mathbf{s}) = 0.15 - 0.65x - 0.1y + 0.9x^2 - xy + 1.2y^2$, with $\mathbf{s} = (x, y)$, and for the covariance function we use the isotropic Matérn model (2.1) with $\sigma^2 = 1$ and $\vartheta = 0.1, 0.3$ and 0.5. We also use two noise–to–signal variance ratios, $\xi = 0$ and 0.5. To compute the approximate reference priors, $f_\nu^{\Delta}(\boldsymbol{\omega})$ is approximated by truncating the series (5.2) so that only the terms with $\max\{|l_1|, |l_2|\} \leq 4$ are retained. These approximate reference priors show very mild sensitivity to the truncation point, and $T = 4$ is an adequate default value; see Appendix E of the Supplementary Material (Han and De Oliveira, 2024). To compute exact reference priors, the entries of $(\partial/\partial\nu)\boldsymbol{\Sigma}_\nu$ are evaluated using the expression derived in De Oliveira and Han (2022).

Figure 3 displays plots of the (normalized) exact and approximate reference priors of ν based on data in the regular design without measurement error ($\xi = 0$). The left panels

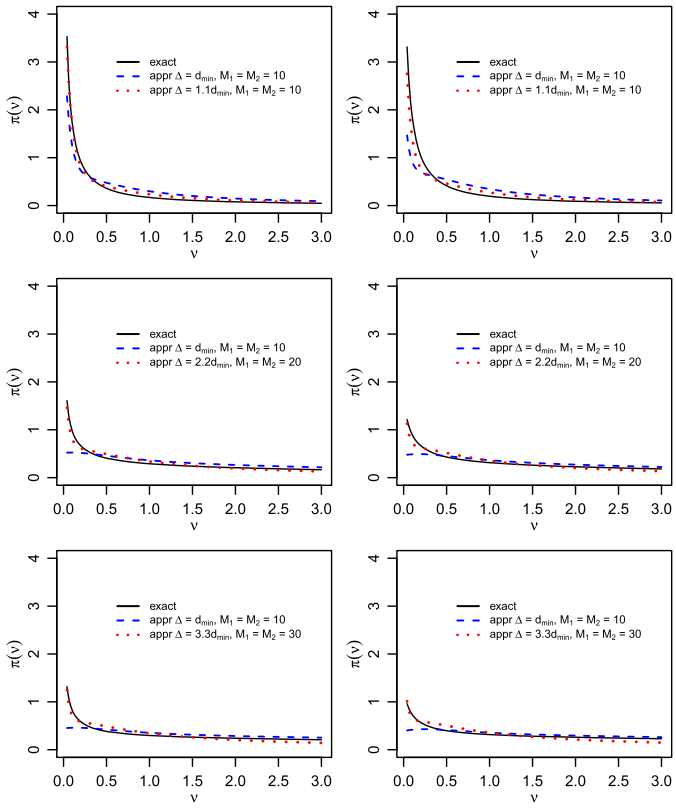

Figure 3: Marginal densities of the exact and approximate reference priors of ν for data without measurement error ($\xi = 0$) in the 10×10 regular design. Left: constant mean model. Right: non–constant mean model. From top to bottom: $\vartheta = 0.1, 0.3$ and 0.5.

are the priors for the constant mean models, and the right panels are the priors for the non–constant mean models. The top, middle, and bottom panels are the priors obtained when $\vartheta = 0.1, 0.3$ and 0.5, respectively. The solid black curves are exact reference priors, and the broken colored curves are approximate reference priors. As the default choice to compute $\pi^{\mathrm{AR}}(\nu)$ we use $\mathcal{U}_M = \mathcal{S}_n$, i.e., we set $M_1 = M_2 = \sqrt{n}$ and $\Delta = d_{\min}$, the distance between adjacent sampling locations. The resulting approximate reference priors (broken blue curves) have similar overall behaviors as the exact reference priors, but display some discrepancies for small values of ν, especially when ϑ is small. However, the approximation improves when M_1, M_2 and Δ are tuned using values $M_1 = M_2 > \sqrt{n}$ and $\Delta > d_{\min}$ (dotted red curves), in which case the two priors are close everywhere.

Figure 4 displays plots of the (normalized) exact and approximate reference priors of ν based on data in the regular design with measurement error ($\xi = 0.5$). The layout and curve labels are the same as those in Figure 3. In this case, the default choices of M_1, M_2 and Δ indicated above result in approximate reference priors that are close

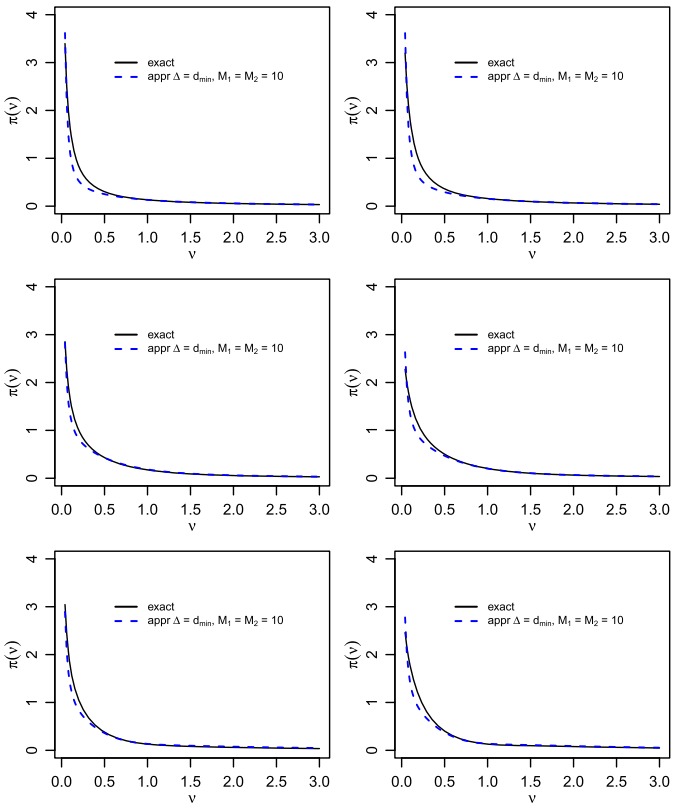

Figure 4: Marginal densities of the exact and approximate reference priors of ν for data with measurement error ($\xi = 0.5$) in the 10×10 regular design. Left: constant mean model. Right: non–constant mean model. From top to bottom: $\vartheta = 0.1, 0.3$ and 0.5.

to their exact counterparts. The same behaviors were found to hold for other positive values of ξ (not shown).

Next we consider the three irregular sampling designs displayed in Figure 5 (left panels). Figure 5 (right panels) displays plots of the exact and approximate reference priors of ν based on these irregular designs for the model with constant mean, $\vartheta = 0.3$ and $\xi = 0$. The choices of M_1, M_2, and Δ were made similarly as for the regular design with a slightly larger spacing Δ. Specifically, setting $M_1 = M_2 = 20$ and Δ equal to 1.5 times a value between the 75th to 95th percentile of $\{d_i\}_{i=1}^n$, with $d_i = \min\{\|\mathbf{s}_i - \mathbf{s}_j\| : j \neq i\}$, provides satisfactory approximations for practical purposes. Additionally, the plots show that these reference priors are not sensitive to the sampling design. Similar behaviors were found to hold for other models with non–constant mean and other values of ϑ and ξ (not shown) Overall, the numerical explorations reported in Figures 3–5 indicate that the approximate reference priors, after properly tuned when needed, provide satisfactory approximations to exact reference priors for a variety of sampling designs and models. The larger the noise–to–signal variance ratio, the less tuning is needed, and approximate reference priors without tuning provide good approximations when ξ is moderate

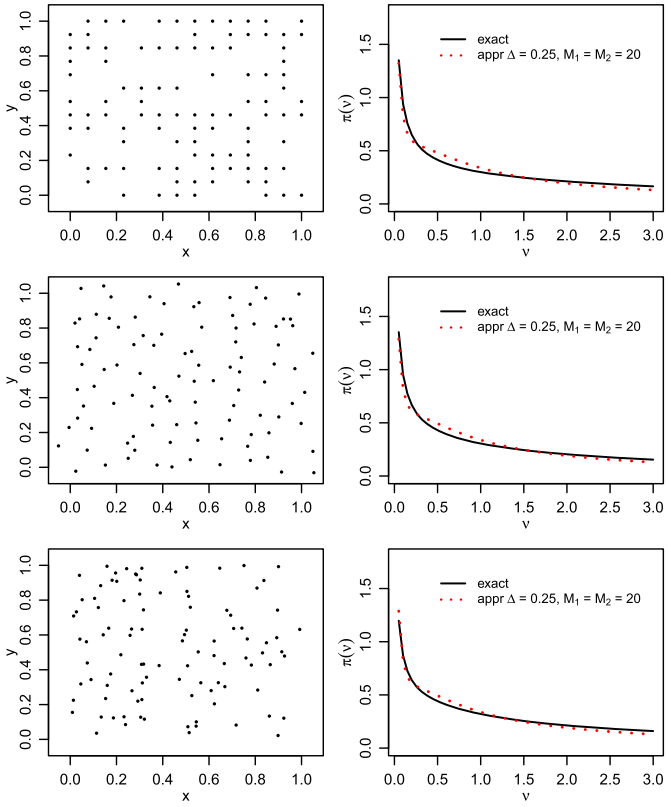

Figure 5: Left: Three irregular sampling designs in $[0, 1]^2$. Right: Corresponding exact and approximate reference priors of ν when the mean is constant, $\vartheta = 0.3$ and $\xi = 0$.

	p	Reference Prior	$n = 100$	$n = 400$	$n = 1600$	$n = 10000$
Regular Design	1	Exact	0.82	14.54	596.17	–
		Approximate	0.02	0.07	0.22	1.82
	6	Exact	0.95	17.18	874.05	–
		Approximate	0.11	1.07	8.04	187.45
Irregular Design	1	Exact	33.28	521.01	6310.31	–
		Approximate	0.02	0.08	0.19	1.90
	6	Exact	37.71	628.11	10840.67	–
		Approximate	0.10	1.14	9.94	204.32

Table 3: Computational times (in seconds) for 100 evaluations of the exact and approximate reference priors of ν based two types of sampling designs of different sizes, for models with constant and non-constant mean functions and $(\vartheta, \xi) = (0.2, 0)$.

or large. In these cases, setting $M_1 = M_2 = \sqrt{n}$ and $\Delta = d_{\min}$ is recommended. Overall, the approximate reference prior is not overly sensitive to the choices of T, M_1, M_2 and Δ, and the above general recommendations provide satisfactory approximations. Additional sensitivity explorations are provided in Appendix E of the Supplementary Material (Han and De Oliveira, 2024).

7.2 Comparison of Computational Effort

To compare the computational complexity of exact and approximate reference priors, we consider two types of sampling designs of size n, a $\sqrt{n} \times \sqrt{n}$ equally spaced grid and an i.i.d. sample from the unif$((0,1)^2)$ distribution. To compute the approximate reference prior we use $M_1 = M_2 = \sqrt{n}$ and $\Delta = 0.1$ for both design types. To compute the exact reference prior for irregular designs, the derivative w.r.t. ν needs to be evaluated $n(n-1)/2$ times, while for regular designs the derivative calculations only need to be done for pairs with distinct distances, whose total number is far less. The computation of the exact reference prior in (4.3) also requires $O(n^3)$ operations due to the matrix inversion, regardless of the design. On the other hand, for process with constant mean the computation of the approximate reference prior in (6.3) only requires $O(n)$ operations. For processes with a non–constant mean function the computation of $\pi^{\mathrm{AR}}(\nu)$ in (6.1) requires $O(n^2)$ operations, and does not involve the evaluation of special functions or their derivatives. The major computational cost is to evaluate $\boldsymbol{X}_1(\boldsymbol{X}_1^\top \tilde{\boldsymbol{\Lambda}}_\nu^{-1} \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^\top \tilde{\boldsymbol{\Lambda}}_\nu^{-1}$, where $\boldsymbol{X}_1$ only needs to be calculated once. Computing the HS prior is almost instantaneous as it does not depend on the design or model.

Table 3 reports the timings for 100 evaluations[1] of exact and approximate reference priors of ν based on the two types of sampling designs of different sizes for models with constant and non–constant mean functions. Depending on the sample size, design type and mean function, the evaluation of approximate reference priors is between one to

[1]The computational times reported in this article are based on a MacBook Pro with 2.3 GHz Intel Core i9 processor under the `R` programming language.

three orders of magnitude faster than the evaluation of exact reference priors. And the computational time gap increases substantially with sample size, at a higher rate for models with constant mean. In particular, when $n = 10000$ the computation of exact reference priors becomes computationally unfeasible on a personal computer due to the challenges of storing and inverting large correlation matrices. It is also noted that, for a given sample size, the time to compute approximate reference priors is about the same regardless of the design type, while the time to compute exact reference priors is considerably larger for irregular designs than for regular designs.

7.3 Comparison of Predictive Inferences

We carry out a simulation experiment to compare frequentist properties of plug–in and Bayesian predictive inferences based on different priors. We use a 10×10 equally spaced regular design in $\mathcal{D} = [0, 1]^2$ and three prediction locations $\mathbf{s}_0$ at $(0.5, 0.5)$, $(0.3, 0.3)$ and $(0.79, 0.1)$ (with decreasing distance to their nearest sampling locations). We consider Gaussian random fields with $\mu = 0$, $\sigma^2 = 1$ and Matérn correlation function with $\vartheta = 0.3$, and $\nu = 0.5$ or 1.5. For each of these models, we simulated 1000 independent data sets at the 103 locations with noise–to–signal variance ratio $\xi = 0$ or 0.2. For each model and data set we computed equal–tail 95% prediction intervals for $Z(\mathbf{s}_0)$ using plug–in and Bayesian procedures. For the plug–in procedures, we used Gaussian predictive distributions with parameters estimated by the following variants of MLE:

- MLE: all parameters were estimated
- MLE–0.5: all parameters were estimated except ν that was fixed at 0.5.
- MLE–1.5: all parameters were estimated except ν that was fixed at 1.5.

The last two variants use a correctly specified value of ν for half of the simulated scenarios and a misspecified value for the other half. For the Bayesian procedures, we used a sample from the posterior distribution of ν (obtained from the algorithm in Appendix B of the Supplementary Material (Han and De Oliveira, 2024) based on the following default priors:

- Bayes–AR: the approximate reference prior in Corollary 6.1.
- Bayes–HS: similar as the previous one, but using $\pi^{\mathrm{HS}}(\nu)$ instead of $\pi^{\mathrm{AR}}(\nu)$.

For each model and prediction location, we evaluated the predictive performance of the aforementioned procedures using two metrics: the frequentist coverage probability and the interval score proposed by Gneiting and Raftery (2007). The latter is given by

$$S^{\mathrm{int}}_{0.05}(l, u;\, Z(\mathbf{s}_0)) = (u - l) + 40\big((l - Z(\mathbf{s}_0))\mathbf{1}\{Z(\mathbf{s}_0) < l\} + (Z(\mathbf{s}_0) - u)\mathbf{1}\{Z(\mathbf{s}_0) > u\}\big), \tag{7.1}$$

where l and u are, respectively, the lower and upper bounds of the 95% prediction interval for $Z(\mathbf{s}_0)$. The latter criterion rewards narrow intervals and imposes a penalty

Method	$\xi = 0$			$\xi = 0.2$		
	Point 1	Point 2	Point 3	Point 1	Point 2	Point 3
			$\nu = 0.5$			
MLE	0.944	0.922	0.857	0.946	0.926	0.889
	[2.463]	[2.656]	[2.221]	[3.377]	[3.846]	[4.066]
MLE–0.5	0.961	0.950	0.950	0.949	0.935	0.889
	[2.352]	[2.514]	[1.675]	[3.267]	[3.657]	[3.524]
MLE–1.5	0.902	0.867	0.696	0.948	0.936	0.904
	[2.526]	[2.958]	[3.164]	[3.384]	[3.746]	[3.830]
Bayes–AR	0.952	0.932	0.898	0.955	0.935	0.908
	[2.441]	[2.619]	[1.980]	[3.329]	[3.708]	[3.601]
Bayes–HS	0.954	0.938	0.911	0.952	0.939	0.908
	[2.429]	[2.596]	[1.940]	[3.317]	[3.700]	[3.529]
			$\nu = 1.5$			
MLE	0.948	0.932	0.914	0.946	0.929	0.876
	[0.963]	[0.939]	[0.447]	[2.582]	[2.946]	[3.450]
MLE–0.5	0.999	0.997	1.000	0.955	0.929	0.851
	[1.402]	[1.327]	[0.941]	[2.467]	[2.841]	[3.659]
MLE–1.5	0.961	0.949	0.953	0.943	0.928	0.882
	[0.927]	[0.897]	[0.398]	[2.579]	[2.918]	[3.248]
Bayes–AR	0.958	0.942	0.936	0.951	0.930	0.892
	[0.965]	[0.928]	[0.440]	[2.559]	[2.891]	[3.194]
Bayes–HS	0.960	0.947	0.939	0.948	0.930	0.889
	[0.968]	[0.929]	[0.439]	[2.544]	[2.874]	[3.216]

Table 4: Coverage probability and [interval score] of MLE–based Wald 95% prediction intervals and Bayesian equal–tail 95% credible intervals for prediction of $Z(\mathbf{s}_0)$ at locations $\mathbf{s}_0 = (0.5, 0.5)$, $(0.3, 0.3)$ and $(0.79, 0.1)$ (Points 1, 2 and 3, respectively).

on intervals that do not capture the true value, with the magnitude of the penalty proportional to the gap between the true value and a boundary of the interval.

Table 4 reports, for each combination of model and procedure, the empirical frequentist coverage and average interval scores for the three prediction locations. Both Bayesian prediction intervals, derived from default priors, exhibit similar performance in terms of coverage and interval score. Notably, the Bayesian approaches slightly outperform the MLE procedure (where all parameters are estimated) in these respects, especially when $\xi = 0$. Additionally, the Bayesian procedures have a much better performance than the MLE procedure when the value of ν is fixed and misspecified, both in terms of coverage and interval score. On the other hand, the Bayesian procedures have a similar or slightly worse performance than the MLE procedure when the value of ν is fixed at the correct value. The above findings suggest that: (a) Predictive inferences based on both default priors are about the same; (b) Typically, setting a fixed smoothness adversely affects predictive performance. The only exception arises when the selected value of ν equals the true value of the smoothness of the random field, a

fortuitous situation that is rare in practice. We also compared the mean square prediction errors of the aforementioned procedures, but all exhibited similar performances based on this criterion (not shown).

Appendices C and D of the Supplementary Material (Han and De Oliveira, 2024) compare frequentist properties of Bayesian inferences about covariance parameters based on the approximate reference and HS priors. It is found that these have similar satisfactory properties. The only exception appears to occur in non–smooth covariance functions for which inference about ν based on the approximate reference prior is better.

8 Example

To illustrate the proposed default Bayesian analysis for the smoothness of random fields, we use the rainfall data set analyzed in Diggle and Ribeiro (2007, Section 5.4.7), available from the `R` package `geoR` (object `SIC`). The data consists of daily rainfall totals that felt in Switzerland on May 8, 1986, measured in 1/10 of a millimeter, and collected over 467 stations with the coordinates of the sampling locations measured in kilometers; see Figure 6. An exploratory analysis reveals no apparent spatial trend. The rainfall at five stations is zero and the sampling distribution of the positive rainfall is skewed. As Diggle and Ribeiro (2007), we replaced the zeros with 0.5 and transformed the rainfall totals using the Box–Cox transformation with parameter $\lambda = 0.5$. The model for the transformed data is the Gaussian random field in (2.2) with constant mean and Matérn covariance function (2.1). Instead of fixing the smoothness of the random field at an arbitrary value, as customarily done in geostatistical analyses, we conduct Bayesian analyses and compare inferences based on default priors $\pi(\nu)/\sigma^2$, with $\pi(\nu) = \pi^{\mathrm{R}}(\nu)$, $\pi^{\mathrm{AR}}(\nu)$ and $\pi^{\mathrm{HS}}(\nu)$.

To compute $\pi^{\mathrm{R}}(\nu)$ and $\pi^{\mathrm{AR}}(\nu)$ we need to determine the range and the noise–to–signal variance ratio. For this we use the integrated likelihood approach proposed by

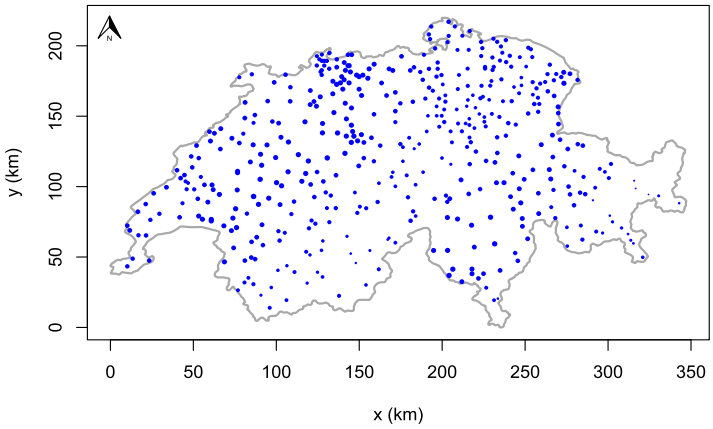

Figure 6: Sampling locations of the Swiss rainfall data. The radius of each circle is proportional to the logarithm of the rainfall amount (in 1/10 millimeter).

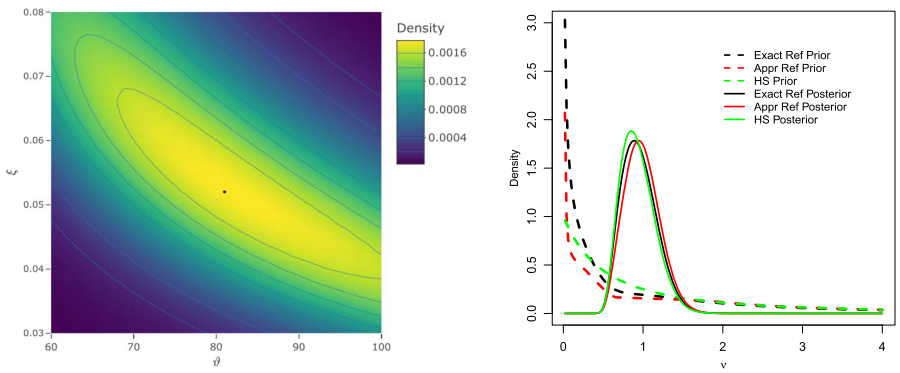

Figure 7: Left: Contour plot of the integrated likelihood of (ϑ, ξ) for the rainfall data. Right: Densities of the three priors and their corresponding posteriors of ν.

Berger et al. (2001) with the approximate reference prior. Specifically, when (ϑ, ξ) are assumed known, the (conditional) approximate reference prior is of the form

$$\pi^{\mathrm{AR}}(\boldsymbol{\beta}, \sigma^2, \nu \mid \vartheta, \xi) = \frac{C(\vartheta, \xi)\pi^{\mathrm{AR}}(\nu \mid \vartheta, \xi)}{\sigma^2},$$

where $\pi^{\mathrm{AR}}(\nu \mid \vartheta, \xi)$ is given in (6.3), with the dependence on the parameters (ϑ, ξ) now being explicit, and $C(\vartheta, \xi) := \left(\int_0^\infty \pi^{\mathrm{AR}}(\nu \mid \vartheta, \xi)\, d\nu\right)^{-1}$. If $\boldsymbol{\vartheta} = (\vartheta, \nu)$ denotes the correlation parameters, then the integrated likelihood of (ϑ, ξ) is given by

$$\begin{aligned} m(\boldsymbol{z} \mid \vartheta, \xi) &= \int_{\mathbb{R}^p \times (0,\infty)^2} L(\boldsymbol{\beta}, \sigma^2, \xi, \vartheta, \nu; \boldsymbol{z}) \pi^{\mathrm{AR}}(\boldsymbol{\beta}, \sigma^2, \nu \mid \vartheta, \xi)\, d\boldsymbol{\beta} d\sigma^2 d\nu \\ &\propto \int_0^\infty |\boldsymbol{\Psi}_{\boldsymbol{\vartheta},\xi}|^{-\frac{1}{2}} |\boldsymbol{X}^\top \boldsymbol{\Psi}_{\boldsymbol{\vartheta},\xi}^{-1} \boldsymbol{X}|^{-\frac{1}{2}} (S^2_{\boldsymbol{\vartheta},\xi})^{-\frac{n-p}{2}} C(\vartheta, \xi) \pi^{\mathrm{AR}}(\nu \mid \vartheta, \xi)\, d\nu, \end{aligned}$$

where $\boldsymbol{\Psi}_{\boldsymbol{\vartheta},\xi}$ and $S^2_{\boldsymbol{\vartheta},\xi}$ are defined circa (4.1), but now we make their dependence on ν, ϑ and ξ explicit. Then (ϑ, ξ) are chosen as the values that maximizes $m(\boldsymbol{z} \mid \vartheta, \xi)$, with $\boldsymbol{z}$ set at the observed data. Figure 7 (left) displays the contour plot of the integrated likelihood of (ϑ, ξ) for the rainfall data, showing that the maximum occurs around $(\vartheta, \xi) = (82, 0.052)$. Additionally, to compute $\pi^{\mathrm{AR}}(\nu)$ we approximate $\tilde{f}^{\Delta}_{\nu}(\boldsymbol{\omega}_j)$ by truncating the series (5.2) so that only the terms for which $\max\{|l_1|, |l_2|\} \leq 4$ are retained and, following the guidelines discussed in Section 7.1, set $M_1 = M_2 = 32$ and $\Delta = 7$ (= 75th percentile of the distances between nearest sampling locations).

We carry out Bayesian analyses based of the aforementioned three default priors, using the Monte Carlo algorithm described in Appendix B of the Supplementary Material (Han and De Oliveira, 2024). For each analysis, we run a simulation of size 12000 from the posterior distribution of (β_1, σ^2, ν), and the first 2000 draws are discarded as burn–in. Figure 7 (right) displays the three priors of ν and their corresponding marginal posteriors. All three priors place large and small probability masses in about the same

Prior	$\hat{\beta}_1$ (95% CI)	$\hat{\sigma}^2$ (95% CI)	$\hat{\nu}$ (95% CI)
Exact Reference	19.790 (10.776, 28.535)	127.517 (79.190, 174.109)	0.888 (0.565, 1.390)
Approximate Reference	19.779 (11.539, 29.224)	131.310 (82.594, 177.237)	0.946 (0.589, 1.424)
HS	19.828 (10.312, 28.172)	125.257 (78.541, 170.664)	0.851 (0.561, 1.346)

Table 5: Parameter estimates from the rainfall data using three different priors. The estimate $\hat{\nu}$ is the posterior mode, $\hat{\sigma}^2$ is the posterior median and $\hat{\beta}_1$ is the posterior mean. The 95% credible intervals are the HPD.

regions of the parameter space, but some discrepancies exist when ν is small. Nevertheless, the posterior distributions corresponding to the exact and approximate reference priors are quite close, while that corresponding to the HS prior is slightly less similar. Table 5 reports the Bayesian estimates of the model parameters and their corresponding 95% highest posterior density (HPD) credible intervals, showing that inferences based on the three priors are similar. The posterior of ν using the HS prior is slightly more concentrated than the other two and, as a result, the credible intervals are slightly narrower.

9 Conclusions and Discussion

This work proposes an easy–to–compute default prior for a class of Gaussian Matérn random fields with unknown smoothness parameter. The prior is obtained by approximating a reference prior using the spectral approximation to stationary random fields. This approximate reference prior has several advantages over the exact reference prior. First, the computation of the former is more stable and considerably less burdensome than that of the latter. Yet, results from extensive simulation experiments in Section 7.3, Appendix C of the Supplementary Material (Han and De Oliveira, 2024) and a data analysis in Section 8 suggest that Bayesian inferences based on both priors are practicably equivalent and have satisfactory frequentist properties. Second, both the marginal prior of the smoothness parameter and the joint posterior of all parameters are proper for the approximate reference prior, while the status of these for the exact reference prior is unknown. This enables the use of approximate reference priors for covariance function selection using Bayes factors, as described in Berger et al. (2001), a very helpful property since few tools are available for this purpose, and covariance selection is often done casually. Recommendations were given for the tuning of the approximate reference prior in terms of quantities that depend of the sampling design. The resulting prior is just mildly sensitive to the tuning choices.

To the best of our knowledge, the proposed default prior for the smoothness parameter is the first based on first principles. Interestingly, the ad–hoc HS prior has similar

overall and tail behaviors, so it is also a reasonable default prior for the smoothness parameter. On the other hand, we advise against using the seemingly 'non–informative' uniform prior with a large subjective upper bound, since Bayesian inferences based on it have poor frequentist properties; see Appendix C of the Supplementary Material (Han and De Oliveira, 2024).

It was also found that the MLE of the smoothness parameter has poor sampling properties when the data contain measurement error. This estimator tends to severely overestimate the smoothness and may not even exist for some data sets. In contrast, the estimator of the smoothness parameter obtained by maximizing its marginal posterior based on the approximate reference prior has good sampling properties as this prior penalizes large smoothness values. This finding resembles that in Gu et al. (2018) who showed that estimators of range parameters obtained by maximizing their marginal posteriors based on reference priors have better sampling properties than those of MLEs.

A drawback of the proposed methodology is that requires assuming the range and noise–to–signal variance ratio are known. This is partly mitigated by choosing these parameters by maximizing their integrated likelihood, as described in Section 8. These estimators are expected to be reasonable for most data sets, and typically better than ML estimators; see Berger et al. (1999). The developments in this work and that in De Oliveira and Han (2023) are expected to serve as the basis for the construction of a default prior for the case when all four parameters of the Matérn covariance function are unknown.

Acknowledgments

The authors would like to thank an anonymous referee, an Associate Editor and an Editor for their constructive comments and suggestions that improved the quality of this paper.

Funding

The first author was partially supported by the National Natural Science Foundation of China (No. 12201112 and No. 12371264), and by the Fundamental Research Funds for the Central Universities, China in UIBE (CXTD14-05). The second author was partially supported by the U. S. National Science Foundation grant DMS–2113375.

Supplementary Material

Supplementary Material of "Default Priors for the Smoothness Parameter in Gaussian Matérn Random Fields" (DOI: 10.1214/24-BA1431SUPP; .pdf). The supplementaty material available online contains five appendices. Appendix A provides auxiliary lemmas and the proofs of Lemma 1 and Theorem 6.2. Appendix B describes the Monte Carlo algorithm used for fitting. Appendix C explores frequentist properties of Bayesian estimative inferences based on several default priors. Appendix D provides additional frequentist properties of Bayesian inferences about the smoothness for non–smooth models. Appendix E explores further the guidelines given in Section 7.1 for the selection of the tuning constants in the approximate reference prior.

References

Anderes, E. and Stein, M. (2008). "Estimating Deformations of Isotropic Gaussian Random Fields on the Plane." *The Annals of Statistics,*, 36: 719–741. MR2396813. doi: https://doi.org/10.1214/009053607000000893. 2

Berger, J., De Oliveira, V., and Sansó, B. (2001). "Objective Bayesian Analysis of Spatially Correlated Data." *Journal of the American Statistical Association*, 96: 1361–1374. MR1946582. doi: https://doi.org/10.1198/016214501753382282. 3, 7, 8, 20, 21

Berger, J., Liseo, B., and Wolpert, R. (1999). "Integrated Likelihood Methods for Eliminating Nuisance Parameters." *Statistical Science*, 14: 1–28. MR1702200. doi: https://doi.org/10.1214/ss/1009211803. 22

Bose, M., Hodges, J., and Banerjee, S. (2018). "Toward a Diagnostic Toolkit for Linear Models with Gaussian–Process Distributed Random Effects." *Biometrics*, 74: 863–873. MR3860707. doi: https://doi.org/10.1111/biom.12848. 9

De Oliveira, V. (2007). "Bayesian Analysis of Spatial Data with Measurement Error." *The Canadian Journal of Statistics*, 35: 283–301. MR2393610. doi: https://doi.org/10.1002/cjs.5550350206. 3

De Oliveira, V. and Han, Z. (2022). "On Information about Covariance Parameters in Gaussian Matérn Random Fields." *Journal of Agricultural, Biological, and Environmental Statistics*, 27: 690–712. MR4498005. doi: https://doi.org/10.1007/s13253-022-00510-5. 2, 8, 13

De Oliveira, V. and Han, Z. (2023). "Approximate reference priors for Gaussian random fields." *Scandinavian Journal of Statistics*, 50: 296–326. MR4558737. 3, 9, 10, 12, 22

De Oliveira, V., Kedem, B., and Short, D. (1997). "Bayesian Prediction of Transformed Gaussian Random Fields." *Journal of the American Statistical Association*, 92: 1422–1433. MR1615252. doi: https://doi.org/10.2307/2965412. 1, 2

Diggle, P. and Ribeiro, P. (2007). *Model–Based Geostatistics*. Springer–Verlag. MR2293378. 1, 19

Geoga, C., Marin, O., Schanen, M., and Stein, M. (2023). "Fitting Matérn Smoothness Parameters Using Automatic Differentiation." *Statistics and Computing*, 33: 48. MR4556305. doi: https://doi.org/10.1007/s11222-022-10127-w. 2

Gneiting, T. and Raftery, A. E. (2007). "Strictly proper scoring rules, prediction, and estimation." *Journal of the American statistical Association*, 102: 359–378. MR2345548. doi: https://doi.org/10.1198/016214506000001437. 17

Gradshteyn, T. and Ryzhik, I. (2000). *Table of Integrals, Series, and Products, 6th edition*. Academic Press. MR0669666. 4

Gu, M., Wang, X., and Berger, J. (2018). "Robust Gaussian Stochastic Process Emulation." *The Annals of Statistics*, 46: 3038–3066. MR3851764. doi: https://doi.org/10.1214/17-AOS1648. 3, 22

Han, Z. and De Oliveira, V. (2016). "On the Correlation Structure of Gaussian Copula Models for Geostatistical Count Data." *Australian & New Zealand Journal of Statistics*, 58: 47–69. MR3499150. doi: https://doi.org/10.1111/anzs.12140. 1

Han, Z. and De Oliveira, V. (2024). "Supplementary Material for "Default Priors for the Smoothness Parameter in Gaussian Matérn Random Fields"." *Bayesian Analysis*. 8, 10, 12, 13, 16, 17, 19, 20, 21, 22

Handcock, M. and Stein, M. (1993). "A Bayesian Analysis of Kriging." *Technometrics*, 35: 403–410. 2, 3, 4, 8

Im, H., Stein, M., and Zhu, Z. (2007). "Semiparametric Estimation of Spectral Density With Irregular Observations." *Journal of the American Statistical Association*, 102: 726–735. MR2381049. doi: https://doi.org/10.1198/016214507000000220. 2

Kazianka, H. (2013). "Objective Bayesian Analysis of Geometrically Anisotropic Spatial Data." *Journal of Agricultural, Biological, and Environmental Statistics*, 18: 514–537. MR3142598. doi: https://doi.org/10.1007/s13253-013-0137-y. 3

Kazianka, H. and Pilz, J. (2012). "Objective Bayesian Analysis of Spatial Data with Uncertain Nugget and Range Parameters." *The Canadian Journal of Statistics*, 40: 304–327. MR2927748. doi: https://doi.org/10.1002/cjs.11132. 3

Kitanidis, P. (1986). "Parameter Uncertainty in Estimation of Spatial Functions: Bayesian Analysis." *Water Resources Research*, 22: 499–507. 2

Loh, W.-L. (2015). "Estimating the Smoothness of a Gaussian Random Field From Irregularly Spaced Data Via Higher–Order Quadratic Variations." *The Annals of Statistics*, 43: 2766–2794. MR3405611. doi: https://doi.org/10.1214/15-AOS1365. 2

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: 3127–3152. MR4352525. doi: https://doi.org/10.1214/21-aos2077. 2

Matérn, B. (1986). *Spatial Variation*. Springer–Verlag, 2nd edition. MR0867886. doi: https://doi.org/10.1007/978-1-4615-7892-5. 1

Paciorek, C. (2007). "Bayesian Smoothing with Gaussian Processes Using Fourier Basis Functions in the spectralGP package." *Journal of Statistical Software*, 19: 1–38. 9

Papritz, A. (2024). "georob: Robust Geostatistical Analysis of Spatial Data." *R package version 0.3-19*. URL https://cran.r-project.org/web/packages/georob/index.html 2

Paulo, R. (2005). "Default Priors for Gaussian Processes." *The Annals of Statistics*, 33: 556–582. MR2163152. doi: https://doi.org/10.1214/009053604000001264. 3

Ren, C., Sun, D., and He, C. (2012). "Objective Bayesian Analysis for a Spatial Model with Nugget Effects." *Journal of Statistical Planning and Inference*, 142: 1933–1946. MR2903403. doi: https://doi.org/10.1016/j.jspi.2012.02.034. 3

Ren, C., Sun, D., and Sahu, S. (2013). "Objective Bayesian Analysis of Spatial Mod-

els with Separable Correlation Functions." *The Canadian Journal of Statistics*, 41: 488–507. MR3101596. doi: https://doi.org/10.1002/cjs.11186. 3

Ribeiro, P. and Diggle, P. (2024). "geoR: Analysis of Geostatistical Data." *R package version 1.9-4*. URL https://cran.r-project.org/web/packages/geoR/index.html 2

Royle, J. A. and Wikle, C. (2005). "Efficient Statistical Mapping of Avian Count Data." *Environmental and Ecological Statistics*, 12: 225–243. MR2144403. doi: https://doi.org/10.1007/s10651-005-1043-4. 9

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer–Verlag. MR1697409. doi: https://doi.org/10.1007/978-1-4612-1494-6. 1, 9

Wu, W.-Y. and Lim, C. (2016). "Estimation of Smoothness of a Stationary Gaussian Random Field." *Statistica Sinica*, 26: 1729–1745. MR3586236. 2

Wu, W.-Y., Lim, C., and Xiao, Y. (2013). "Tail Estimation of the Spectral Density for a Stationary Gaussian Random Field." *Journal of Multivariate Analysis*, 116: 74–91. MR3049893. doi: https://doi.org/10.1016/j.jmva.2012.11.014. 2

Yaglom, A. (1987). *Correlation Theory of Stationary and Related Random Functions I: Basic Results*. Springer–Verlag. MR0893393. 9