Inverses of Matérn Covariances on Grids

Joseph Guinness

Cornell University, Department of Statistics and Data Science

Abstract

We conduct a study of the aliased spectral densities of Matérn covariance functions on a regular grid of points, providing clarity on the properties of a popular approximation based on stochastic partial differential equations. While others have shown that it can approximate the covariance function well, we find that it assigns too much power at high frequencies and does not provide increasingly accurate approximations to the inverse as the grid spacing goes to zero, except in the one-dimensional exponential covariance case.

1. Introduction

The Matérn covariance between two points in $\mathbb{R}^d$ separated by lag $h \in \mathbb{R}^d$ is

$$M[\,h\,;\,\nu, d\,] = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\alpha\|h\|)^{\nu} K_{\nu}(\alpha\|h\|),$$

where σ^2 is a variance parameter, α is an inverse range parameter, ν is a smoothness parameter, and K_ν is the modified Bessel function of the second kind. Guttorp and Gneiting (2006) provide a summary of its important properties and a detailed discussion of its history. Our article presents a theoretical and numerical study of properties of the spectral density of the Matérn covariance when aliased to regular grids of points in one and two dimensions. We apply our results to study a popular approximation to the inverse of Matérn covariance matrices that is motivated by connections between the Matérn covariance and a class of stochastic partial differential equations (SPDEs) (Lindgren et al., 2011).

Building on work by Whittle (1954), Whittle (1963), and Besag (1981), Lindgren et al. (2011) proposed that the inverse of Matérn covariance matrices can be represented by sparse matrices whenever $\nu + d/2$ is an integer, which is why our notation for M includes ν and d. The resulting approximation is commonly referred to as the SPDE approximation. In this paper, we use the terms "SPDE approach" and "SPDE approximation" to refer specifically to the methods in Lindgren et al. (2011). We investigate the sparsity of Matérn inverses and find that there is nothing particularly special with regards to sparsity about the $d = 1, \nu = 3/2$ case or the $d = 2, \nu = 1$, relative to other values of ν. Further, by studying the spectral densities implied by the SPDE approximation, we show that the SPDE over-approximates power at the highest frequencies by a factor of 3 in the $d = 1$, $\nu = 3/2$ case and by as much as a factor of 2.7 in the $d = 2$, $\nu = 1$ case.

In the discussion of Lindgren et al. (2011), Lee and Kaufman noted that the likelihood implied by the SPDE approximation overestimates spatial range parameters. The discussion of the bias was centered on boundary effects. Though boundary effects are important for approximations to the inverse covariance matrix, the present paper suggests instead that the overestimation stems from the fact that the SPDE approximation has too much power at the highest frequencies, causing the likelihood to select a larger range parameter in order to compensate. Our results also suggest an explanation for why

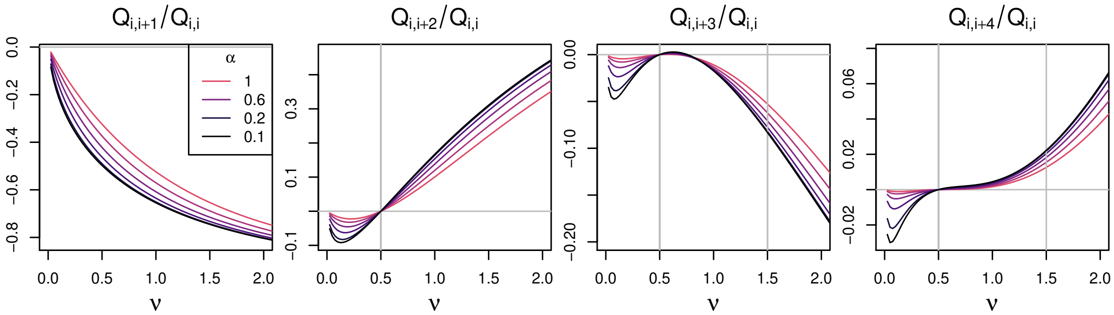

Fig. 1. For dimension $d = 1$, $Q_{i,i+h}/Q_{i,i}$ as a function of ν for various values of h and several inverse range parameters α. Vertical lines indicate values set to zero in the SPDE approximation.

Guinness (2018) found that SPDE approximations were less accurate in terms of KL-divergence than Vecchia's approximation (Vecchia, 1988). The supplementary material of the present paper contains a simulation study that corroborates Lee and Kaufman's results.

To set the stage, consider a Matérn covariance matrix Σ for a large grid of points in one dimension with spacing 1, ordered left to right. Then let $Q = \Sigma^{-1}$, and consider the values $Q_{i,i+h}/Q_{i,i}$, where i is the index for a location near the center of the domain. Figure 1 plots these values for $h = 1, 2, 3$, and 4 over a range of smoothness and inverse range parameters. When $\nu = 0.5$, the values are zero for $h > 1$, and they generally appear to converge as the inverse range decreases, but not to zero when $\nu = 3/2$, which is the approximation used in the SPDE approach. The supplement contains analogous numerical results for the two-dimensional case. The rest of the present paper aims to explore properties of the Matérn model and the SPDE approximation, with an aim of understanding these numerical results.

2. Background

The details for all derivations in this section are spelled out in the supplementary material. Let $Y : \mathbb{R}^d \to \mathbb{R}$ be a stationary process with autocovariance function $A[\,h\,] = \mathrm{Cov}\{Y[\,x+h\,], Y[\,x\,]\}$. Due to Bochner's theorem (cf. Stein, 1999), $A[\,]$ is positive definite when

$$\int_{\mathbb{R}^d} A[\,h\,] e^{-i2\pi\omega\cdot h} dh := A(\omega) > 0 \quad \text{for all } \omega \in \mathbb{R}^d.$$

We call $A()$ the spectral density for $A[\,]$. Our notational convention uses the same letter for the spectral density and covariance function, distinguishing the two with the type of bracket: round for spectral densities and square for covariances. For $\Delta > 0$, define the interval $\mathbb{T}_\Delta = [0, 1/\Delta]$ and hypercube $\mathbb{T}_\Delta^d$. When $h \in \mathbb{Z}^d$, the inverse Fourier transform can be rewritten as

$$A[\,\Delta h\,] = \int_{\mathbb{T}_\Delta^d} \sum_{k \in \mathbb{Z}^d} A(\omega + k/\Delta) e^{i2\pi\Delta\omega\cdot h} d\omega =: A_\Delta[\,h\,],$$

which uses the aliasing property of complex exponentials and introduces a notation $A_\Delta[\,]:\mathbb{Z}^d\to\mathbb{R}$ for covariances on a grid of points with spacing Δ. We define

$$A_\Delta(\omega)=\sum_{k\in\mathbb{Z}^d}A(\omega+k/\Delta)$$

to be the aliased spectral density for A on a grid with spacing Δ. The discrete covariances and the aliased spectral density are related via

$$A_\Delta[\,h\,]=\int_{\mathbb{T}^d_\Delta}A_\Delta(\omega)e^{i2\pi\Delta\omega\cdot h}d\omega,\quad A_\Delta(\omega)=\Delta^d\sum_{h\in\mathbb{Z}^d}A_\Delta[\,h\,]e^{-i2\pi\Delta\omega\cdot h},$$

so that $A_\Delta[\,]$ is the integral Fourier transform of $A_\Delta()$ over $\mathbb{T}^d_\Delta$, and $A_\Delta()$ is the infinite discrete Fourier transform of $A_\Delta[\,]$. We say that $A_\Delta^{-1}[\,]$ is the inverse of $A_\Delta[\,]$ if

$$\Delta^d\sum_{k\in\mathbb{Z}^d}A_\Delta[\,h-k\,]A_\Delta^{-1}[\,k\,]=\mathbb{1}[\,h\,].\tag{1}$$

where $\mathbb{1}[\,h\,]=1$ when $h=0$ and 0 otherwise. We call $A_\Delta^{-1}[\,]$ the inverse operator. Taking the infinite discrete Fourier transform of both sides of (1) reveals that

$$A_\Delta(\omega)A_\Delta^{-1}(\omega)=\Delta^d,$$

meaning that the spectrum of A_Δ^{-1} is the Δ^d times the reciprocal of the spectrum of A_Δ.

3. Matérn Covariances and SPDE Approximations

3.1. *General Representation*

The stationary Matérn covariance function is

$$M[\,h\,;\,\nu,d\,]=\frac{\sigma^2(\alpha\|h\|)^\nu K_\nu(\alpha\|h\|)}{\Gamma(\nu)2^{\nu-1}}=\int_{\mathbb{R}^d}\frac{\sigma^2N_{\alpha,\nu,d}}{\left(\alpha^2+4\pi^2\|\omega\|^2\right)^{\nu+d/2}}e^{i2\pi\omega\cdot h}d\omega,$$

where $N_{\alpha,\nu,d}=2^d\pi^{d/2}\alpha^{2\nu}\Gamma(\nu+d/2)/\Gamma(\nu)$ is a normalizing constant (Williams and Rasmussen, 2006). The aliased spectral density is

$$M_\Delta(\omega\,;\,\nu,d)=\sum_{k\in\mathbb{Z}^d}\sigma^2N_{\alpha,\nu,d}\left(\alpha^2+4\pi^2\|\omega+k/\Delta\|^2\right)^{-\nu-d/2}.$$

3.2. *One Dimension, $\nu=1/2$*

From here on, we set $\sigma^2=1$ to simplify the expressions. When $d=1$ and $\nu=1/2$, the aliased spectral density has the closed form

$$M_\Delta(\omega\,;\,1/2,1)=\Delta\frac{1-e^{-2\Delta\alpha}}{1+e^{-2\Delta\alpha}-e^{-\Delta\alpha}e^{-i\omega2\pi\Delta}-e^{-\Delta\alpha}e^{+i\omega2\pi\Delta}},\tag{2}$$

which can be proven by taking the discrete Fourier transform of the covariance function. The inverse spectral density is Δ times the reciprocal,

$$M_\Delta^{-1}(\omega\,;\,1/2,1)=\frac{1+e^{-2\Delta\alpha}-e^{-\Delta\alpha}e^{-i\omega2\pi\Delta}-e^{-\Delta\alpha}e^{+i\omega2\pi\Delta}}{1-e^{-2\Delta\alpha}},$$

and thus the inverse operator is

$$M_\Delta^{-1}[\,h\,;\,1/2,1\,] = \begin{cases} (1+e^{-2\Delta\alpha})/(1-e^{-2\Delta\alpha}) & h=0, \\ -e^{-\Delta\alpha}/(1-e^{-2\Delta\alpha}) & |h|=1, \\ 0 & |h|>1. \end{cases}$$

The SPDE approximation for the inverse operator in Lindgren et al. (2011) is

$$\widetilde{M}_\Delta^{-1}[\,h\,;\,1/2,1\,] = \begin{cases} \frac{\alpha\Delta}{2} + \frac{1}{\alpha\Delta} & h\ =0 \\ -\frac{1}{2\alpha\Delta} & |h|=1 \\ 0 & |h|>1. \end{cases}$$

which corresponds to spectral density

$$\widetilde{M}_\Delta(\omega\,;\,1/2,1) = \Delta\Big(\frac{\alpha\Delta}{2} + \frac{1}{\alpha\Delta} - \frac{1}{2\alpha\Delta}e^{-i\omega 2\pi\Delta} - \frac{1}{2\alpha\Delta}e^{+i\omega 2\pi\Delta}\Big)^{-1}.$$

Our first theorem establishes that the true and SPDE spectral densities for $\nu = 1/2$ converge to the same values at frequencies 0 and $1/(2\Delta)$ for small $\alpha\Delta$.

THEOREM 1.

$$\frac{M_\Delta(0\,;\,1/2,1)}{2/\alpha} = 1 + O(\alpha^2\Delta^2) \qquad \frac{M_\Delta\big(\frac{1}{2\Delta}\,;\,1/2,1\big)}{2/\alpha} = \frac{\alpha^2\Delta^2}{4} + O(\alpha^4\Delta^4)$$

$$\frac{\widetilde{M}_\Delta(0\,;\,1/2,1)}{2/\alpha} = 1 \qquad \frac{\widetilde{M}_\Delta\big(\frac{1}{2\Delta}\,;\,1/2,1\big)}{2/\alpha} = \frac{\alpha^2\Delta^2}{4} + O(\alpha^4\Delta^4).$$

This provides evidence that the SPDE approximation for $\nu = 1/2$, $d = 1$ is a good approximation to the true model when $\alpha\Delta$ is small; their spectral densities are similar at the lowest frequency ($\omega = 0$) when the power is greatest and at the highest frequency ($\omega = \Delta^{-1}/2$) when the power is smallest, implying that the both the SPDE spectral density and its reciprocal may be good approximations to the truth, which in turn implies that both the covariance operator and its inverse may be good approximations.

3.3. *One Dimension, $\nu = 3/2$*

When $\nu = 3/2$, the aliased spectral density is

$$M_\Delta(\omega\,;\,3/2,1) = \sum_{k\in\mathbb{Z}} \frac{4\alpha^3}{[\alpha^2 + 4\pi^2(\omega + k/\Delta)^2]^2}. \tag{3}$$

The SPDE approximation in Lindgren et al. (2011) to the inverse operator is simply the convolution of the $\nu = 1/2$ approximation

$$(\alpha\Delta)\widetilde{M}_\Delta^{-1}[\,h\,;\,3/2,1\,] = \begin{cases} \Big(\frac{\alpha\Delta}{2} + \frac{1}{\alpha\Delta}\Big)^2 + \frac{1}{2\alpha^2\Delta^2} & h\ =0 \\ -\frac{1}{2} - \frac{1}{\alpha^2\Delta^2} & |h|=1 \\ \frac{1}{4\alpha^2\Delta^2} & |h|=2 \\ 0 & |h|>2 \end{cases},$$

which means that the spectral density for the $\nu = 3/2$ SPDE inverse operator is simply the square of spectral density for the $\nu = 1/2$ SPDE inverse operator,

$$\widetilde{M}_\Delta^{-1}(\omega\,;\,3/2,1) = \frac{1}{\alpha\Delta}\Big(\frac{\alpha\Delta}{2} + \frac{1}{\alpha\Delta} - \frac{1}{2\alpha\Delta}e^{-i\omega 2\pi\Delta} - \frac{1}{2\alpha\Delta}e^{+i\omega 2\pi\Delta}\Big)^2,$$

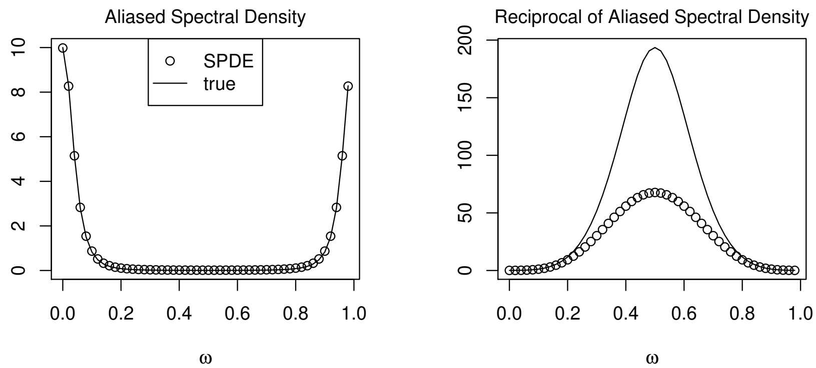

Fig. 2. For $d = 1$, $\nu = 3/2$, $\alpha = 0.4$, true aliased spectral density and its reciprocal (lines) and SPDE approximation to the aliased spectral density and its reciprocal (circles).

and the spectral density for the $\nu = 3/2$ SPDE covariance operator is

$$\widetilde{M}_\Delta(\omega\,;\,3/2,1) = \alpha\Delta^2\Big(\frac{\alpha\Delta}{2} + \frac{1}{\alpha\Delta} - \frac{1}{2\alpha\Delta}e^{-i\omega 2\pi\Delta} - \frac{1}{2\alpha\Delta}e^{+i\omega 2\pi\Delta}\Big)^{-2}.$$

Normalizing constants are chosen so that $M_\Delta(0\,;\,3/2,1) \to \widetilde{M}_\Delta(0\,;\,3/2,1)$ as $\alpha\Delta \to 0$.

Note that the aliased Matérn spectral density for $\nu = 3/2$ in (3) is not simply the square of the aliased $\nu = 1/2$ spectral density in (2); rather, we alias the square of the unaliased $\nu = 1/2$ spectral density. The SPDE appproximation reverses the order of operations, squaring the aliased spectral density. This subtle difference leads to the SPDE approximation assigning too much power at the highest frequencies, made explicit in the following theorem:

Theorem 2.

$$\frac{M_\Delta(0\,;\,3/2,1)}{4/\alpha} = 1 + O(\alpha^4\Delta^4) \qquad \frac{M_\Delta\big(\frac{1}{2\Delta}\,;\,3/2,1\big)}{4/\alpha} = \frac{\alpha^4\Delta^4}{48} + O(\alpha^6\Delta^6)$$

$$\frac{\widetilde{M}_\Delta(0\,;\,3/2,1)}{4/\alpha} = 1 \qquad \frac{\widetilde{M}_\Delta\big(\frac{1}{2\Delta}\,;\,3/2,1\big)}{4/\alpha} = \frac{\alpha^4\Delta^4}{16} + O(\alpha^6\Delta^6).$$

When scaled by $4/\alpha$, both spectral densities converge to 1 when $\omega = 0$ and $\alpha\Delta \to 0$, but they converge to two different values, $\alpha^4\Delta^4/48$ and $\alpha^4\Delta^4/16$, when $\omega = \Delta^{-1}/2$, meaning that the SPDE spectral density assigns three times too much power at the highest frequency. The inaccuracy of the spectral density at high frequencies impacts the quality of the approximation to the reciprocal of the spectral density, seen in Figure 2, and to the inverse operator, as evidenced in Figure 1.

3.4. *Two Dimensions*

The aliased spectral density for the Matérn in two dimensions is

$$M_\Delta(\omega\,;\,\nu,2) = 4\pi\alpha^2 \sum_{k\in\mathbb{Z}^2} \big[\alpha^2 + 4\pi^2(\omega_1 + k_1/\Delta)^2 + 4\pi^2(\omega_2 + k_2/\Delta)^2\big]^{-\nu-1}.$$

The following theorem establishes properties of the aliased Matérn spectral density for $\nu = 1$ at the lowest frequency and at high frequencies in one and both spatial dimensions.

THEOREM 3.

$$\frac{M_\Delta((0,0)\,;\,1,2)}{4\pi/\alpha^2} = 1 + \frac{\alpha^4\Delta^4}{258.6} + O(\alpha^6\Delta^6)$$
$$\frac{M_\Delta\big((\frac{1}{2\Delta},0)\,;\,1,2\big)}{4\pi/\alpha^2} = \frac{\alpha^4\Delta^4}{43.10} + O(\alpha^6\Delta^6)$$
$$\frac{M_\Delta\big((\frac{1}{2\Delta},\frac{1}{2\Delta})\,;\,1,2\big)}{4\pi/\alpha^2} = \frac{\alpha^4\Delta^4}{86.20} + O(\alpha^6\Delta^6)$$

The numbers 258.6, 43.10, and 86.20 are the result of numerical calculations and are rounded to one or two decimals. They are available to higher accuracy. Details are given in the proof in the supplementary material. The SPDE approximation to the inverse operator is

$$4\pi(\alpha\Delta)^2\widetilde{M}_\Delta^{-1}[\,h\,;\,1,2\,] = \begin{cases} (4+\alpha^2\Delta^2)^2+4 & h=(0,0) \\ -2(4+\alpha^2\Delta^2) & h=(0,1),(0,-1),(1,0),(-1,0) \\ 2 & h=(1,1),(1,-1),(-1,1),(-1,-1) \\ 1 & h=(2,0),(0,2),(-2,0),(0,-2) \\ 0 & \text{otherwise,} \end{cases}$$

which corresponds to the spectral density

$$\widetilde{M}_\Delta(\omega\,;\,1,2) = 4\pi\alpha^2\Delta^4\Big(4+\alpha^2\Delta^2 - e^{i2\pi\Delta\omega_1} - e^{-i2\pi\Delta\omega_1} - e^{i2\pi\Delta\omega_2} - e^{-i2\pi\Delta\omega_2}\Big)^{-2}.$$

The normalizing constants are chosen so that $M_\Delta((0,0)\,;\,1,2) \to \widetilde{M}_\Delta((0,0)\,;\,1,2) = 4\pi/\alpha^2$ as $\alpha\Delta \to 0$. The following theorem establishes the behavior of $\widetilde{M}_\Delta(\omega\,;\,1,2)$ at the same frequencies in Theorem 3.

THEOREM 4.

$$\frac{\widetilde{M}_\Delta((0,0)\,;\,1,2)}{4\pi/\alpha^2} = 1$$
$$\frac{\widetilde{M}_\Delta\big((\frac{1}{2\Delta},0)\,;\,1,2\big)}{4\pi/\alpha^2} = \frac{\alpha^4\Delta^4}{16} + O(\alpha^6\Delta^6)$$
$$\frac{\widetilde{M}_\Delta\big((\frac{1}{2\Delta},\frac{1}{2\Delta})\,;\,1,2\big)}{4\pi/\alpha^2} = \frac{\alpha^4\Delta^4}{64} + O(\alpha^6\Delta^6)$$

Theorems 3 and 4 imply that when $\alpha\Delta$ is small, the SPDE approach over-approximates the spectral density by a factor of $43.1/16 = 2.69$ at $\omega = (\Delta^{-1}/2, 0)$ and $86.2/64 = 1.35$ at $\omega = (\Delta^{-1}/2, \Delta^{-1}/2)$. Figure 3 contains an example where the ratio between the SPDE spectral density and the true spectral density varies between 0.999 at $\omega = (0,0)$, 2.680 at $\omega = (\Delta^{-1}/2, 0)$ and 1.345 at $\omega = (\Delta^{-1}/2, \Delta^{-1}/2)$.

4. DISCUSSION

The SPDE approximation has proven useful as a computational tool and as a conceptual tool for defining extensions of the stationary gridded model to models for irregularly-

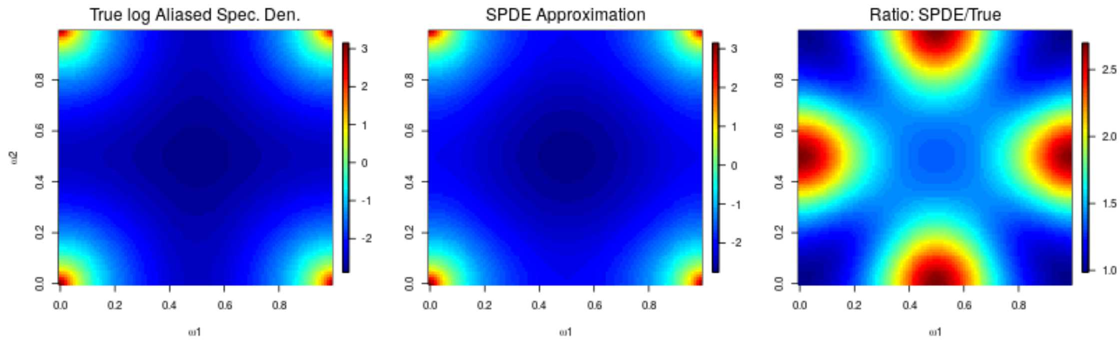

Fig. 3. True spectral density for $\nu = 1$, $\alpha = 0.5$, SPDE approximation to the spectral density, and the ratio of the two. The ratio is near 1.00 at $(0, 0)$, near 2.69 at $(1/2\Delta, 0)$ and near 1.35 at $(1/2\Delta, 1/2\Delta)$, as predicted by the theory.

spaced data, models on manifolds, and to non-stationary models (Fuglstad et al., 2015; Bakka et al., 2018). This paper does not question the usefulness of the SPDE approach as a tool for data analysis. Rather, it is a study of the SPDE approximation to Matérn models on grids. The supplementary material explores the implications of our results, namely a simulation study showing that the SPDE approximation overestimates range parameters. The supplementary material also includes numerical results for the true inverse operator in the one- and two-dimensional cases.

We study SPDE approximations to Matérn fields observed at point locations on a grid, as opposed to observations of gridbox averages. While SPDE approximations have been applied in both cases, the spectral properties of gridbox average fields are different, and it is not clear to the author whether the SPDE approximations would be more or less accurate in the gridbox average case. This is certainly an interesting question worthy of future study. In addition, it would be interesting to explore extensions to irregularly sampled locations. In all cases, a study of the impact of the approximations on predictions is also warranted. It seems plausible that if the approximate model is used for both inferring parameters and generating predictions, the resulting predictions would be reasonably accurate.

Supplementary Material

The appendices contain additional numerical and simulation studies, background material, and proofs. R code for reproducing all numerical results and figures has been uploaded as online supplementary material.

Acknowledgements

This work was supported by the National Science Foundation under grant No. 1916208 and the National Institutes of Health under grant No. R01ES027892. The author is grateful to have received helpful feedback on an early draft from Matthias Katzfuss, Michael Stein, and Ethan Anderes, and from an Associate Editor and two anonymous reviewers, one of which pointed out an mistake in the first version.

References

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6):e1443.

Besag, J. (1981). On a system of two-dimensional recurrence equations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43(3):302–309.

Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2015). Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. *Statistica Sinica*, 25:115–133.

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

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

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Stein, M. L. (1999). *Interpolation of spatial data: some theory for Kriging*. Springer.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):297–312.

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

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

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