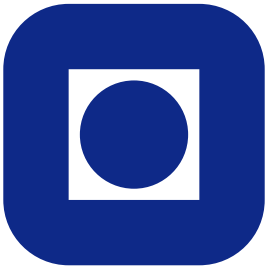
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach

by

Finn Lindgren, Johan Lindström and Håvard Rue

PREPRINT
STATISTICS NO. 5/2010



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL http://www.math.ntnu.no/preprint/statistics/2010/S5-2010.pdf
Håvard Rue has homepage: http://www.math.ntnu.no/∼hrue
E-mail: hrue@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach

Finn Lindgren & Johan Lindström
Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Sweden

Håvard Rue
Department of Mathematical Sciences
Norwegian University of Science and Technology
NTNU, Norway

March 25, 2010

Abstract

Continuously indexed Gaussian fields (GFs) is the most important ingredient in spatial statistical modelling and geo-statistics. The specification through the covariance function gives an intuitive interpretation of its properties. On the computational side, GFs are hampered with the *big-n* problem, since the cost of factorising dense matrices is cubic in the dimension. Although the computational power today is all-time-high, this fact seems still to be a computational bottleneck in applications. Along with GFs, there is the class of Gaussian Markov random fields (GMRFs) which are discretely indexed. The Markov property makes the involved precision matrix sparse which enables the use of numerical algorithms for sparse matrices, that for fields in $\mathbb{R}^2$ only use the square-root of the time required by general algorithms. The specification of a GMRF is through its full conditional distributions but its marginal properties are not transparent in such a parametrisation.

In this paper, we show that using an approximate stochastic weak solution to (linear) stochastic partial differential equations (SPDEs), we can, for some GFs in the Matérn class, provide an *explicit link*, for any triangulation of $\mathbb{R}^d$, between GFs and GMRFs. The consequence is that we can take the best from the two worlds and do the modelling using GFs but do the computations using GMRFs. Perhaps more importantly, our approach generalises to other covariance functions generated by SPDEs, including oscillating and non-stationary GFs, as well as GFs on manifolds. We illustrate our approach by analysing global temperature data with a non-stationary model defined on a sphere.

ACKNOWLEDGEMENT: The authors are listed in alphabetical order.

KEYWORDS: Approximate Bayesian inference, Gaussian Markov random fields, Generalised additive mixed models, Laplace approximation, Sparse matrices, Parallel computing, Stochastic partial differential equations, Structured additive regression models.

AMS SUBJECT CLASSIFICATION:: Primary 62F15; secondary 62H99

ADDRESS FOR CORRESPONDENCE: H. Rue, Department of Mathematical Sciences, The Norwegian University for Science and Technology, N-7491 Trondheim, Norway. Email: hrue@math.ntnu.no, WWW-address: http://www.math.ntnu.no/∼hrue, Voice: +47-7359-3533, Fax: +47-7359-3524.

1 Introduction

Gaussian fields (GFs) has a dominant role in spatial statistics and especially in the traditional field of geostatistics (Chilés and Delfiner, 1999; Cressie, 1993; Diggle and Ribeiro, 2006; Stein, 1999), and forms an important building block in modern hierarchical spatial models (Banerjee et al., 2004). GFs is one of a few appropriate multivariate models with an explicit and computable normalising constant and has otherwise good analytic properties. In a domain $\mathcal{D} \in \mathbb{R}^d$ with coordinate $\boldsymbol{s} \in \mathcal{D}$, $x(\boldsymbol{s})$ a continuously indexed GF if all finite collections $\{x(\boldsymbol{s}_i)\}$ are jointly Gaussian distributed. In most cases, the Gaussian field is specified using a mean function $\mu(\cdot)$ and a covariance function $\mathrm{C}(\cdot,\cdot)$, so the mean is $\boldsymbol{\mu} = (\mu(\boldsymbol{s}_i))$ and the covariance matrix is $\boldsymbol{\Sigma} = (\mathrm{C}(\boldsymbol{s}_i, \boldsymbol{s}_j))$. Often the covariance function is only a function of the Euclidean distance between two locations in which the deviation from the mean is said to be isotropic, and it is stationary if the covariance function only depends on any distance measure between two locations. Since the covariance matrix is positive definite, the covariance function must be a positive definite function. This restriction makes it less easy to "invent" analytical covariance functions. Bochner's theorem is often used in this context, as it characterises all continuous positive definite functions in $\mathbb{R}^d$ as the Fourier transforms of non-negative Borel measures.

Although GFs are convenient from both an analytical and a practical point of view, the computational issues has always been a bottleneck. This is due to the general cost of $\mathcal{O}(n^3)$ to factorise dense $n \times n$ (covariance) matrices. Although the computational power today is all-time-high, the tendency seems to be that the dimension n is always set, or we want to set it, a bit higher than the value that gives a reasonable computation time. The increasing popularity of hierarchical Bayesian models has made this issue more important, as "repeated computations (as for simulation-based model fitting) can be very slow, perhaps infeasible" (Banerjee et al., 2004, p.387), and the authors continue with referring to this situation informally as "the big n problem".

There are several approaches trying to overcome or avoid "the big n problem". The spectral approach representation of the likelihood (Whittle, 1954) makes it possible to estimate the (power-)spectrum (using discrete Fourier transforms calculations) and compute the log-likelihood from it (Dahlhaus and Künsch, 1987; Fuentes, 2008; Guyon, 1982) but this is only possible for directly observed stationary GFs on a (near-)regular lattice locations. Stein et al. (2004); Vecchia (1988) propose to use an approximate likelihood constructed using a sequential representation and then simplify the conditioning set, and similar ideas also apply when computing conditional expectation (Kriging). An alternative approach, is to do exact computations on a simplified Gaussian model of low rank (Banerjee et al., 2008; Cressie and Johannesson, 2008). Their difference is essentially how the basis for the low rank approximation is constructed. Furrer et al. (2006) apply covariance tapering to zero-out parts of the covariance matrix to gain computational speedup. However, the sparsity pattern will depend on the range of the GFs, and the potential in a related approach, named "lattice methods" by (Banerjee et al., 2004, A.5.3), is superior to the covariance tapering idea. In this approach the GF is replaced by a Gaussian Markov random field (GMRF); see Rue and Held (2005) for a detailed introduction and Rue et al. (2009, Sec. 2.1) for a condensed review. A GMRF is a (discretely indexed) Gaussian field $\boldsymbol{x}$ where the full conditionals

$$\pi(x_i \mid \boldsymbol{x}_{-i}) = \pi(x_i \mid \boldsymbol{x}_{\partial i}), \quad i = 1, \ldots, n, \tag{1}$$

only depend on a set of neighbours ∂i to each site i (where consistency requirements imply that if $i \in \partial j$ then also $j \in \partial i$). The computational gain comes from the fact that the zero-pattern of the precision matrix $\boldsymbol{Q}$ (the inverse covariance matrix), relates directly to the notion of neighbours,

$$Q_{ij} \neq 0 \quad \Longleftrightarrow \quad i \in \partial j \cup j$$

see for example Rue and Held (2005, Sec 2.2). Algorithms for MCMC will repeatedly update from these simple full conditionals, which explains to a large extent the popularity of GMRFs in recent years, starting already with the seminal papers by J. Besag (Besag, 1974, 1975). However, GMRFs also allow for fast direct numerical algorithms Rue (2001), as numerical factorisation of the matrix $\boldsymbol{Q}$ can be done using sparse matrix algorithms (Davis, 2006; Duff et al., 1989; George and Liu, 1981) at a typical cost of $\mathcal{O}(n^{3/2})$ for two-dimensional GMRFs; see (Rue and Held, 2005) for detailed algorithms. GMRF has very good computational properties and its major importance in Bayesian inferential methods is based on nested integrated Laplace

approximations (INLA) (Rue et al., 2009), which allow for fast and accurate Bayesian inference for structured additive regression models.

Although GMRFs have very good computational properties, there are good reasons for why statistical models based on GMRFs are relatively simple and often applied to area data like regions or counties. First, there is has been no good way to parametrise the precision matrix of a GMRF to achieve a predefined behaviour in terms of correlation between two sites and to control marginal variances. In matrix terms, this is kind of obvious since the covariance matrix is the inverse of the precision matrix. Therefore, often simple approaches are taken, like letting Q_{ij} be related to the reciprocal distance between sites i and j (Arjas and Gasbarra, 1996; Besag et al., 1991; Gschlößl and Czado, 2007; Pettitt et al., 2002; Weir and Pettitt, 2000), however a more detailed analysis shows that such a rationale is suboptimal (Besag and Kooperberg, 1995; Rue and Tjelmeland, 2002) and can give surprising effects (Wall, 2004). Secondly, it is unclear how large the class of useful GMRF models really is using only a simple neighbourhood. A complicating issue here is the global positive definiteness constraint, and it might not be evident how this influences the parametrisation of the full conditionals.

Rue and Tjelmeland (2002) demonstrated empirically, that GMRFs could approximate very well most of the commonly used covariance functions in geo-statistics, and proposed to use them as computational replacements for GFs for computational reasons like doing Kriging (Hartman and Hössjer, 2008). (It must be emphasised, that GFs used in geo-statistics when discretised do not have the desired Markov property we desire for GMRFs, but there seems to be a GMRF with approximately the same covariance function.) However, there were several drawbacks with their approach; First, the fitting of GMRFs to GFs was restricted to a regular lattice (or torus) and the fit itself had to be precomputed for a discrete set of parameter values (like smoothness and range), but the fit is invariant to the lattice size, using a time-consuming numerical optimisation. Despite these 'proof-of-concept' results, several authors have followed up this idea without any large progress in the methodology (Cressie and Verzelen, 2008; Hrafnkelsson and Cressie, 2003; Song et al., 2008), but the approach itself has shown useful even for spatio-temporal models (Allcroft and Glasbey, 2003).

The discussion so far has revealed an seemingly optimal modelling/computational strategy for approaching the "big n-problem" is a good way:

1. Do the modelling using a GF on a set of locations $\{\boldsymbol{s}_i\}$, to construct a discretised GF with covariance matrix $\boldsymbol{\Sigma}$

2. Find a GMRF with local neighbourhood and precision matrix $\boldsymbol{Q}$ that *represents* the GF in the best possible way; i.e. $\boldsymbol{Q}^{-1}$ is close to $\boldsymbol{\Sigma}$ in some norm. (We deliberately use the phrase "represents" instead of approximate.)

3. Do the computations using the GMRF representation using numerical methods for sparse matrices.

Such an approach relies on several assumptions. First the GF must be of such a type that there exists a GMRF with local neighbourhood that can represent it sufficiently accurate in order to keep the interpretation of the parameters and the results. Secondly, we must be able to compute the GMRF representation from the GF, at any collections of locations, so fast, that we still achieve a considerable speedup compared to treating the GF directly.

The purpose of this paper, is to demonstrate that these requirements can indeed be met for certain members of GF with the Matérn covariance function in $\mathbb{R}^d$, where the GMRF representation is available explicitly. Although these results are seemingly restrictive at first sight, they do cover the most important and used covariance model in spatial statistics; see also Stein (1999, p.14) which concluded his detailed theoretical analysis with "Use the Matérn model." The GMRF representation can be computed using a certain stochastic partial differential equation (SPDE) which has GFs with Matérn covariance function as the solution when driven by Gaussian white noise.

Rather surprisingly, extending this basic result seems to open new doors and opportunities, and provide rather simple answers to rather difficult modelling problems. We will show how we can treat some GFs beyond the Matérn covariances, how to use this approach to define GFs on curved spaces/manifold and how

to represent non-stationary GFs. Our basic task, to do the modelling using GF and the computations using the GMRF representation, still holds for these extensions as the GMRF representation is still available explicitly.

The plan of rest of this paper is as follows. In Section 2, we discuss the relationship between Matérn covariances and a specific stochastic partial differential equation, and present the two main results for explicitly constructing the precisions of GMRFs based on this relationship. In Section 3, the results are extended to fields on triangulated manifolds, as well as non-stationary and oscillating models. The extensions are illustrated with a non-stationary analysis of global temperature data in Section 4, and we conclude the main part of the paper with a brief discussion in Section 5. Thereafter follows five technical appendices, with explicit representation results (A), theory for random fields on manifolds (B), the Hilbert space representation details (C), proofs of the technical details (D), and theory for spherical harmonics (E) needed for models on a globe.

2 Preliminaries and main results

This section will introduce the Matérn covariance model and discuss its representation through a linear SPDE. We will state explicit results for the GMRF representation of Matérn fields on a regular lattice and do an informal summary of the main results.

2.1 The Matérn covariance model and its SPDE

Let $\|\cdot\|$ denote the Euclidean distance in $\mathbb{R}^d$. The Matérn covariance function between locations $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$, is defined as

$$r(\boldsymbol{u}, \boldsymbol{v}) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa\|\boldsymbol{v}-\boldsymbol{u}\|)^\nu K_\nu(\kappa\|\boldsymbol{v}-\boldsymbol{u}\|). \tag{2}$$

Here, K_ν is the modified Bessel function of second kind and order $\nu > 0$, $\kappa > 0$ is a scaling parameter and σ^2 is the marginal variance. The integer value of ν determines the mean square differentiability of the underlying process, which matters for predictions made using such a model. However, ν is usually fixed since it is poorly identified in typically applications. A more natural interpretation of the scaling parameter κ is as a range parameter ρ; the Euclidean distance where $x(\boldsymbol{u})$ and $x(\boldsymbol{v})$ is almost independent. In lack of a simple relationship, we will throughout this paper use the definition $\rho = \sqrt{8\nu}/\kappa$, corresponding to correlations near 0.1 at the distance ρ, for all ν.

The Matérn covariance function appears naturally in a number of scientific fields (Guttorp and Gneiting, 2006), but the important relationship that we will make use of is that the solution $x(\boldsymbol{u})$, of the following (fractional) stochastic partial differential equation (SPDE)

$$(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}), \quad \boldsymbol{u} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{3}$$

is a Gaussian field with the Matérn covariance (Whittle, 1954, 1963). The innovation process $\mathcal{W}$ is (spatial) Gaussian white noise with unit variance, Δ is the Laplace operator (*Laplacian*)

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} \tag{4}$$

and the marginal variance is

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}}. \tag{5}$$

We will name any solution to (3) *a Matérn field* in the following. However, the limiting solutions to the SPDE (3) as $\kappa \longrightarrow 0$ or $\nu \longrightarrow 0$ do not have Matérn covariance functions, but the SPDE still has solutions when $\kappa = 0$ or $\nu = 0$ which are well-defined random measures. We will return to this issue in Section C.3. Further, there is an implicit assumption of "proper" boundary conditions for the SPDE, as for $\alpha \geq 2$ then $\exp(\kappa \boldsymbol{e}^\mathsf{T}\boldsymbol{u})$ is in the null-space of the differential operator, for all $\|\boldsymbol{e}\| = 1$.

The proof of Whittle (1954, 1963), is to show that the wave-number spectrum of the solution is

$$R(\boldsymbol{k}) = \frac{(2\pi)^{-d}}{\left(\kappa^2 + \boldsymbol{k}^\mathsf{T}\boldsymbol{k}\right)^\alpha} \tag{6}$$

using the Fourier transform definition of the fractional Laplacian in $\mathbb{R}^d$

$$(\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2}\phi)(\boldsymbol{k}) = (\kappa^2 + \boldsymbol{k}^\mathsf{T}\boldsymbol{k})^{\alpha/2}(\mathcal{F}\phi)(\boldsymbol{k}) \tag{7}$$

where ϕ is a function on $\mathbb{R}^d$ for which the right-hand side of the definition has a well-defined inverse Fourier transform.

2.2 Main results

This section contains our main results, however in a loose and imprecise form. In the Appendices, our statements are made precise and the proofs are given. In the discussion we will restrict us to dimension $d = 2$ although our results are general.

2.2.1 Main result 1

For our first result, we will use some hand-waving arguments and a simple but powerful consequence of a (partly) analytic result of Besag (1981). We will later (in the Appendices) show that these results are true. Let $\boldsymbol{x}$ be a GMRF on a regular (tending to infinite) two-dimensional lattice indexed by ij, where the Gaussian full conditionals are

$$\mathrm{E}(x_{ij} \mid \boldsymbol{x}_{-ij}) = \frac{1}{a}\left(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}\right), \qquad \mathrm{Var}(x_{ij} \mid \boldsymbol{x}_{-ij}) = 1/a \tag{8}$$

and $|a| > 4$. To simplify later notation, we will write this particular model as

$$\begin{array}{|lll} & -1 & \\ a & & -1 \end{array} \tag{9}$$

where symmetry is required to interpret this model. The approximate result (Besag, 1981, Eq. (14)) is that

$$\mathrm{Cov}(x_{ij}, x_{i'j'}) \approx \frac{a}{2\pi} K_0(l\sqrt{a-4}), \quad l \neq 0 \tag{10}$$

where l is the Euclidean distance between ij and $i'j'$. Comparing with (2), we find that $\kappa^2 = a - 4$, $\nu = 0$ and $\sigma^2 = a/(4\pi)$, even though (2) requires $\nu > 0$. Informally, this means that the discrete model defined by (8) generates approximate solutions to the SPDE in (3) on a unit-distance regular grid, with $\nu = 0$.

Passing Gaussian noise through (3) for $\alpha = 1$ gives the spectrum

$$R_1 \propto \frac{1}{(a-4) + \boldsymbol{k}^\mathsf{T}\boldsymbol{k}} \tag{11}$$

meaning that (some discretised version of) the SPDE acts like a linear filter with squared transfer-function equal to R_1. Passing Gaussian noise with spectrum R_1 into the same filter produces output with spectrum $R_2 = R_1^2$ and so on. The consequence is GMRF representations for the Matérn fields for $\nu = 1$ and $\nu = 2$, as convolutions of the coefficients in (9),

$$\nu = 1: \quad \begin{array}{|lll} & 1 & \\ -2a & & 2 \\ 4 + a^2 & -2a & 1 \end{array} \qquad \nu = 2: \quad \begin{array}{|llll} & -1 & & \\ 3a & & -3 & \\ -3(a^2+3) & & 6a & -3 \\ a(a^2+12) & -3(a^2+3) & 3a & -1 \end{array} \tag{12}$$

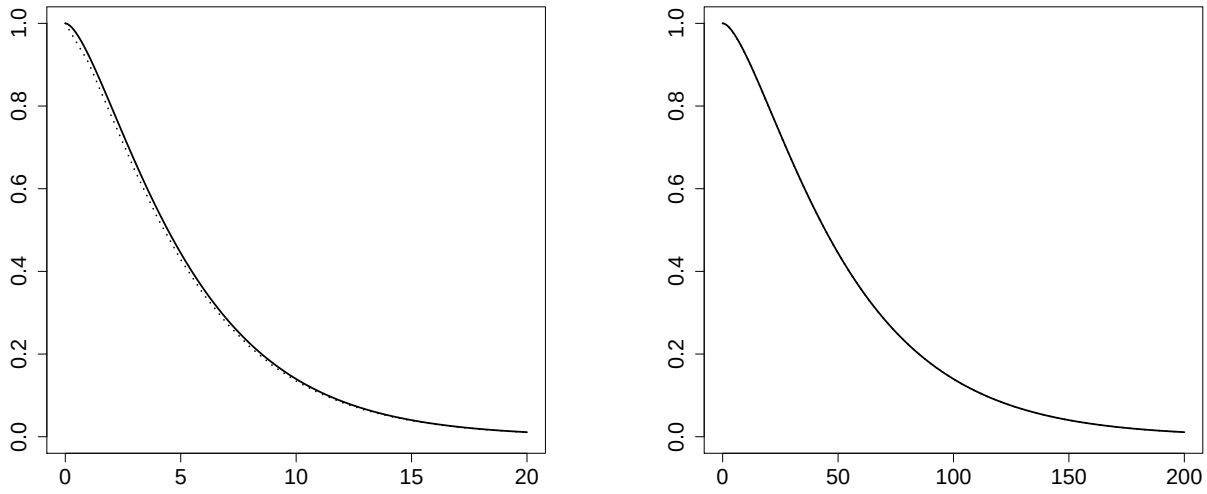


Figure 1: The Matérn correlations (solid line) for range 10 (left) and 100 (right), and the correlations for the GMRF representation (dotted line).

The marginal variance is $1/(4\pi\nu(a-4)^\nu)$. Figure 1 shows how accurate these approximations are for $\nu = 1$ and range 10 and 100, displaying the Matérn correlations and the linearly interpolated correlations for integer lags for the GMRF representation. For range 100 the two curves are indistinguishable. The root-mean-square error between correlations up to twice the range, is 0.01 and 0.0003 for range 10 and 100, respectively. The error is the marginal variance is 4% for range 10 and negligible for range 100.

Our first result confirms the above heuristics.

Main Result 1 *The coefficients in the GMRF representation of (3) on a regular unit-distance two-dimensional infinite lattice for $\nu = 1, 2, \ldots$, is found by convolving (9) by itself ν times.*

Simple extensions of this result are also interesting, like anisotropy along one of the main axes, with details in Appendix A. Extensions to irregular lattices are discussed in the next main result.

2.3 Main result 2

Although Result 1 is useful in itself, is not yet fully practical since often one does not want to have a regular grid, to avoid interpolating the locations of observations to the nearest grid-point, and to allow for finer resolution where details are required. We therefore extend the regular grid to irregular grids, by subdividing $\mathbb{R}^2$ into a set of non-intersecting triangles, where any two triangles meet in at most a common edge or corner. The three corners of a triangle are named *vertices*. In most cases we place initial vertices at the locations for the observations, and add additional vertices to satisfy overall soft constraints of the triangles; see for example Edelsbrunner (2001); Hjelle and Dæhlen (2006) for details. In short, to construct such a triangulation is a standard problem in engineering for solving differential equations using finite element methods (Brenner and Scott, 2007; Quarteroni and Valli, 2008), and many good free implementations exists.

To illustrate the process of triangulation of $\mathbb{R}^2$, we will use an example from Henderson et al. (2002) which models spatial variation in leukaemia survival data in Northwest England. Figure 2(a) displays the 1043 locations of 1043 cases of acute myeloid leukaemia in adults who have been diagnosed between 1982 and 1998 in Northwest England. Panel (b) displays the triangulation of the area of interest, using fine resolution around the data locations and rough resolution outside the area of interest. Further, we place vertices at all data locations. The number of vertices in this example is 2535 and the number of triangles is 5054.

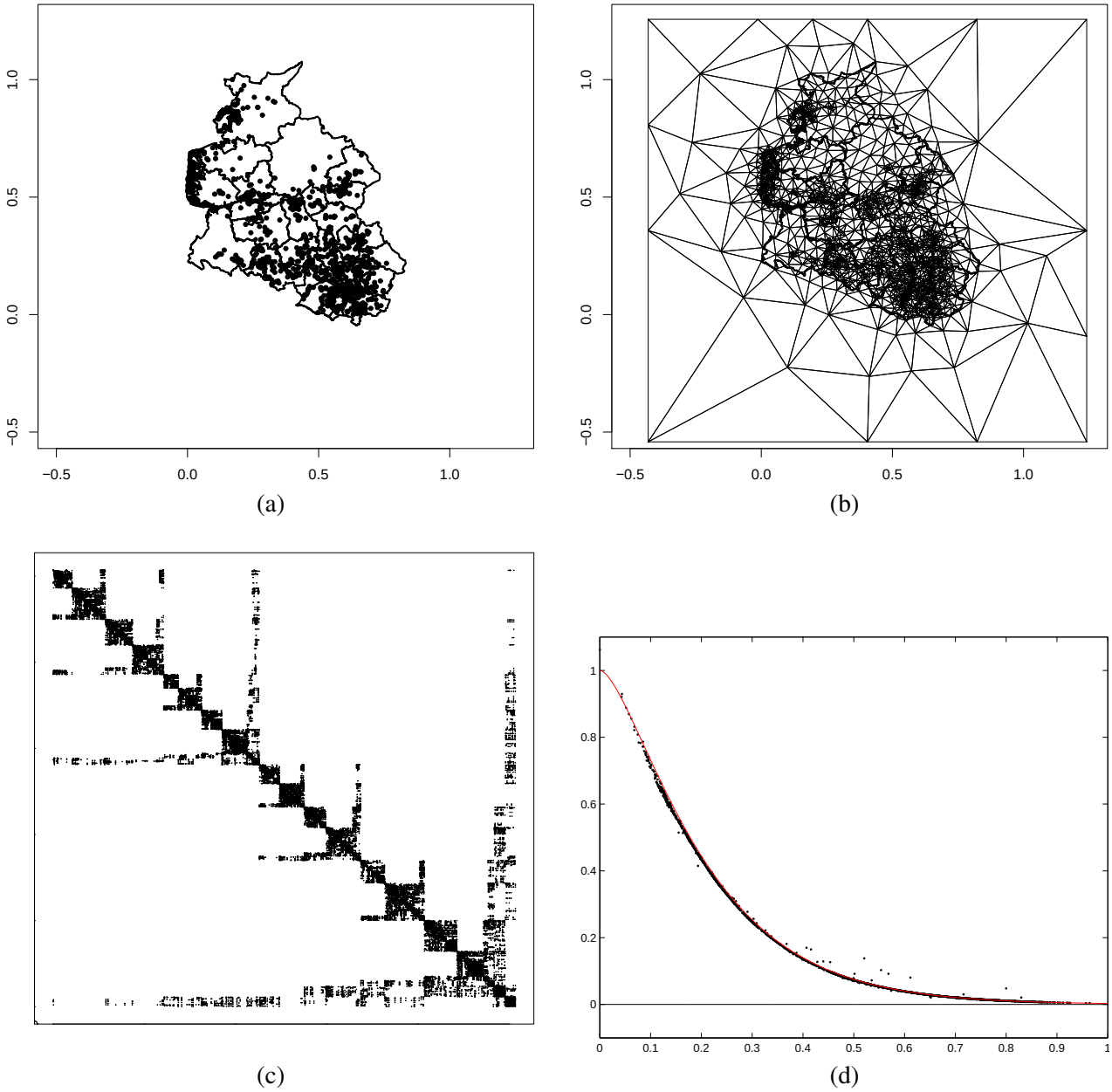


Figure 2: Panel (a) displays the locations of the observations, panel (b) the triangulation using 2535 triangles, panel (c) shows the (reordered) sparse precision matrix for $\nu = 1$, and panel (d) displays the accuracy of the GMRF representation through the numerical and true correlation function.

In order to compute a GMRF representation of the Matérn field on the triangulated lattice, we will compute (an approximation to) a stochastic weak solution of the SPDE (3) (Kleoden and Platen, 1999; Kotelenez, 1999). This part borrows ideas from finite element analysis and starts by *representing* the solution of the SPDE as

$$x(\boldsymbol{u}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{u}) w_k \tag{13}$$

for some chosen basis-functions $\{\psi_k\}$ and Gaussian distributed weights $\{w_k\}$. Here, n is the number of vertices in the triangulation. We chose to use functions ψ_k that are piecewise linear in each triangle, defined such that ψ_k is 1 at vertex k and zero at all other vertices. Another interpretation of (13), is that we try to get the joint distribution of the solution of the SPDE at all vertices and then approximate the continuously indexed solution using local interpolation guided by the triangles. With our choice of basis-functions, the local interpolation is linear.

Define the inner product

$$\langle f, g \rangle = \int f(\boldsymbol{u}) g(\boldsymbol{u}) d\boldsymbol{u} \tag{14}$$

where the integral is over the region of interest. The approximate *stochastic weak solution* of the SPDE is found by requiring that

$$\langle \phi_j, (\kappa^2 - \Delta)^{\alpha/2} x \rangle \stackrel{d}{=} \langle \phi_j, \boldsymbol{\epsilon} \rangle, \qquad j = 1, 2, \ldots$$

for some appropriate set of *test functions* $\{\phi_j(\boldsymbol{u})\}$, where "$\stackrel{d}{=}$" denotes equality in distribution. The choice of test functions, in relation to the basis functions, governs the properties of the approximation.

We now use our representation (13) and choose $\phi_k = (\kappa^2 - \Delta)^{1/2} \psi_k$ for $\alpha = 1$ and $\phi_k = \psi_k$ for $\alpha = 2, 3, 4, \ldots$. These two approximations are denoted the *least squares* approximation and the *Galerkin* approximation, respectively. In addition, for $\alpha \geq 3$, we let $\alpha = 2$ in the left-hand side of (13), and replace the right-hand side with a field generated by $\alpha - 2$. In essence, this generates a recursive formulation, terminating in either $\alpha = 1$ or $\alpha = 2$. The background for these choices are given in Appendix C where a proper derivations of the results are given.

Define the $n \times n$-matrices $\boldsymbol{C}$, $\boldsymbol{G}$, and $\boldsymbol{K}$ with entries

$$\begin{aligned} C_{ij} &= \langle \psi_i, \psi_j \rangle & (15) \\ G_{ij} &= \langle \nabla \psi_i, \nabla \psi_j \rangle & (16) \\ K_{ij}(\kappa^2) &= \kappa^2 C_{ij} + G_{ij}. & (17) \end{aligned}$$

Using Neumann boundary conditions (zero normal-derivative at the boundary), we get our second main result, expressed here for $\mathbb{R}^1$ and $\mathbb{R}^2$.

Main Result 2 *Let $\boldsymbol{Q}_\alpha(\kappa^2)$ be the precision matrix for the Gaussian weights $\boldsymbol{w}$ as defined in (13) for $\alpha = 1, 2, \ldots$, as a function of κ^2. Then*

$$\begin{aligned} \boldsymbol{Q}_1(\kappa^2) &= \boldsymbol{K}(\kappa^2) & (18) \\ \boldsymbol{Q}_2(\kappa^2) &= \boldsymbol{K}(\kappa^2) \boldsymbol{C}^{-1} \boldsymbol{K}(\kappa^2) & (19) \end{aligned}$$

and for $\alpha = 3, 4, \ldots$

$$\boldsymbol{Q}_\alpha(\kappa^2) = \boldsymbol{K}(\kappa^2) \boldsymbol{C}^{-1} \boldsymbol{Q}_{\alpha-2} \boldsymbol{C}^{-1} \boldsymbol{K}(\kappa^2). \tag{20}$$

Some remarks to this result:

1. The matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ are easy to compute as their elements are non-zero only for pairs of basis-functions which share common triangles (a line segment in $\mathbb{R}^1$), and their values do not depend on κ^2. Analytic formulas are give in Appendix A.

2. A consequence of the previous remark is that we have an explicit mapping (the computational cost is $\mathcal{O}(n)$) between a GF and the GMRF representation in (13), for any triangulation, involving no computing at all!

3. The matrix $\boldsymbol{C}^{-1}$ is dense, which makes the precision matrix dense as well. In Section C.5, we give good arguments why $\boldsymbol{C}^{-1}$ should be replaced by the diagonal matrix $\widetilde{\boldsymbol{C}}^{-1}$ where $\widetilde{C}_{ij} = \langle \psi_i, 1 \rangle$, which makes the precision matrices sparse.

4. For the special case where all the vertices are points on a regular lattice, the Main Result 2 reduces to Main Result 1. Note that the neighbourhood of the corresponding GMRF in $\mathbb{R}^2$, is 3×3 for $\alpha = 1$, is 5×5 for $\alpha = 2$, and so on. Increased smoothness of the random field induces a larger neighbourhood in the GMRF representation.

5. In terms of the smoothness parameter ν in the Matérn covariance function, these results correspond to $\nu = 1/2, 3/2, 5/2, \ldots$, in $\mathbb{R}^1$ and $\nu = 0, 1, 2, \ldots$, in $\mathbb{R}^2$.

6. We are currently unable to provide results for other values of α; the main obstacle is the fractional derivative in the SPDE which is defined using the Fourier transform (7). A result of Rozanov (1982, Chapter 3.1) for the continuously indexed random field, says that a random field has a Markov property if and only if the reciprocal of the spectrum is a polynomial. Our SPDE (3) corresponds to $\alpha = 1, 2, 3, \ldots$; see (6). This result indicates that a different approach may be needed to provide representation results when α is not an integer.

Although our approach does give a GMRF representation of the Matérn field on the triangulated region, it is truly an approximation to the stochastic weak solution as we only use a subset of the test functions $\{\phi_k\}$. To see how well we now are able to approximate the Matérn covariance, panel (c) displays the non-zero entries in the (reordered) sparse precision matrix derived using our GMRF representation and panel (d) displays the empirical correlation function (dots) and the theoretical one for range equals 0.4 and $\nu = 1$. The accuracy of the GMRF-approximation is quite good as the empirical correlation function match quite well the true one. Some dots shows some discrepancy from the true correlations, but these can be identified to be due to the rather rough triangulation outside the area of interest included to reduce edge effects.

In practice there is a trade-off between accuracy of the GMRF representation and the number of vertices used. In Figure 2(a) we chose to use a fine resolution in the study-area and a reduced resolution outside. A minor drawback using GMRFs is that there are boundary effects, which is due to the boundary conditions of the SPDE. In Main Result 2 we used Neumann conditions (see Section A.4 for details) but other choices are also possible.

2.4 Leukaemia example

We will now return to the example from Henderson et al. (2002) which models spatial variation in leukaemia survival data in Northwest England. The model for the log-hazard is

$$\log(\text{hazard}) = \log(\text{basline}(\text{time})) + \text{intercept} + \text{sex} + \text{age} + \text{wbc} + \text{tpi} + \text{spatial}(\text{location}) \qquad (21)$$

where the smooth time-varying $\log(\text{baseline})$-model is piecewise constant model using 20 bins, "wbc" is the white blood-cell count at diagnosis, "tpi" is the Townsend deprivation index (which measures the deprivation for the related district) and "spatial" is the spatial component depending on the spatial location for each measurement. The hyper-parameters in this model, is the marginal precision and range for the spatial component, and the precision for the smoothness model for the $\log(\text{baseline})$-hazard.

Kneib and Fahrmeir (2007) reanalysed the same data-set using a similar model but was unable, for computational reasons, to include a spatial model using a dense covariance matrix but had to rely on a low-rank approximation. With our GMRF representation we easily work with a sparse 2535×2535 precision matrix for the spatial component. We ran the model in `R-inla` (`www.r-inla.org`) using integrated nested Laplace

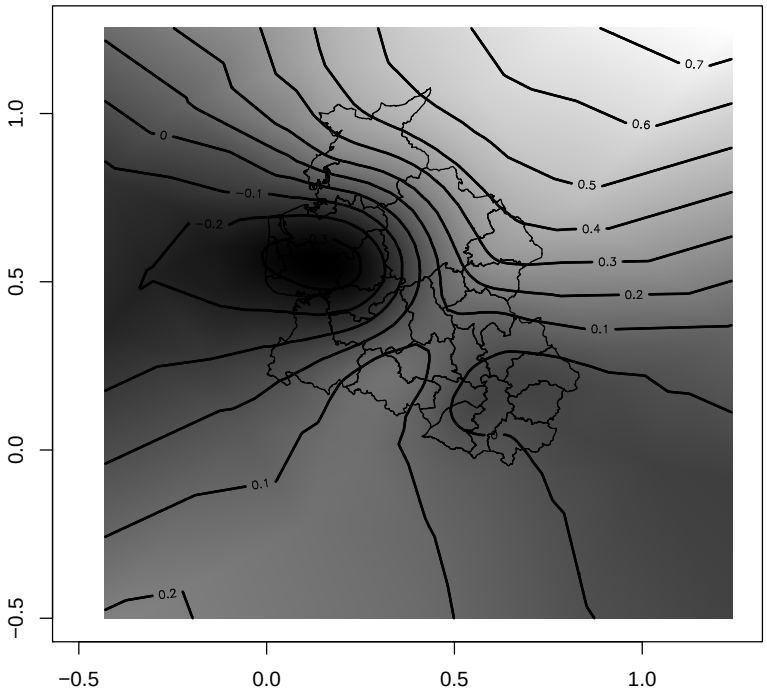


Figure 3: The estimated spatial effect of the log-hazard using the GMRF representation.

approximations to do the full Bayesian analysis (Rue et al., 2009) and thus completely avoiding any MCMC. Figure 3 displays the posterior mean of the spatial term, which shows a reduced hazard in the north-east corner. A full Bayesian analysis took about 60 seconds on a dual-core laptop, and factorising the 5258×5258 (total) precision matrix took about 0.07 seconds on average. We also note that the results are sensitive to prior specifications of the marginal precision and range.

3 Extensions

In this section we will discuss four extensions to the SPDE, widening the usefulness of the results in various ways. The first extension is to *define* the SPDE on a (regular) manifold, like the sphere, to define Matérn fields on the sphere. The second extension is to allow for space-varying parameters in the SPDE which allows us to construct non-stationary locally isotropic Gaussian fields. The third extension is to study a complex version of (3) which makes it possible to construct oscillating fields. Finally, the fourth extension generalises the non-stationary SPDE to non-isotropic fields.

An important feature in our approach, is that all these extensions still give explicit GMRF representations similar to (13) and (20), even if all the extensions are combined. The rather amazing consequence, is that we can construct the GMRF representations of non-stationary oscillating GFs on the sphere, still not requiring any computation. In Section 4, we will illustrate the use of these extensions, with a non-stationary model for global temperatures.

3.1 Matérn fields on manifolds

We will now move away from $\mathbb{R}^2$ and consider Matérn fields on differentiable manifolds. Our main objective is to construct Matérn fields on the sphere, which is important for the analysis of spatial and spatio-temporal models. To simplify the current discussion we will restrict the construction of Matérn fields to a unit radius sphere $\mathbb{S}^2$ in three dimensions, leaving the general case for the technical appendices.

One way of defining covariance models on a sphere is to interpret the two-dimensional space, $\mathbb{S}^2$, as a surface embedded in $\mathbb{R}^3$. Any three-dimensional covariance function can then be used to define the model

on the sphere, considering only the restriction of the function to the surface. This has the interpretational disadvantage of using chordal distances to determine the correlation between points. The simple alternative of plugging in the great circle distances into the original covariance function does not work, since this does not yield a valid positive definite covariance function. Thus, the Matérn covariance function in $\mathbb{R}^d$ can not be used to define GFs on a unit sphere embedded in $\mathbb{R}^3$ with distance naturally defined with respect to distances within the surface. However, we can still use its origin, the SPDE! For this purpose, we simply reinterpret the SPDE to be defined on $\mathbb{S}^2$ instead or $\mathbb{R}^d$, and the solution is still what we mean by a Matérn field, but defined directly for the given manifold. The Gaussian white noise which drives the SPDE can easily be defined on $\mathbb{S}^2$ as a (zero mean) random Gaussian measure $W(\cdot)$ with the property that the covariance between $W(A)$ and $W(B)$ for any subsets A and B of $\mathbb{S}^2$, is proportional to the surface integral over $A \cap B$. Any regular 2-manifold behaves locally like $\mathbb{R}^2$, which heuristically explains why the GMRF-representation of the weak solution only needs to change the definition of the inner product (14) to a surface integral on $\mathbb{S}^2$. Details appear in the technical appendices.

To illustrate the continuous index definition and the Markov representation of Matérn fields on a sphere, Figure 4 shows an irregular triangulation of the globe, with small triangles in the vicinity of meteorological measurement stations, the sparse structure of the optimally ordered precision matrix, a random realisation of the field, and the resulting numerically calculated correlation values as a function of great circle distances. Despite the highly irregular grid, the numerical correlations are tightly gathered around the theoretical values. The discrepancy between the approximation and the theoretical model can be more easily interpreted through the covariances, shown in Figure 5. The values deviate only for short distances, related to the local structure of the triangles. Panel (b) of the figure shows numerically calculated variances, showing that the deviation is approximately a linear function of the local triangle areas. This is to be expected, since the approximation construction is adapted to approximate distributions of integrals rather than point-wise distributions, and the variance at the vertices need then to be higher than the within-triangle variances.

3.2 Non-stationary fields

The most surprising extension within the SPDE-framework, is how we can model non-stationarity. Many applications do require non-stationarity in the correlation function and there is a vast literature on this subject (Cressie and Huang, 1999; Fuentes, 2001; Gneiting, 2002; Higdon, 1998; Higdon et al., 1999; Hughes-Oliver et al., 1998; Jun and Stein, 2008; Paciorek and Schervish, 2006; Stein, 2005). The SPDE approach has the additional huge advantage that the resulting (non-stationary) Gaussian field is a GMRF which allow for swift computations and can additionally be defined on a regular manifold.

In the SPDE defined in (3), the parameters κ^2 and the innovation variance are constant in space. In general, we can allow both parameters to depend on the coordinate $\boldsymbol{u}$, and we write

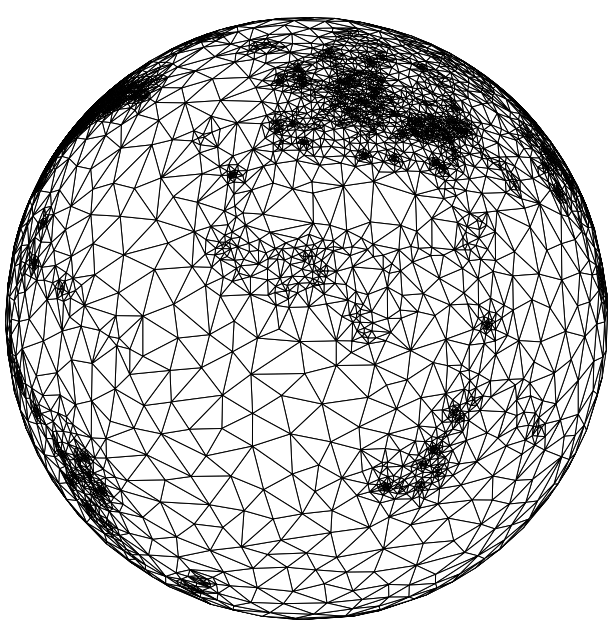
$$(\kappa^2(\boldsymbol{u}) - \Delta)^{\alpha/2}(\tau(\boldsymbol{u})x(\boldsymbol{u})) = \mathcal{W}(\boldsymbol{u}). \tag{22}$$

For simplicity, we chose to keep the variance for the innovation constant and scale the resulting process $x(\boldsymbol{u})$ with a scaling parameter $\tau(\boldsymbol{u})$. Non-stationarity is gained when one or both parameters is non-constant. Of particular interest is the case where they vary slowly with $\boldsymbol{u}$, for examples using a low-dimensional representation like

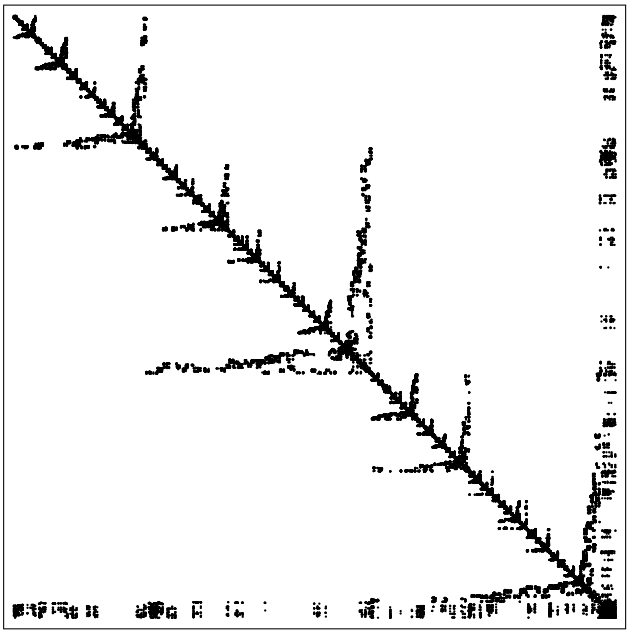
$$\log(\kappa^2(\boldsymbol{u})) = \sum_i \beta_i^{(\kappa^2)} B_i^{(\kappa^2)}(\boldsymbol{u}) \qquad \text{and} \qquad \log(\tau(\boldsymbol{u})) = \sum_i \beta_i^{(\tau)} B_i^{(\tau)}(\boldsymbol{u}) \tag{23}$$

where the basis functions $\{B_i^{(\cdot)}(\cdot)\}$ are smooth over the domain of interest. In the global temperatures on the earth example considered in Section 4, it is natural to use the spherical harmonics as the basis functions.

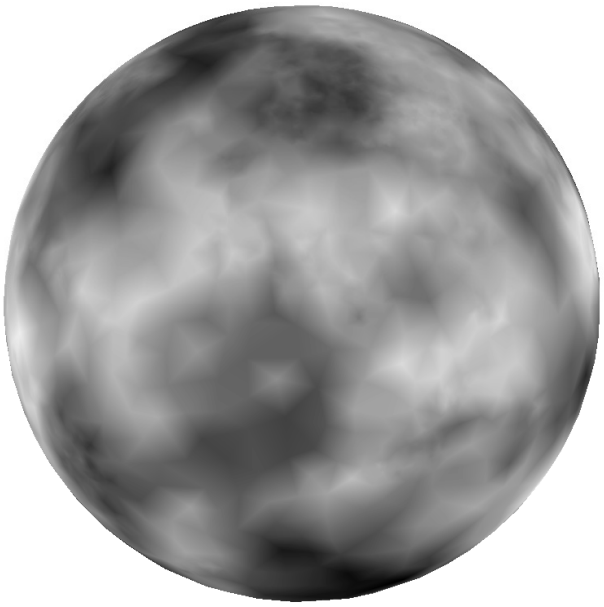
With slowly varying parameters $\kappa^2(\boldsymbol{u})$ and $\tau(\boldsymbol{u})$, the appealing *local* interpretation of (22) as a Matérn field remains unchanged, whereas the actual form of the achieved non-stationary correlation function is unknown. The actual process of "combining all local Matérn fields into a consistent global field", is done automatically by the SPDE.



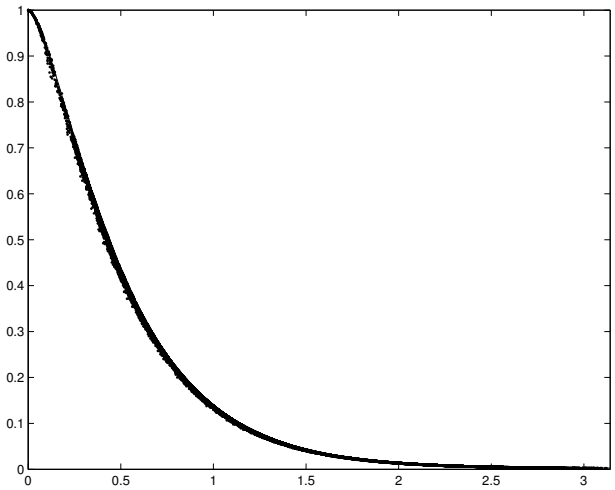
(a)



(b)

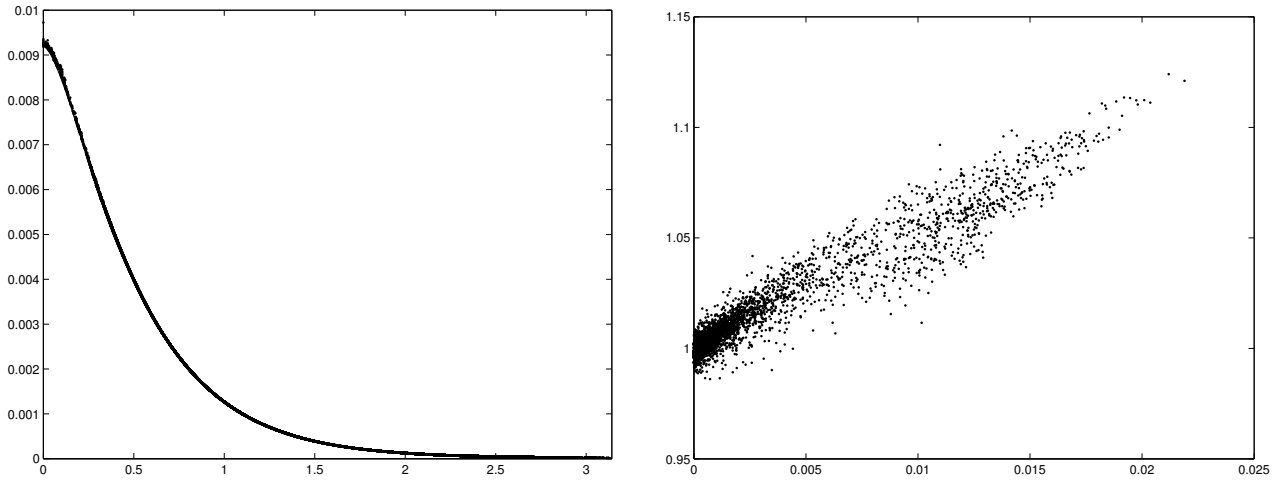


(c)



(d)

Figure 4: Panel (a) displays a triangulation of the earth (assuming it's a perfect sphere), panel (b) shows the (reordered) precision matrix for $\nu = 1$, panel (c) shows a sample of the Matérn field with $\kappa^2 = 9$ and in panel (c) we display the numerically calculated correlation function, together with the theoretical correlation function.



(a) (b)

Figure 5: Panel (a) displays the numerically calculated covariance function for the model in Figure 4, together with the theoretical covariance function, and panel (b) displays the ratio between the numerical and theoretical variances for each node, as a function of the local triangle areas.

The GMRF representation of (22) is found using the same approach as for the stationary case, with minor changes. For convenience, we assume that both κ^2 and τ can be considered as constant within the support of the basis-functions $\{\psi_k\}$, hence inner products like

$$\langle \psi_i, \kappa^2 \psi_j \rangle = \int \psi_i(\boldsymbol{u}) \psi_j(\boldsymbol{u}) \kappa^2(\boldsymbol{u}) \, d\boldsymbol{u} \approx C_{ij} \kappa^2(\boldsymbol{u}_j^*) \tag{24}$$

for a naturally defined $\boldsymbol{u}_j^*$ in the support of ψ_i and ψ_j. The consequence is a simple scaling of the matrices in (20) at no additional cost, see Section A.2. If we improve the integral approximation (24) from considering $\kappa^2(\boldsymbol{u})$ locally constant to locally planar, the computational preprocessing cost increases, but is still $\mathcal{O}(1)$ for each element in the precision matrix $\boldsymbol{Q}_\alpha$.

3.3 Oscillating covariance functions

In addition to non-stationary models, the SPDE framework admits many other extensions. One such possibility is to consider a complex version of the basic equation (3). For simplicity, we only consider the case $\alpha = 2$. With $\mathcal{W}_1$ and $\mathcal{W}_2$ as two independent white noise fields, and an *oscillation* parameter θ, the complex version becomes

$$(\kappa^2 \mathrm{e}^{i\pi\theta} - \Delta)(x_1(\boldsymbol{u}) + i x_2(\boldsymbol{u})) = \mathcal{W}_1(\boldsymbol{u}) + i\mathcal{W}_2(\boldsymbol{u}), \quad 0 \leq \theta < 1. \tag{25}$$

The real and imaginary stationary solution components $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent, with spectral densities

$$\frac{1}{(2\pi)^d} \cdot \frac{1}{\kappa^4 + 2\cos(\pi\theta)\kappa^2 \boldsymbol{k}^\mathsf{T}\boldsymbol{k} + (\boldsymbol{k}^\mathsf{T}\boldsymbol{k})^2}$$

on $\mathbb{R}^d$. The corresponding covariance functions for $\mathbb{R}$ and $\mathbb{R}^2$ are given in Appendix A. For general manifolds, no closed form expression can be found. In Figure 6, we illustrate the resonance effects obtained for compact domains by comparing oscillating covariances for $\mathbb{R}^2$ and the unit sphere, $\mathbb{S}^2$. The precision matrices for the resulting fields are obtained by a simple modification of the construction for the regular case, the precise expression given in Appendix A.

For $\theta = 0$, the regular Matérn covariance with $\nu = 2 - d/2$ is recovered, with oscillations increasing with θ. The limiting case $\theta = 1$ generates intrinsic stationary random fields, invariant to addition of cosine functions of arbitrary direction, with wave number κ.

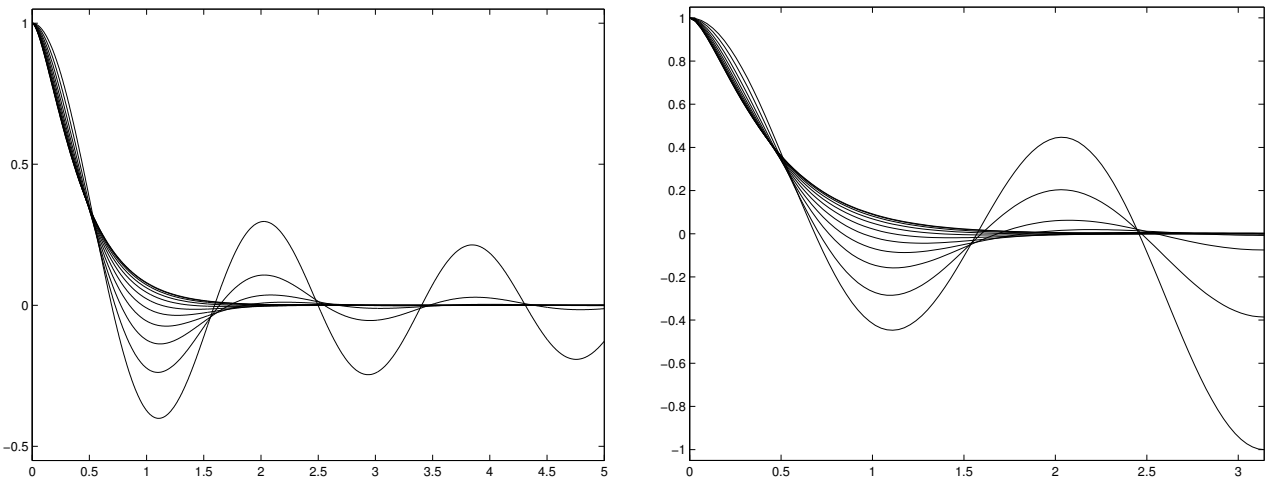


Figure 6: Correlation functions from oscillating SPDE models, for $\theta = 0, 0.1, \ldots, 1$, on $\mathbb{R}^2$ (panel (a)) and on $\mathbb{S}^2$ (panel (b)), with $\kappa^2 = 12$, $\nu = 1$.

3.4 Non-isotropic models and spatial deformations

The non-stationary model defined above, in Section 3.2, has locally isotropic correlations, despite having globally non-stationary correlations. This can be relaxed by widening the class of considered SPDE:s, allowing a non-isotropic Laplacian, and also by including a directional derivative term. The resulting models have close links to the deformation method for non-stationary covariances introduced by Sampson and Guttorp (1992).

In the deformation method, the domain is deformed into a space where the field is stationary, resulting in a non-stationary covariance model in the original domain. If the stationary field is generated by a stationary SPDE, the non-stationary model is generated by a corresponding non-stationary SPDE.

For notational simplicity, assume that the deformation is between two d-manifolds $\Omega \subseteq \mathbb{R}^d$ to $\widetilde{\Omega} \subseteq \mathbb{R}^d$, with $\boldsymbol{u} = \boldsymbol{f}(\widetilde{\boldsymbol{u}})$, $\boldsymbol{u} \in \Omega$, $\widetilde{\boldsymbol{u}} \in \widetilde{\Omega}$. Restricting to the case $\alpha = 2$, consider the stationary SPDE on the deformed space $\widetilde{\Omega}$

$$(1 - \widetilde{\nabla}^{\mathsf{T}}\boldsymbol{m} - \widetilde{\nabla}^{\mathsf{T}}\widetilde{\nabla})\widetilde{x}(\widetilde{\boldsymbol{u}}) = \widetilde{\sigma}\widetilde{\mathcal{W}}(\widetilde{\boldsymbol{u}}), \tag{26}$$

where $\boldsymbol{m}$ is a vector specifying a directional derivative, and $\widetilde{\sigma}^2$ is the variance of the driving white noise. Note that the model is stationary but non-isotropic if $\boldsymbol{m} \neq \boldsymbol{0}$. A change of variables onto the undeformed space Ω yields (Smith, 1934)

$$\frac{1}{\det(\boldsymbol{F}(\boldsymbol{u}))}\left(1 - \nabla^{\mathsf{T}}\boldsymbol{F}(\boldsymbol{u})\boldsymbol{m} - \det(\boldsymbol{F}(\boldsymbol{u}))\nabla^{\mathsf{T}}\frac{\boldsymbol{F}(\boldsymbol{u})\boldsymbol{F}(\boldsymbol{u})^{\mathsf{T}}}{\det(F(\boldsymbol{u}))}\nabla\right)x(\boldsymbol{u}) = \frac{\widetilde{\sigma}}{\det(\boldsymbol{F}(\boldsymbol{u}))^{1/2}}\mathcal{W}(\boldsymbol{u}), \tag{27}$$

where $\boldsymbol{F}(\boldsymbol{u})$ is the Jacobian of the deformation function $\boldsymbol{f}$. When $\boldsymbol{m} = \boldsymbol{0}$, this non-stationary SPDE exactly reproduces the deformation method with Matérn covariances introduced by Sampson and Guttorp (1992). A sparse GMRF approximation can be constructed using the same principles as for simpler non-stationary model in Section 3.2. A possible option for parameterising the model without explicit construction of a deformation function is to control the major axis of the local deformation given by $F(\boldsymbol{u})$ through a vector field. For example, on the sphere, this vector field could be parameterised as a weighted sum of vector spherical harmonics. By also re-introducing a spatially varying κ^2-parameter, a more general class of non-stationary models is obtained.

4 Example: Global temperature reconstruction

When analysing past observed weather and climate, the Global Historical Climatology Network (GHCN) data set[1] (Peterson and Vose, 1997) is commonly used. As of 2010-02-24, the data contains of meteorological observations from 7280 stations spread across continents, where each of the 422615 rows of observations contains the monthly mean temperatures from a specific station and year. The data spans the period 1702 through 2009, though counting, for each year, only stations with no missing values, yearly averages can be calculated only as far back as 1835. The spatial coverage varies from less than 400 stations prior to 1880 up to 3700 in the 1970s. For each station, covariate information such as location, elevation, and land use is available.

The GHCN data, or variants thereof, is used to analyse regional and global temperatures in the GISS (Hansen et al., 1999, 2001) and HadCRUT3 (Brohan et al., 2006) global temperature series, together with additional data such as ocean based measurements. Various techniques for handling station specific effects are applied, and the information about the temperature anomaly (the difference in weather to the local climate, the latter defined as the average weather over a 30 year reference period) is then aggregated to latitude-longitude grid boxes. The grid box anomalies are then combined using area based weights into an estimate of the average global anomaly for each year. The analysis is accompanied by a derivation of the resulting uncertainty of the estimates.

Though different in details, the gridding procedures are algorithmically based, i.e. there is no underlying statistical model for the weather and climate, only for the observations themselves. We will here present a basis for a stochastic model based approach to the problem of estimating past regional and global temperatures, as an example of how the non-stationary SPDE models can be used in practice. The ultimate goal is reconstruct the entire spatio-temporal yearly average temperature field, with appropriate measures of uncertainty taking the model parameter uncertainty into account.

Since most of the spatial variation is linked to the rotational nature of the globe in relation to the sun, we will here restrict ourselves to a rotationally invariant model, which reduces the initial computational burden. The model separates weather from climate by assuming that the climate can be parameterised by rotationally invariant expectation and covariance parameters $\mu(\boldsymbol{u})$, $\kappa(\boldsymbol{u})$, and $\tau(\boldsymbol{u})$, for $\boldsymbol{u} \in \mathbb{S}^2$, and assuming that the yearly weather follows the model defined by (22), given the climate. Using the triangulation in Figure 4 with piecewise linear basis function, the GMRF representation given in Section A.2 will be used, with $\boldsymbol{x}_t$ denoting the discretised field at time t.

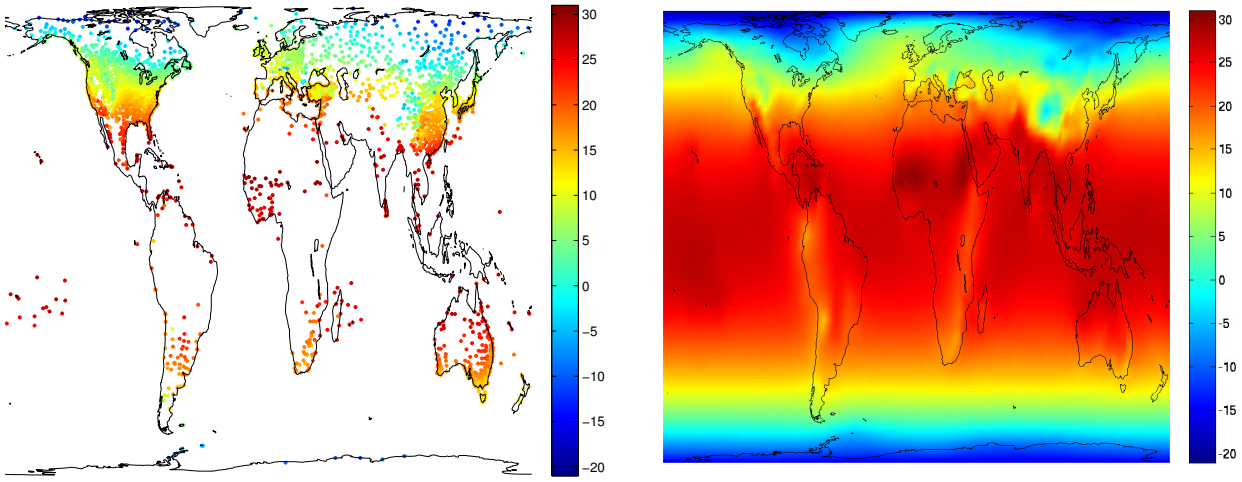
Introducing *observation matrices* $\boldsymbol{A}_t$, that extract the nodes from $\boldsymbol{x}_t$ for each observation, the full model can be summarised as follows:

- Climate ($\boldsymbol{\theta}$, $\boldsymbol{B}_\cdot$ = rotationally invariant spherical harmonics up to order 3):
  - Expectation field $\mu(\boldsymbol{u})$: $\boldsymbol{\mu} = \boldsymbol{B}_\mu \theta_\mu$
  - Local spatial dependence $\kappa(\boldsymbol{u})$: $\log \kappa^2 = \boldsymbol{B}_\kappa \theta_\kappa$
  - Local variance scaling $\tau(\boldsymbol{u})$: $\log \tau = \boldsymbol{B}_\tau \theta_\tau$
- Weather/spatial yearly temperature means ($\boldsymbol{x}_t$):

$$(\boldsymbol{x}_t|\boldsymbol{\theta}) \sim N(\boldsymbol{\mu}(\theta_\mu), \boldsymbol{Q}_x^{-1}(\theta_\kappa, \theta_\tau))$$

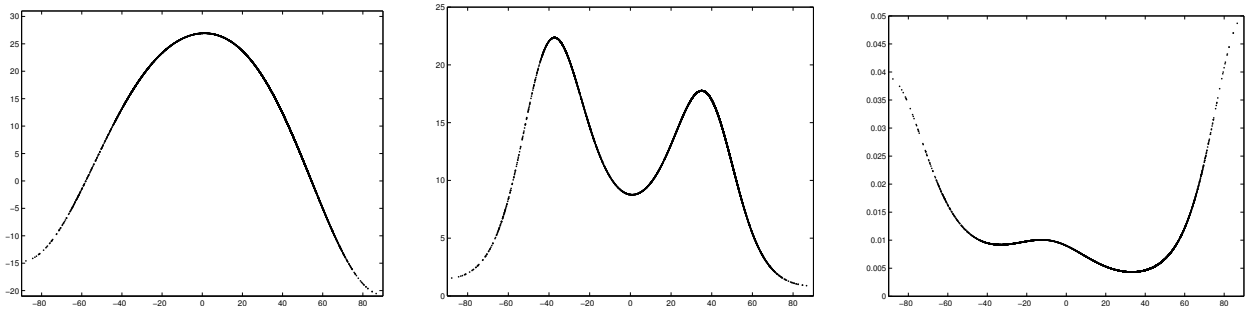
- Temperature data ($\boldsymbol{y}_t$):
  - Station specific effects: $\boldsymbol{S}_t \theta_y$
  - Observation precision: $\boldsymbol{Q}_y(\theta_\epsilon) = \mathrm{e}^{\theta_\epsilon} \boldsymbol{I}$
  - Observed weather: $(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}) \sim N(\boldsymbol{A}_t \boldsymbol{x}_t + \boldsymbol{S}_t \theta_y, \boldsymbol{Q}_y^{-1}(\theta_\epsilon))$

[1] `http://www.ncdc.noaa.gov/ghcn/ghcn.html`



(a) (b)

Figure 7: Temperature data and global reconstruction for 1980.



(a) (b) (c)

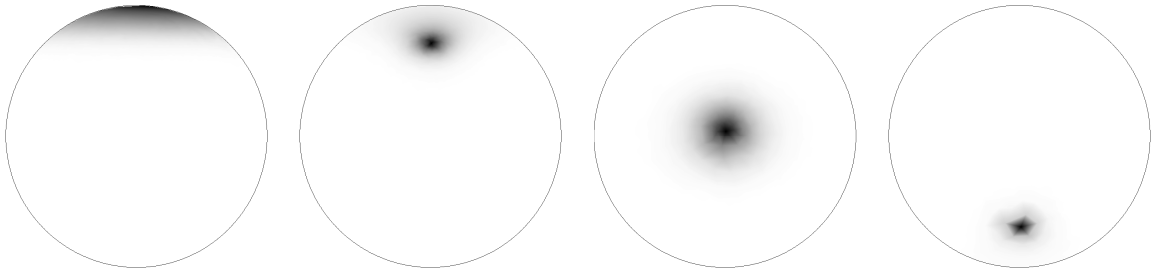
Figure 8: Estimated parameter fields, μ, κ and τ.

Since we only use the data for illustrative purposes here, we will ignore the station specific effects, that need to be included in a future more complete analysis. We also ignore any spatial dependencies between consecutive years, which for yearly data is small, analysing only the marginal distribution properties of each year.

Just as for the example in Section 2.4, we implemented the model in the INLA framework. Due to the large size of the data set, this initial analysis is based on data only from the period 1980 to 1989, requiring on the order of 120000 nodes in a joint model for the temperature fields and measurements, with 9500 nodes in each field, and on average 2500 observations per year. Further enhancements to the INLA implementation are needed to smoothly handle the full data set in an automated joint estimation.

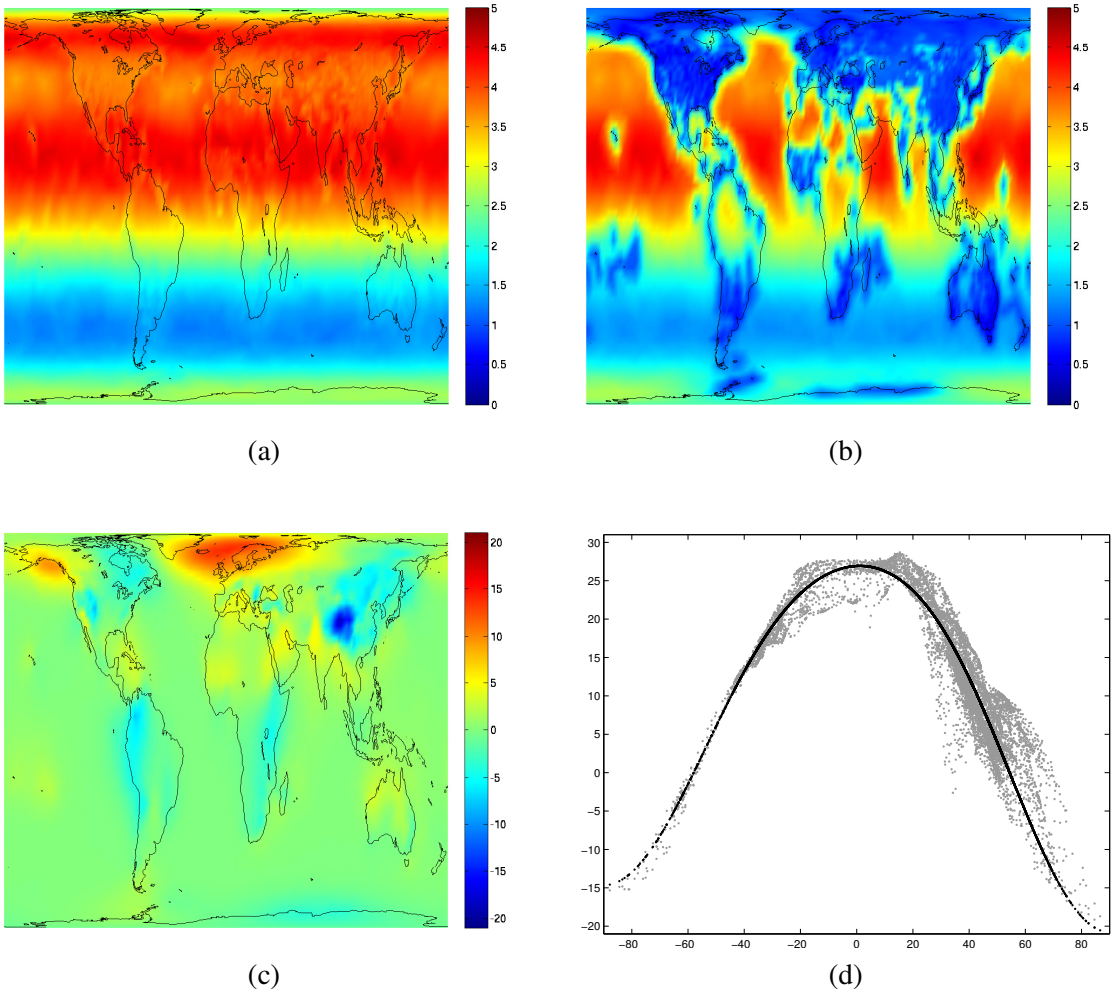
In Figure 7(a), the measured temperatures for 1980 are shown, with the corresponding posterior reconstruction $\mathrm{E}(\boldsymbol{x}_{1980}(\boldsymbol{u})|\boldsymbol{y})$ shown in panel (b). The point estimates of the climate parameters in shown in Figure 8. As expected, the temperatures are low near the poles and high near the equator. The covariance parameters are more difficult to interpret, but they indicate that the correlation range is high at the poles, smaller at the equator, and small around the ± 40 latitudes, as illustrated in Figure 9.

The estimated model standard deviations are shown in Figure 10(a). Note the small deviations from rotational symmetry, due to the simple approximation of the non-stationary GMRF representation. Integrating out the parameter uncertainty with the INLA method, we calculated posterior point-wise variances $\mathrm{Var}(\boldsymbol{x}_t(\boldsymbol{u}_i)|\boldsymbol{y})$, shown as standard deviations in Figure 10(b). To see how well the simple expectation model captures the temporal averages, we calculated the difference between the posterior temporal averages, $\frac{1}{30}\sum_{t=1980}^{1989}\mathrm{E}(\boldsymbol{x}_t(\boldsymbol{u})|\boldsymbol{y})$



(a) (b) (c) (d)

Figure 9: Correlation functions centered at latitudes 90N, 45N, equator, and 45S. The correlation range varies with latitude.



(a) (b)

(c) (d)

Figure 10: Standard deviation, posterior standard deviation, spatial differences between average (over 1980–1989) of posterior means and the expectation.

and the estimated $\boldsymbol{\mu}$, as shown in Figure 10(c) and (d). The overall structure has been captured, apart from some extra warming around the North Atlantic and north-east Pacific coasts, due to ocean currents. Not surprisingly, a clear effect of regional topography can be seen, showing cold areas for high elevations such as in the Himalayas. This effect can be handled naturally using the station elevations as covariate information. The uncertain extent of the ocean-induced warm areas might be accounted for by using a more general expectation model, but that would require incorporating ocean based data as well, to improve the identifiability of the parameters due to the sparse spatial coverage.

5 Discussion

The main result in this work is that we can construct an explicit link between (some) Gaussian fields and Gaussian Markov random fields using an approximate weak solution of the corresponding stochastic partial differential equation. Although this result is not generally applicable for all covariance functions, the subclass of models where this result is applicable is substantial, and we expect to find additional versions and extensions in the future; see for example Bolin and Lindgren (2009a). The explicit link makes these Gaussian fields much more practically applicable, as we might model and interpret the model using covariance functions while doing the computations using the GMRF representation which allow for sparse-matrix numerical linear algebra. In most cases, we can make use of the INLA approach for doing (approximate) Bayesian inference (Rue et al., 2009), which require the latent field to be a GMRF. It is our hope that the SPDE link might help bridging the literature of (continuously indexed) Gaussian fields and Geostatistics on one side, and Gaussian Markov random fields/Conditional auto-regressions on the other.

Furthermore, the simplicity of the SPDE parameter specifications provides a new modelling approach that is not dependent on the theory for constructing positive definite covariance functions. Although still slightly puzzled by the apparent simplicity of the SPDE approach, we are still in a mild shock of what this result implies. The ease of how non-stationary models can be defined using spatially varying parameters of the SPDE is both natural, gives good local interpretation, and is computationally very efficient, as we still obtain a GMRF representation. The extension to manifolds is also useful, with fields on the globe as the main example.

A third issue not yet discussed, is that the SPDE approach might help interpret external covariates (for example wind speed) as an appropriate drift term or similar in the related SPDE and then this covariate would enter correctly in the spatial dependence models. This is again an argument for more physics based spatial modelling, but as we have shown in this report, such an approach can also provide a huge computational benefit.

On the negative side, our approach comes with an implementation and preprocessing cost for setting up the models, as it involves the SPDE, triangulations, and GMRF representations, but we firmly believe that such costs are unavoidable when efficient computations are required.

A Explicit coefficients and covariances

This section includes some explicit expressions and results not included in the main text.

A.1 Grid based precisions

We will here give some explicit precision expressions for grid based models on $\mathbb{R}$ and $\mathbb{R}^2$. Consider the SPDE

$$(\kappa^2 - \nabla^\mathsf{T} \boldsymbol{F} \boldsymbol{F}^\mathsf{T} \nabla)^{\alpha/2} x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}), \quad \Omega = \mathbb{R}^d, \, d = 1 \text{ or } 2, \tag{28}$$

where $\boldsymbol{F}$ is a diagonal d-dimensional matrix.

For any given ordered discretisation $u_1, \ldots, u_n$ on $\mathbb{R}$, let $\gamma_i = u_i - u_{i-1}$, $\delta_i = u_{i+1} - u_i$, and $s_i = (\gamma_i + \delta_i)/2$. Then, the elements on row i, around the diagonal, of the precision are given by

$$\boldsymbol{Q}_1 : s_i \cdot \begin{bmatrix} -a_i & c_i & -b_i \end{bmatrix}$$
$$\boldsymbol{Q}_2 : s_i \cdot \begin{bmatrix} a_i a_{i-1} & -a_i(c_{i-1} + c_i) & a_i b_{i-1} + c_i^2 + b_i a_{i+1} & -b_i(c_i + c_{i+1}) & b_i b_{i+1} \end{bmatrix}$$

where $a_i = F^2/(\gamma_i s_i)$, $b_i = F^2/(\delta_i s_i)$, and $c_i = \kappa^2 + a_i + b_i$. If the spacing is regular, $s = \delta = \gamma$, and the expressions simplify to

$$\begin{aligned}
\boldsymbol{Q}_1 &: s \cdot \begin{bmatrix} -a & c & -a \end{bmatrix} \\
\boldsymbol{Q}_2 &: s \cdot \begin{bmatrix} a^2 & -2ac & 2a^2 + c^2 & -2ac & a^2 \end{bmatrix}
\end{aligned}$$

where $a = F^2/\delta^2$ and $c = \kappa^2 + 2a$. The special case $\alpha = 2$ with $\kappa = 0$ and irregular spacing is a generalisation of Lindgren and Rue (2008).

For $\mathbb{R}^2$, assume a given regular grid discretisation, with horizontal (coordinate component 1) distances γ and vertical (coordinate component 2) distances δ. Let $s = \gamma\delta$, $a = F_{11}^2/\gamma^2$, $b = F_{22}^2/\delta^2$, and $c = \kappa^2 + 2a + 2b$. The precision elements are then given by

$$\begin{aligned}
\boldsymbol{Q}_1 &: s \cdot \begin{bmatrix} -b & \\ c & -a \end{bmatrix} \\
\boldsymbol{Q}_2 &: s \cdot \begin{bmatrix} b^2 & & \\ -2bc & 2ab & \\ 2a^2 + 2b^2 + c^2 & -2ac & a^2 \end{bmatrix} \\
\boldsymbol{Q}_3 &: s \cdot \begin{bmatrix} -b^3 & & & \\ 3b^2c & -3ab^2 & & \\ -3b(2a^2 + b^2 + c^2) & 6abc & -3a^2b & \\ c(6a^2 + 6b^2 + c^2) & -3a(a^2 + 2b^2 + c^2) & 3a^2c & -a^3 \end{bmatrix} \\
\boldsymbol{Q}_4 &: s \cdot \begin{bmatrix} b^4 & & & & \\ -4b^3c & 4ab^3 & & & \\ 2b^2(6a^2 + 2b^2 + 3c^2) & -12ab^2c & 6a^2b^2 & & \\ -4bc(6a^2 + 3b^2 + c^2) & 12ab(a^2 + b^2 + c^2) & -12a^2bc & 4a^3b & \\ 6(a^4 + b^4 + 4a^2b^2) + 12(a^2 + b^2)c^2 + c^4 & -4ac(3a^2 + 6b^2 + c^2) & 2a^2(2a^2 + 6b^2 + 3c^2) & -4a^3c & a^4 \end{bmatrix}
\end{aligned}$$

If the grid distances are proportional to the corresponding diagonal elements of $\boldsymbol{F}$ (such as in the isotropic case $\gamma = \delta$ and $F_{11} = F_{22}$), the expressions simplify to

$$\begin{aligned}
&s = \gamma\delta \\
&a = F_{11}^2/\gamma^2 = F_{22}/\delta^2 \\
&c = \kappa^2 + 4a \\
\boldsymbol{Q}_2 &: s \cdot \begin{bmatrix} a^2 & & \\ -2ac & 2a^2 & \\ 4a^2 + c^2 & -2ac & a^2 \end{bmatrix} \\
\boldsymbol{Q}_3 &: s \cdot \begin{bmatrix} -a^3 & & & \\ 3a^2c & -3a^3 & & \\ -3a(3a^2 + c^2) & 6a^2c & -3a^3 & \\ c(12a^2 + c^2) & -3a(3a^2 + c^2) & 3a^2c & -a^3 \end{bmatrix} \\
\boldsymbol{Q}_4 &: s \cdot \begin{bmatrix} a^4 & & & & \\ -4a^3c & 4a^4 & & & \\ 2a^2(8a^2 + 3c^2) & -12a^3c & 6a^4 & & \\ -4ac(9a^2 + c^2) & 12a^2(2a^2 + c^2) & -12a^3c & 4a^4 & \\ 36a^4 + 24a^2c^2 + c^4 & -4ac(9a^2 + c^2) & 2a^2(8a^2 + 3c^2) & -4a^3c & a^4 \end{bmatrix}
\end{aligned}$$

A.2 Non-stationary and oscillating precision matrices

For easy reference, we give specific precision matrix expressions for the case $\alpha = 2$ for arbitrary triangulated manifold domains Ω.

The stationary and simple oscillating models for $\alpha = 2$ have precision matrices given by

$$\boldsymbol{Q}_2(\kappa^2, \theta) = \kappa^4 \boldsymbol{C} + 2\kappa^2 \cos(\pi\theta)\boldsymbol{G} + \boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G}, \tag{29}$$

where $\theta = 0$ corresponds to the regular Matérn case and $0 < \theta < 1$ are oscillating models.

Using the approximation from (24), the non-stationary model (22) with $\alpha = 2$ has precision matrix given by

$$\boldsymbol{Q}_2(\kappa^2(\cdot), \tau(\cdot)) = \boldsymbol{\tau}\left(\boldsymbol{\kappa}^2\boldsymbol{C}\boldsymbol{\kappa}^2 + \boldsymbol{\kappa}^2\boldsymbol{G} + \boldsymbol{G}\boldsymbol{\kappa}^2 + \boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G}\right)\boldsymbol{\tau} \tag{30}$$

where $\boldsymbol{\kappa}^2$ and $\boldsymbol{\tau}$ are diagonal matrices, with $\boldsymbol{\kappa}^2_{ii} = \kappa(\boldsymbol{u}_i)^2$ and $\boldsymbol{\tau}_{ii} = \tau(\boldsymbol{u}_i)$.

A.3 Precision building blocks for triangulated domains

In this section, we derive explicit expressions for the building blocks for the precision matrices, for general triangulated domains with piecewise linear basis functions. We need to calculate

$$\widetilde{C}_{i,i} = \langle \psi_i, 1 \rangle_\Omega \,, \tag{31}$$
$$C_{i,j} = \langle \psi_i, \psi_j \rangle_\Omega \,, \tag{32}$$
$$G_{i,j} = \langle \nabla\psi_i, \nabla\psi_j \rangle_\Omega \,, \tag{33}$$
$$B_{i,j} = \langle \psi_i, \partial_{\boldsymbol{n}}\psi_j \rangle_{\partial\Omega} \,. \tag{34}$$

For 2-manifolds such as regions in $\mathbb{R}^2$ or on $\mathbb{S}^2$, we require a triangulation with a set of vertices $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, embedded in $\mathbb{R}^3$. Each vertex $\boldsymbol{v}_k$ is assigned a continuous piecewise linear basis function ψ_k with support on the triangles attached to $\boldsymbol{v}_k$. In order to obtain explicit expressions for (31)–(34), we need to introduce some notation for geometry of an arbitrary triangle. For notational convenience, we number the corner vertices of a given triangle $T = (\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2)$. The edge vectors opposite each corner are

$$\boldsymbol{e}_0 = \boldsymbol{v}_2 - \boldsymbol{v}_1, \quad \boldsymbol{e}_1 = \boldsymbol{v}_0 - \boldsymbol{v}_2, \quad \boldsymbol{e}_2 = \boldsymbol{v}_1 - \boldsymbol{v}_0,$$

and the corner angles are θ_0, θ_1, and θ_2.

The triangle area $|T|$ can be obtained from the formula

$$|T| = \frac{1}{2}\|\boldsymbol{e}_0 \times \boldsymbol{e}_1\| \tag{35}$$

i.e. half the length of the vector product in $\mathbb{R}^3$. The contributions from the triangle to the $\widetilde{\boldsymbol{C}}$ and $\boldsymbol{C}$ matrices are given by

$$\left[\widetilde{C}_{i,i}(T)\right]_{i=0,1,2} = \frac{|T|}{3}\begin{bmatrix}1 & 1 & 1\end{bmatrix}, \tag{36}$$

$$\left[C_{i,j}(T)\right]_{i,j=0,1,2} = \frac{|T|}{12}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & 1\\ 1 & 1 & 2\end{bmatrix} \tag{37}$$

The contribution to $G_{0,1}$ from the triangle T is

$$G_{0,1}(T) = |T|(\nabla\psi_0)^\mathsf{T}(\nabla\psi_1) = -\frac{\cot(\theta_2)}{2} = \frac{1}{4|T|}\boldsymbol{e}_0^\mathsf{T}\boldsymbol{e}_1,$$

and the entire contribution from the triangle is

$$\left[G_{i,j}(T)\right]_{i,j=0,1,2} = \frac{1}{4|T|}\begin{bmatrix}\|\boldsymbol{e}_0\|^2 & \boldsymbol{e}_0^\mathsf{T}\boldsymbol{e}_1 & \boldsymbol{e}_0^\mathsf{T}\boldsymbol{e}_2\\ \boldsymbol{e}_1^\mathsf{T}\boldsymbol{e}_0 & \|\boldsymbol{e}_1\|^2 & \boldsymbol{e}_1^\mathsf{T}\boldsymbol{e}_2\\ \boldsymbol{e}_2^\mathsf{T}\boldsymbol{e}_0 & \boldsymbol{e}_2^\mathsf{T}\boldsymbol{e}_1 & \|\boldsymbol{e}_2\|^2\end{bmatrix} = \frac{1}{4|T|}\begin{bmatrix}\boldsymbol{e}_0 & \boldsymbol{e}_1 & \boldsymbol{e}_2\end{bmatrix}^\mathsf{T}\begin{bmatrix}\boldsymbol{e}_0 & \boldsymbol{e}_1 & \boldsymbol{e}_2\end{bmatrix}. \tag{38}$$

For the boundary integrals in (34), the contribution from the triangle is

$$\left[B_{i,j}(T)\right]_{i,j=0,1,2} = \frac{-1}{4|T|} \begin{bmatrix} \mathbf{0} & \boldsymbol{e}_0 & \boldsymbol{e}_0 \\ \boldsymbol{e}_1 & \mathbf{0} & \boldsymbol{e}_1 \\ \boldsymbol{e}_2 & \boldsymbol{e}_2 & \mathbf{0} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} e_0\boldsymbol{I} \\ e_1\boldsymbol{I} \\ e_2\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_0 & \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{bmatrix}, \tag{39}$$

where $e_k = \mathbb{I}(\text{Edge } k \text{ in } T \text{ lies on } \partial\Omega))$. Summing the contributions from all the triangles yields the complete $\widetilde{\boldsymbol{C}}$, $\boldsymbol{C}$, $\boldsymbol{G}$, and $\boldsymbol{B}$ matrices.

For the anisotropic version, parameterised as in Section C.4, with $\boldsymbol{H} = \boldsymbol{F}\boldsymbol{F}^{\mathsf{T}}$ in (28), the modified $\boldsymbol{G}$ matrix elements are given by

$$\left[G_{i,j}(T)\right]_{i,j=0,1,2} = \frac{1}{4|T|} \begin{bmatrix} \boldsymbol{e}_0 & \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{bmatrix}^{\mathsf{T}} \text{adj}(\boldsymbol{H}) \begin{bmatrix} \boldsymbol{e}_0 & \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{bmatrix}, \tag{40}$$

where $\text{adj}(\boldsymbol{H})$ is the adjugate matrix of $\boldsymbol{H}$, for non-singular matrices defined as $\det(\boldsymbol{H})\boldsymbol{H}^{-1}$.

A.4 Neumann boundary effects

The effects on the covariance functions resulting from using Neumann boundary conditions can be explicitly expressed as a folding effect. When the full SPDE is

$$\begin{cases} (\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}), & \boldsymbol{u} \in \Omega \\ \partial_{\boldsymbol{n}} (\kappa^2 - \Delta)^j x(\boldsymbol{u}) = 0, & \boldsymbol{u} \in \partial\Omega,\ j = 0, 1, \ldots, \lfloor (\alpha - 1)/2 \rfloor, \end{cases} \tag{41}$$

the following theorem provides a direct answer, in terms of the Matérn covariance function.

Theorem 1 *If x is a solution to the boundary value problem* (41) *for $\Omega = [0, L]$ and a positive integer α, then*

$$\mathit{Cov}(x(u), x(v)) = \sum_{k=-\infty}^{\infty} \left(r_M(u, v - 2kL) + r_M(u, 2kL - v) \right) \tag{42}$$

where r_M is the Matérn covariance as defined on the whole of $\mathbb{R}$.

The theorem, that extends naturally to arbitrary generalised rectangles in $\mathbb{R}^d$, is proved in Appendix D.1. In practice, when the effective range is small, only the three main terms need to be included for a very close approximation:

$$\begin{aligned} \text{Cov}(x(u), x(v)) &\approx r_M(u, v) + r_M(u, -v) + r_M(u, 2L - v) && (43) \\ &= r_M(0, v - u) + r_M(0, v + u) + r_M(0, 2L - (v + u)). && (44) \end{aligned}$$

Moreover, the resulting covariance is nearly indistinguishable from the stationary Matérn covariance at distances greater than twice the range away form the borders of the domain.

A.5 Oscillating covariances

The covariances for the oscillating model can be calculated explicitly for $\mathbb{R}$ and $\mathbb{R}^2$, from the spectrum. On $\mathbb{R}$, complex analysis gives

$$r(u, v) = \frac{1}{2\sin(\pi\theta)\kappa^3} e^{-\kappa\cos(\pi\theta/2)|v-u|} \sin\left(\pi\theta/2 + \kappa\sin(\pi\theta/2)|v - u|\right), \tag{45}$$

which has variance $(4\cos(\pi\theta/2)\kappa^3)^{-1}$. On $\mathbb{R}^2$, complicated Bessel function integrals yield

$$r(\boldsymbol{u}, \boldsymbol{v}) = \frac{1}{4\pi\sin(\pi\theta)\kappa^2} \Im\left(K_0(\kappa\|\boldsymbol{v} - \boldsymbol{u}\| e^{-i\pi\theta/2}) - K_0(\kappa\|\boldsymbol{v} - \boldsymbol{u}\| e^{i\pi\theta/2}) \right) \tag{46}$$

which has variance $(4\pi\kappa^2\,\text{sinc}(\theta))^{-1}$.

B Manifolds, random measures, and operator identities

B.1 Manifold basics

Although the theory needed for the analysis of stochastic partial differential equations on manifolds is well developed in principle, it can be difficult to locate a concise summary of the needed concepts and definitions. Therefore, we here give a brief summary of the necessary theory, aimed at statisticians familiar with measure theory and stochastic calculus on $\mathbb{R}^d$. For more details on manifolds, differential calculus and geometric measure theory see for example Auslander and MacKenzie (1977), Federer (1978) and Krantz and Parks (2008).

Loosely, we say that a space Ω is a *d-manifold* if it locally behaves as $\mathbb{R}^d$. We only consider manifolds with well-behaved boundaries, in the sense that the boundary $\partial\Omega$ of a manifold, if present, is required to be a $(d-1)$-manifold. By also requiring the manifolds to be *orientable*, possibly problematic domains such as Möbius strips are ruled out.

Definition 1 (Metric manifolds) *A* Riemannian manifold (Ω, m) *is a smooth manifold with a metric m. The metric defines a smoothly varying family of inner products for the collection of tangent spaces to points in Ω. The metric induces natural notions of lengths of curves, areas or volumes of subregions, angles, as well as rules for calculus. Common terms related to the topology of a manifold:*

1. *A manifold is* connected *if every pair of points in Ω can be joined by a curve in Ω.*
2. *The* geodesic distance $d(\boldsymbol{u}, \boldsymbol{v})$ *between two points $\boldsymbol{u}, \boldsymbol{v} \in \Omega$ is the infimum of the lengths of all curves in Ω connecting $\boldsymbol{u}$ and $\boldsymbol{v}$.*
3. *A* geodesic *is a curve in Ω with length $d(\boldsymbol{u}, \boldsymbol{v})$ that connects two points $\boldsymbol{u}, \boldsymbol{v} \in \Omega$.*
4. *If the geodesic distances on a manifold are bounded, i.e. $\sup_{\boldsymbol{u},\boldsymbol{v}\in\Omega} d(\boldsymbol{u}, \boldsymbol{v}) < \infty$, the manifold is* bounded. *The maximal geodesic distance is called the* diameter *of Ω.*
5. *A* compact *manifold is a bounded, complete manifold, i.e. every Cauchy sequence on Ω has a limit in Ω, so Ω includes its boundary, if any, denoted $\partial\Omega$.*
6. *A* closed *manifold is a compact manifold with no boundary.*
7. *The metric m_∂ for the $(d-1)$-manifold boundary is defined by the restriction of m to the tangent spaces of $\partial\Omega$. If Ω is compact, $\partial\Omega$ is closed.*

The most common Riemannian manifolds are subsets of $\mathbb{R}^d$ equipped with the Euclidean metric. For the most part, we will consider submanifolds of $\mathbb{R}^d$, with metrics defined by the restriction of the Euclidean metric to each manifold. For brevity of notation, we write "a manifold Ω" throughout. Examples of manifolds include intervals on $\mathbb{R}$, regions in $\mathbb{R}^2$, the surface of the unit radius sphere $\mathbb{S}^2$ embedded in $\mathbb{R}^3$, a torus embedded in $\mathbb{R}^3$, and a flat torus, i.e. a manifold with torus topology, but with a different metric. The flat torus commonly appears as a tool to make a periodic continuation of a rectangle in $\mathbb{R}^2$.

Some useful definitions for calculus on Ω follows.

Definition 2 (Manifold differential calculus) *Let ϕ denote a function $\phi : \Omega \mapsto \mathbb{R}$.*

1. *The* gradient *of ϕ at $\boldsymbol{u}$ is a vector $\nabla\phi(\boldsymbol{u})$ in the tangent space of Ω at $\boldsymbol{u}$, such that for any tangent vector $\boldsymbol{t}$ at $\boldsymbol{u}$, the inner product is $m(\nabla\phi, \boldsymbol{t}) = \partial_{\boldsymbol{t}}\phi$ where $\partial_{\boldsymbol{t}}\phi$ denotes the directional derivative along $\boldsymbol{t}$. In $\mathbb{R}^d$ with Euclidean metric, the gradient operator ∇ is formally given by the column vector $\left[\frac{\partial}{\partial u_1} \cdots \frac{\partial}{\partial u_d}\right]^\mathsf{T}$.*
2. *The* Laplacian Δ *of ϕ at $\boldsymbol{u}$ (or the Laplace or Laplace-Beltrami operator) is defined as the sum of the second order directional derivatives, with respect to a local orthonormal basis for the tangent space. In Euclidean metric on $\mathbb{R}^d$, we can write $\Delta\phi = \nabla^\mathsf{T}\nabla\phi$, and formally, $\Delta = \frac{\partial^2}{\partial u_1^2} + \cdots + \frac{\partial^2}{\partial u_d^2}$.*

3. *The vector $\boldsymbol{n}_\partial(\boldsymbol{u})$ denotes the unit length outward normal vector at a point $\boldsymbol{u}$ on the boundary of Ω, i.e. $\boldsymbol{n}_\partial$ is orthogonal to the tangent space of $\partial\Omega$ at $\boldsymbol{u}$. The* normal derivative *of a function ϕ at $\boldsymbol{u} \in \partial\Omega$ is the directional derivative, i.e. $\partial_{\boldsymbol{n}}\phi(\boldsymbol{u}) = \boldsymbol{n}_\partial(\boldsymbol{u})^\mathsf{T}\nabla\phi(\boldsymbol{u})$ in Euclidean metric.*

4. *For notational convenience, we let ∇_∂ and Δ_∂ denote the restrictions of ∇ and Δ to a boundary manifold $\partial\Omega$. In particular, we have $\nabla_\partial = \nabla - \boldsymbol{n}_\partial\partial_{\boldsymbol{n}}$ and, for the Euclidean metric, $\Delta_\partial = \nabla_\partial^\mathsf{T}\nabla_\partial$.*

5. *The length of a differentiable curve $\gamma(t) : [a, b] \mapsto \Omega$ is given by*

$$|\gamma|_\Omega = \int_a^b \left\{ m\left(\frac{\mathrm{d}\gamma}{\mathrm{d}t}, \frac{\mathrm{d}\gamma}{\mathrm{d}t}\right)\right\}^{1/2} \mathrm{d}t.$$

Definition 3 (Measure and integration theory on manifolds) *Unless noted otherwise, let ϕ denote a function $\phi : \Omega \mapsto \mathbb{R}$, and we state commonly used concepts in measure and integration theory.*

1. *On general metric d-manifolds, normalised Hausdorff measures (Federer, 1951, 1978), here denoted $H^d_\Omega(\mathrm{d}\boldsymbol{u})$, can be used to define areas. This leads to a natural generalisation of Lebesgue measure and integration, that coincides with the regular theory on $\mathbb{R}^d$. A common equivalent alternative is to define integration on manifolds by mapping to $\mathbb{R}^d$. We write the area of an d-dimensional Hausdorff measurable subset A of Ω as*

$$|A|_\Omega = H^d_\Omega(A) = \int_{\boldsymbol{u}\in A} H^d_\Omega(\mathrm{d}\boldsymbol{u}),$$

and the (Hausdorff or Lebesgue) integral of a function ϕ as

$$H^d_\Omega(\phi) = \int_{\boldsymbol{u}\in\Omega} \phi(\boldsymbol{u}) H^d_\Omega(\mathrm{d}\boldsymbol{u}).$$

2. *The inner product of scalar and vector valued functions with respect to Hausdorff measure on Ω is defined as*

$$\langle\phi, \psi\rangle_\Omega = H^d_\Omega(\phi^\mathsf{T}\psi) = \int_{\boldsymbol{u}\in\Omega} \phi(\boldsymbol{u})^\mathsf{T}\psi(\boldsymbol{u}) H^d_\Omega(\mathrm{d}\boldsymbol{u}).$$

A function $\phi : \Omega \mapsto \mathbb{R}^d$ is said to be square integrable if and only if $\langle\phi, \phi\rangle_\Omega < \infty$, denoted $\phi \in \mathrm{L}^2(\Omega)$.

3. *A fundamental relation, that corresponds to integration by parts for functions on $\mathbb{R}$, is* Green's first identity*: Whenever $\nabla\phi \in \mathrm{L}^2(\Omega)$ and $\Delta\psi \in \mathrm{L}^2(\Omega)$,*

$$\langle\phi, -\Delta\psi\rangle_\Omega = \langle\nabla\phi, \nabla\psi\rangle_\Omega - \langle\phi, \partial_{\boldsymbol{n}}\psi\rangle_{\partial\Omega}\,.$$

4. *If μ is a measure relative to the Hausdorff measure on Ω (i.e. μ is absolutely continuous with respect to the Hausdorff measure on Ω), and ϕ is integrable, these notational conventions are equivalent:*

$$\begin{aligned}\langle\phi, \mu\rangle_\Omega &= \langle\phi(\boldsymbol{u}), \mu(\boldsymbol{u})\rangle_{H^d_\Omega(\mathrm{d}\boldsymbol{u})} = \int_{\boldsymbol{u}\in\Omega} \phi(\boldsymbol{u})\mu(\boldsymbol{u}) H^d_\Omega(\mathrm{d}\boldsymbol{u}) \\ &= \int_{\boldsymbol{u}\in\Omega} \phi(\boldsymbol{u})\mu(\mathrm{d}\boldsymbol{u}) = \langle\phi(\boldsymbol{u}), 1\rangle_{\mu(\mathrm{d}\boldsymbol{u})} = \langle\phi, 1\rangle_\mu = \mu(\phi)\end{aligned}$$

5. *Convenient notation for the inner products with respect to a measure $\mu(\boldsymbol{u}, \boldsymbol{v})$ on a product space $\Omega_1 \times \Omega_2$,*

$$\langle\phi, \psi\rangle_\mu = \int_{\boldsymbol{u}\in\Omega_1}\int_{\boldsymbol{v}\in\Omega_2} \phi(\boldsymbol{u})^\mathsf{T}\psi(\boldsymbol{v})\mu(\mathrm{d}\boldsymbol{u}, \mathrm{d}\boldsymbol{v})$$

Definition 4 (Generalised Fourier representations) *The Fourier transform pair for functions* $\{\phi \in \mathrm{L}^2 : \mathbb{R}^d \mapsto \mathbb{R}\}$ *is given by*

$$\begin{cases} \widehat{\phi}(\boldsymbol{k}) = (\mathcal{F}\phi)(\boldsymbol{k}) = \dfrac{1}{(2\pi)^d} \left\langle \phi(\boldsymbol{u}), \mathrm{e}^{-i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{u})}, \\ \phi(\boldsymbol{u}) = (\mathcal{F}^{-1}\widehat{\phi})(\boldsymbol{u}) = \left\langle \widehat{\phi}(\boldsymbol{k}), \mathrm{e}^{i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k})}. \end{cases}$$

(Here, we briefly abused our notation by including complex functions in the inner products.)

If Ω *is a compact manifold, a countable subset* $\{E_k, k = 0, 1, 2, \ldots\}$ *of orthogonal and normalised eigenfunctions to the negated Laplacian,* $-\Delta E_k = \lambda_k E_k$*, can be chosen as basis, and the Fourier representation for functions* $\{\phi \in \mathrm{L}^2 : \Omega \mapsto \mathbb{R}\}$ *is given by*

$$\begin{cases} \widehat{\phi}(k) = (\mathcal{F}\phi)(k) = \langle \phi, E_k \rangle_\Omega, \\ \phi(\boldsymbol{u}) = (\mathcal{F}^{-1}\widehat{\phi})(\boldsymbol{u}) = \displaystyle\sum_{k=0}^{\infty} \widehat{\phi}(k) E_k(\boldsymbol{u}). \end{cases}$$

The normalisation for $\mathbb{R}^d$ was chosen for convenient scaling in Bochner's theorem, where a positive definite covariance function ϕ for some stationary random field is written as the inverse Fourier transform of some spectral power measure $\widehat{\phi}$. For the case $\Omega = \mathbb{S}^2$, a corresponding Bochner-adapted scaling can be obtained, see Appendix E. However, for more general manifolds it is not possible to define global stationarity.

Finally, we define a subspace of L^2 functions, with inner product adapted to the differential operators we will study in the remainder of this paper.

Definition 5 *The Hilbert space* $\mathcal{H}^1(\Omega, \kappa)$*, for a given* $\kappa \geq 0$*, is the space of functions* $\{\phi : \Omega \mapsto \mathbb{R}\}$ *with* $\nabla\phi \in \mathrm{L}^2(\Omega)$*, equipped with inner product*

$$\langle \phi, \psi \rangle_{\mathcal{H}^1(\Omega,\kappa)} = \kappa^2 \langle \phi, \psi \rangle_\Omega + \langle \nabla\phi, \nabla\psi \rangle_\Omega .$$

The inner product induces a norm, given by $\|\phi\|_{\mathcal{H}^1(\Omega,\kappa)} = \langle \phi, \phi \rangle^{1/2}_{\mathcal{H}^1(\Omega,\kappa)}$*. The boundary case* $\kappa = 0$ *is also well defined, since* $\|\phi\|_{\mathcal{H}^1(\Omega,0)}$ *is a semi-norm, and* $\mathcal{H}^1(\Omega, 0)$ *is a space of equivalence classes of functions, that can be identified by functions with* $\langle \phi, 1 \rangle_\Omega = 0$*.*

The Hilbert space $\mathcal{H}^1$ is a quintessential Sobolev space.

B.2 Gaussian measures

We now turn to the problem of characterising *random* measures on Ω. We restrict ourselves to Gaussian measures that are at most as irregular as white noise. The distributions of such Gaussian measures are determined by the properties of expectations and covariances of integrals of functions with respect to the random measures, the so called finite dimensional distributions.

Definition 6 *A* Gaussian random measure $\mathcal{E}$ *on* Ω *is a measure such that for every finite set of* test functions $\{\phi_k \in \mathrm{L}^2(\Omega) : \Omega \mapsto \mathbb{R}, k = 1, \ldots, n\}$*, the inner products*

$$\langle \phi_k, \mathcal{E} \rangle_\Omega , \quad k = 1, \ldots, n,$$

are jointly Gaussian. More precisely, if there is a constant $b \geq 0$ *such that* $\mathsf{Var}(\langle \phi, \mathcal{E} \rangle_\Omega) \leq b \|\phi\|^2_\Omega$ *for every* $\phi \in \mathrm{L}^2(\Omega)$*, we say that* $\mathcal{E}$ *is an* $\mathrm{L}^2(\Omega)$-bounded Gaussian measure*. For convenience, we denote the measure of a function* ϕ *by* $\mathcal{E}(\phi) = \langle \phi, \mathcal{E} \rangle_\Omega$*.*

Definition 7 Gaussian white noise $\mathcal{W}$ *on a manifold* Ω *is an* $\mathrm{L}^2(\Omega)$*-bounded random measure such that for any set of test functions* $\{\phi_i \in \mathrm{L}^2 : \Omega \mapsto \mathbb{R},\, i = 1, \ldots, n\}$*, the integrals* $\langle \phi_i, \mathcal{W} \rangle_\Omega$*,* $i = 1, \ldots, n$*, are jointly Gaussian, with expectation- and covariance-measures given by*

$$\begin{aligned} \mathsf{E}(\langle \phi_i, \mathcal{W} \rangle_\Omega) &= \mathsf{E}(\mathcal{W}(\phi_i)) = 0, \\ \mathsf{Cov}\big(\langle \phi_i, \mathcal{W} \rangle_\Omega \,,\, \langle \phi_j, \mathcal{W} \rangle_\Omega\big) &= \mathsf{Cov}(\mathcal{W}(\phi_i), \mathcal{W}(\phi_j)) = \langle \phi_i, \phi_j \rangle_\Omega \,. \end{aligned}$$

In particular, the covariance-measure of $\mathcal{W}$ *over two subregions* $A, B \subseteq \Omega$ *is equal to the area measure of their intersection,* $|A \cap B|_\Omega$*, so that the variance-measure of* $\mathcal{W}$ *over a region is proportional to the area.*

We note that the popular approach to defining white noise on $\mathbb{R}^d$ via a *Brownian sheet* is not applicable for general manifolds, since the notion of *globally orthogonal directions* is not present. The closest equivalent would be to define a set-indexed Gaussian random function $\mathcal{W}(A) : \{A; A \subseteq \Omega\} \mapsto \mathbb{R}$, such that $\mathsf{E}(\mathcal{W}(A)) = 0$ and $\mathsf{Cov}(\mathcal{W}(A), \mathcal{W}(B)) = |A \cap B|_\Omega$. This definition is equivalent to the one above, and the Brownian sheet is a special case that only considers rectangular regions along the axes of $\mathbb{R}^d$, with one corner fixed at the origin.

B.3 Operator identities

In defining and solving the considered SPDEs, the half-Laplacian operator needs to be characterised in a way that permits practical calculations on general manifolds, and the regular notion of integration by parts needs to be extended not only to manifolds, but to random measures that are not differentiable in the classic sense.

B.3.1 The half-Laplacian

The fractional modified Laplacian operators $(\kappa^2 - \Delta)^{\alpha/2}$, $\kappa, \alpha \geq 0$, are commonly (Samko et al., 1992, p. 483) defined through the Fourier transform, as defined above:

$$\begin{aligned} (\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2}\phi)(\boldsymbol{k}) &= (\kappa^2 + \boldsymbol{k}^\mathsf{T}\boldsymbol{k})^{\alpha/2}(\mathcal{F}\phi)(\boldsymbol{k}), \quad \text{on } \mathbb{R}^d \\ (\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2}\phi)(k) &= (\kappa^2 + \lambda_k)^{\alpha/2}(\mathcal{F}\phi)(k), \quad \text{on compact } \Omega. \end{aligned}$$

The formal definition if mostly of theoretical interest, since in practice, the generalised Fourier basis and eigenvalues for the Laplacian are unknown. The following Lemma provides an integration identity that allows practical calculations involving the half-Laplacian.

Lemma 1 *Let* ϕ *and* ψ *be functions in* $\mathcal{H}^1(\Omega, \kappa)$*. Then, the Fourier-based modified half-Laplacians satisfy*

$$\left\langle (\kappa^2 - \Delta)^{1/2}\phi, (\kappa^2 - \Delta)^{1/2}\psi \right\rangle_\Omega = \langle \phi, \psi \rangle_{\mathcal{H}^1(\Omega,\kappa)}$$

whenever either

1. $\Omega = \mathbb{R}^d$*,*
2. Ω *is closed, or*
3. Ω *is compact, and* $\langle \phi, \partial_{\boldsymbol{n}}\psi \rangle_{\partial\Omega} = \langle \partial_{\boldsymbol{n}}\phi, \psi \rangle_{\partial\Omega} = 0$*.*

It would be tempting to eliminate the qualifiers by subtracting the average of the two boundary integrals to the relation, and extend the Lemma to an equivalence relation. However, the motivation is problematic, since the half-Laplacian is defined for a wider class of functions than the Laplacian, and it is yet unclear whether such a generalisation yields the same half-Laplacian as the Fourier definition for functions that are not of the class $\Delta\phi \in \mathrm{L}^2(\Omega)$. An obvious solution would be to reduce the admitted class of functions, but that would eliminate the usefulness of the result, as we will later have that $\nabla\psi$ is an $\mathrm{L}^2(\Omega)$-bounded random measure. It is however clear that the Lemma can be used also for suitably well-behaved unbounded manifolds with no boundary. For such manifolds the issue is instead that of defining the Fourier representation.

B.3.2 Stochastic Green's first identity

Identities for differentiation and integration on manifolds are usually stated as requiring functions in C^1, C^2, or even C^∞, which is much too restrictive to be applied to general random measures, and unnecessarily so. We state without proof a generalisation of Green's first identity, noting that the identity applies to general Gaussian measures, as opposed to only differentiable functions.

Proposition 1 *If x and y are Gaussian measures on Ω such that ∇x and Δy are $\mathrm{L}^2(\Omega)$-bounded,*

$$\langle x, -\Delta y\rangle_\Omega = \langle \nabla x, \nabla y\rangle_\Omega - \langle x, \partial_{\boldsymbol{n}} y\rangle_{\partial\Omega}\,.$$

C Hilbert space approximation

We are now ready to formulate the main results of the paper in more technical detail. The idea is to approximate the full SPDE solutions with functions in finite Hilbert spaces, showing that the approximations converge to the true solutions as the finite Hilbert space approaches the full space. In Section C.1, we state the needed convergence and stochastic Finite Element definitions. The main result for Matérn covariance models is stated in Section C.2, followed by generalisations to intrinsic and oscillating fields in Section C.3 and Section C.4. Finally, the full Finite Element constructions are modified to Markov models in Section C.5.

C.1 Weak convergence and stochastic FEM

The following definitions concern the formal definitions of convergence of Hilbert spaces and random measures in such spaces (Definition 8 and 9) and the definition of the Finite Element constructions that will be used (Definition 10).

Definition 8 *A finite subspace $\mathcal{H}^1_n(\Omega,\kappa) \subset \mathcal{H}^1(\Omega,\kappa)$ is spanned by a finite set of basis functions $\Psi_n = \{\psi_1,\dots,\psi_n\}$. We say that a sequence of subspaces $\{\mathcal{H}^1_n\}$ is dense in $\mathcal{H}^1$ if for any $f \in \mathcal{H}^1$ there is a sequence $\{f_n\}$, $f_n \in \mathcal{H}^1_n$, such that $\lim_{n\to\infty} \|f - f_n\|^2_{\mathcal{H}^1(\Omega,\kappa)} = 0$.*

Definition 9 *A sequence of $\mathrm{L}^2(\Omega)$-bounded Gaussian random measures $\{x_n\}$ is said to* converge weakly *to an $\mathrm{L}^2(\Omega)$-bounded Gaussian random measure x if for all $f, g \in \mathrm{L}^2(\Omega)$,*

$$\begin{aligned}\mathsf{E}\left(\langle f, x_n\rangle_\Omega\right) &\to \mathsf{E}\left(\langle f, x\rangle_\Omega\right),\\ \mathsf{Cov}\left(\langle f, x_n\rangle_\Omega, \langle g, x_n\rangle_\Omega\right) &\to \mathsf{Cov}\left(\langle f, x\rangle_\Omega, \langle g, x\rangle_\Omega\right),\end{aligned}$$

as $n\to\infty$. We denote such convergence $x_n \xrightarrow[n\to\infty]{D(\mathrm{L}^2(\Omega))} x$.

Definition 10 *Let $\mathcal{L}$ be a linear differential operator, and let $\mathcal{E}$ be a Gaussian random measure on Ω. Let $x_n = \sum_j \psi_j w_j \in \mathcal{H}^1_n(\Omega,\kappa)$ denote approximate weak solutions to the SPDE $\mathcal{L}x = \mathcal{E}$ on Ω.*

a) The weak Galerkin approximations *are given by random $\boldsymbol{w} = \{w_1,\dots,w_n\}$ such that*

$$\begin{aligned}\mathsf{E}(\langle f_n, \mathcal{L}x_n\rangle_\Omega) &= \mathsf{E}(\langle f_n, \mathcal{E}\rangle_\Omega)\\ \mathsf{Cov}(\langle f_n, \mathcal{L}x_n\rangle_\Omega, \langle g_n, \mathcal{L}x_n\rangle_\Omega) &= \mathsf{Cov}(\langle f_n, \mathcal{E}\rangle_\Omega, \langle g_n, \mathcal{E}\rangle_\Omega)\end{aligned}$$

for any pair of test functions $f_n, g_n \in \mathcal{H}^1_n(\Omega,\kappa)$.

b) The weak least squares solutions *are given by random $\boldsymbol{w} = \{w_1,\dots,w_n\}$ such that*

$$\begin{aligned}\mathsf{E}(\langle \mathcal{L}f_n, \mathcal{L}x_n\rangle_\Omega) &= \mathsf{E}(\langle \mathcal{L}f_n, \mathcal{E}\rangle_\Omega)\\ \mathsf{Cov}(\langle \mathcal{L}f_n, \mathcal{L}x_n\rangle_\Omega, \langle \mathcal{L}g_n, \mathcal{L}x_n\rangle_\Omega) &= \mathsf{Cov}(\langle \mathcal{L}f_n, \mathcal{E}\rangle_\Omega, \langle \mathcal{L}g_n, \mathcal{E}\rangle_\Omega)\end{aligned}$$

for any pair of test functions $f_n, g_n \in \mathcal{H}^1_n(\Omega,\kappa)$.

C.2 The basic Matérn-like cases

In the classic Matérn case, the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} x = \mathcal{W} \tag{47}$$

can be unravelled into a recursive formulation

$$\begin{aligned}(\kappa^2 - \Delta)^{1/2} y_1 &= \mathcal{W}, \\ (\kappa^2 - \Delta) y_2 &= \mathcal{W}, \\ (\kappa^2 - \Delta) y_k &= y_{k-2}, \quad k = 3, 4, \ldots, \alpha.\end{aligned}$$

For integers $\alpha = 1, 2, 3, \ldots$, y_α is a solution to the original SPDE. To avoid solutions in the null-space of $(\kappa^2 - \Delta)$, we require Neumann boundaries, i.e. the solutions must have zero normal derivatives at the boundary of Ω. In the Hilbert space approximation, this can be achieved by requiring that all basis functions have zero normal derivatives. For piecewise linear basis functions, arbitrarily small regions near the boundary of Ω can be altered to fulfil the requirement.

We now formulate the three main theorems of the paper, that show what the precision matrices should look like for given basis functions (Theorem 2), that the finite Hilbert representations converge to the true distributions for $\alpha = 1$ and $\alpha = 2$ and dense Hilbert space sequences (Theorem 3), and finally that the iterative constructions for $\alpha \geq 3$ also converge (Theorem 4). Note that a sequence $\mathcal{H}^1_n(\Omega, \kappa)$ of piecewise linear Hilbert spaces defined on triangulations of Ω is a dense sequence in $\mathcal{H}^1(\Omega, \kappa)$ if the maximal triangle diameter decreases to 0. Thus, the theorems are applicable for piecewise linear basis functions, showing weak convergence of the field itself and its derivatives up to order $\min(2, \alpha)$.

Theorem 2 *Define matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ through*

$$C_{i,j} = \langle \psi_i, \psi_j \rangle_\Omega, \quad G_{i,j} = \langle \nabla\psi_i, \nabla\psi_j \rangle_\Omega,$$

and denote the distribution for $\boldsymbol{w}$ with $\mathsf{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1})$, where the precision matrix $\boldsymbol{Q}$ is the inverse of the covariance matrix, and let $x_n = \sum_k \psi_k w_k$ be a weak $\mathcal{H}^1_n(\Omega, \kappa)$ approximation to $\mathcal{L}x = \mathcal{E}$ with Neumann boundaries, and $\partial_{\boldsymbol{n}} \psi_k = 0$ on $\partial\Omega$.

a) When $\mathcal{L} = \kappa^2 - \Delta$ and $\mathcal{E}$ is Gaussian white noise, the weak Galerkin approximation is obtained for $\boldsymbol{Q} = \boldsymbol{K}^\mathsf{T} \boldsymbol{C}^{-1} \boldsymbol{K}$, where $\boldsymbol{K} = \kappa^2 \boldsymbol{C} + \boldsymbol{G}$.

b) When $\mathcal{L} = (\kappa^2 - \Delta)^{1/2}$ and $\mathcal{E}$ is Gaussian white noise, the weak least squares approximation is obtained for $\boldsymbol{Q} = \kappa^2 \boldsymbol{C} + \boldsymbol{G}$.

c) When $\mathcal{L} = \kappa^2 - \Delta$ and $\mathcal{E}$ is a Gaussian measure on $\mathcal{H}^1_n(\Omega, \kappa)$ with mean zero and precision $\boldsymbol{Q}_{\mathcal{E},n}$, the weak Galerkin approximation is obtained for $\boldsymbol{Q} = \boldsymbol{K}^\mathsf{T} \boldsymbol{C}^{-1} \boldsymbol{Q}_{\mathcal{E},n} \boldsymbol{C}^{-1} \boldsymbol{K}$, where $\boldsymbol{K} = \kappa^2 \boldsymbol{C} + \boldsymbol{G}$.

Theorem 3 *Let x be a weak solution to an SPDE $\mathcal{L}x = \mathcal{W}$ with Neumann boundaries on a manifold Ω, and let x_n be a weak $\mathcal{H}^1_n(\Omega, \kappa)$ approximation, when $\mathcal{E}$ is an $\mathrm{L}^2(\Omega)$-bounded random measure. Then,*

$$x_n \xrightarrow[n\to\infty]{D(\mathrm{L}^2(\Omega))} x, \tag{I}$$

$$\mathcal{L}x_n \xrightarrow[n\to\infty]{D(\mathrm{L}^2(\Omega))} \mathcal{L}x, \tag{II}$$

if the sequence $\{\mathcal{H}^1_n(\Omega, \kappa), n \to \infty\}$ is dense in $\mathcal{H}^1(\Omega, \kappa)$, and either

1. *$\mathcal{L} = (\kappa^2 - \Delta)$, and x_n is the Galerkin approximation, or*
2. *$\mathcal{L} = (\kappa^2 - \Delta)^{1/2}$ and x_n is the least squares approximation.*

Theorem 4 *Let y be a weak solution to an linear SPDE $\mathcal{L}_y y = \mathcal{E}$ on a manifold Ω, for some $L^2(\Omega)$-bounded random measure $\mathcal{E}$, and let x be a weak solution to the SPDE $\mathcal{L}_y\mathcal{L}_x x = \mathcal{E}$, where $\mathcal{L}_x = \kappa^2 - \Delta$. Further, let y_n be a weak $\mathcal{H}^1_n(\Omega, \kappa)$ approximation to y such that*

$$y_n \xrightarrow[n\to\infty]{D(\mathrm{L}^2(\Omega))} y, \tag{I}$$

and let x_n be the weak Galerkin approximation in $\mathcal{H}^1_n(\Omega, \kappa)$ to the SPDEs $\mathcal{L}_x x = y_n$ on Ω. Then,

$$x_n \xrightarrow[n\to\infty]{D(\mathrm{L}^2(\Omega))} x. \tag{II}$$

$$\mathcal{L}_x x_n \xrightarrow[n\to\infty]{D(\mathrm{L}^2(\Omega))} \mathcal{L}_x x. \tag{III}$$

C.3 The intrinsic cases

When $\kappa = 0$, the Hilbert space from Definition 5 is a space of equivalence classes of functions, corresponding to SPDE solutions where arbitrary functions in the null-space of $(-\Delta)^{\alpha/2}$ can be added. Such solution fields are known as *intrinsic* fields, and have well-defined properties. With piecewise linear basis functions, the intrinsicness can be exactly reproduced for $\alpha = 1$ for all manifolds, and for $\alpha = 2$ on subsets of $\mathbb{R}^2$, by relaxing the boundary constraints to free boundaries. For larger α or more general manifolds, the intrinsicness will only be approximately represented. To approximate intrinsic fields, the matrix $\boldsymbol{K}$ in Theorem 2 should be replaced by $\boldsymbol{G} - \boldsymbol{B}$ (due too Green's identity), where the elements of the (possibly asymmetric) boundary integral matrix $\boldsymbol{B}$ are given by

$$B_{i,j} = \langle \psi_i, \partial_{\boldsymbol{n}} \psi_j \rangle_{\partial\Omega}.$$

The formulations and proofs of Theorem 3 and Theorem 4 remain unchanged, but with the convergence only defined with respect to test functions f and g orthogonal to the null-space of the linear SPDE operator.

The notion of non-null-space convergence allows us to formulate a simple proof of the result from Besag and Mondal (2005), that says that a 1st order intrinsic CAR model on infinite lattices in $\mathbb{R}^2$ converge to the measure valued *de Wij* process. As seen in Section A.1, for $\alpha = 1$ and $\kappa = 0$, the $\boldsymbol{Q}$-matrix for a triangulated regular grid matches the ordinary intrinsic 1st order CAR model. The null-space of the half-Laplacian are constant functions. Choose non-trivial test functions f and g that integrate to zero, and apply Theorem 3 and Definition 9. This shows that the regular CAR model, seen as a Hilbert space representation with linear basis functions, converges to the de Wij process, which is the special SPDE case $\alpha = 1$, $\kappa = 0$ in $\mathbb{R}^2$.

C.4 The oscillating and non-isotropic cases

To construct the Hilbert space approximation for the oscillating model introduced in Section 3.3, as well as non-isotropic versions, we introduce a coupled system of SPDEs for $\alpha = 2$,

$$\begin{bmatrix} h_1 - \nabla^{\mathsf{T}} \boldsymbol{H}_1 \nabla & -h_2 + \nabla^{\mathsf{T}} \boldsymbol{H}_2 \nabla \\ h_2 - \nabla^{\mathsf{T}} \boldsymbol{H}_2 \nabla & h_1 - \nabla^{\mathsf{T}} \boldsymbol{H}_1 \nabla \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \end{bmatrix} \tag{48}$$

which is equivalent to the complex SPDE

$$(h_1 + ih_2 - \nabla^{\mathsf{T}}(\boldsymbol{H}_1 + i\boldsymbol{H}_2)\nabla)(x_1(\boldsymbol{u}) + ix_2(\boldsymbol{u})) = \mathcal{E}_1(\boldsymbol{u}) + i\mathcal{E}_2(\boldsymbol{u}). \tag{49}$$

The model in Section 3.3 corresponds to $h_1 = \kappa^2 \cos(\pi\theta)$, $h_2 = \kappa^2 \sin(\pi\theta)$, $\boldsymbol{H}_1 = \boldsymbol{I}$, and $\boldsymbol{H}_2 = \boldsymbol{0}$.

To solve the coupled SPDE system (48) we take a set $\{\psi_k, k = 1, \ldots, n\}$ of basis functions for $\mathcal{H}^1_n(\Omega, \kappa)$ and construct a basis for the solution space for $\begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathsf{T}}$ as

$$\begin{bmatrix} \psi_1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} \psi_n \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \psi_1 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ \psi_n \end{bmatrix}.$$

The definitions of the $\boldsymbol{G}$ and $\boldsymbol{K}$ matrices are modified as follows:

$$(\boldsymbol{G}_k)_{i,j} = \left\langle \boldsymbol{H}_k^{1/2}\nabla\psi_i, \boldsymbol{H}_k^{1/2}\nabla\psi_j \right\rangle_\Omega, \quad k = 1, 2,$$
$$\boldsymbol{K}_k = h_k\boldsymbol{C} + \boldsymbol{G}_k, \quad k = 1, 2.$$

Using the same construction as in the regular case, the precision for the solutions are given by

$$\begin{bmatrix} \boldsymbol{K}_1 & -\boldsymbol{K}_2 \\ \boldsymbol{K}_2 & \boldsymbol{K}_1 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{Q}_{\mathcal{E}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_{\mathcal{E}} \end{bmatrix} \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{K}_1 & -\boldsymbol{K}_2 \\ \boldsymbol{K}_2 & \boldsymbol{K}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{bmatrix},$$

where $\boldsymbol{Q} = \boldsymbol{Q}_2(h_1, \boldsymbol{H}_1) + \boldsymbol{Q}_2(h_2, \boldsymbol{H}_2)$ and $\boldsymbol{Q}_2(\cdot,\cdot)$ is the precision generated for the regular Matérn model with $\alpha = 2$ and the given parameters. Surprisingly, regardless of the choice of parameters, the solution components are independent.

C.5 Markov approximation

By choosing piecewise linear basis functions, the practical calculation of the matrix elements in the precision construction is straightforward, and the local support make the basic matrices sparse. However, since they are not orthogonal, the $\boldsymbol{C}$ matrix will be non-diagonal, and therefore the FEM-construction does not directly yield Markov fields for $\alpha \geq 2$, since $\boldsymbol{C}^{-1}$ is not sparse. However, $\boldsymbol{C}$ can be approximated with a diagonal matrix as follows. Let $\widetilde{\boldsymbol{C}}$ be a diagonal matrix, with $\widetilde{C}_{ii} = \sum_j C_{ij} = \langle \phi_i, 1\rangle_\Omega$, and note that this can be interpreted as a lower order integration matrix. Substituting $\boldsymbol{C}^{-1}$ with $\widetilde{\boldsymbol{C}}^{-1}$ yields a Markov approximation to the FEM solution.

Let f and g be test functions in $\mathcal{H}^1(\Omega, \kappa)$ and let f_n and g_n be their projections onto $\mathcal{H}^1_n(\Omega, \kappa)$, with basis weights $\boldsymbol{w}_f$ and $\boldsymbol{w}_g$. Taking the difference between the covariances for the FEM and the Markov models in the worst case scenario $\kappa \to \infty$ yields

$$\boldsymbol{w}_f(\boldsymbol{C} - \widetilde{\boldsymbol{C}})\boldsymbol{w}_g \to 0$$

as $n \to \infty$. An even better Markov approximation is obtained by also replacing the $\boldsymbol{C}$ matrix in $\boldsymbol{K}$ with $\widetilde{\boldsymbol{C}}$. In this case, the covariance difference is

$$\boldsymbol{w}_f(\boldsymbol{C} - \boldsymbol{C}\widetilde{\boldsymbol{C}}^{-1}\boldsymbol{C})\boldsymbol{w}_g$$

which has a smaller error in numerical comparisons. The intuitive explanation for this is that the approximation errors cancel, but further study is needed to determine the precise effects of the approximation. See Bolin and Lindgren (2009b) for a comparison of the resulting Kriging errors for different methods, showing negligible differences between the exact FEM representation and the Markov approximation.

D Proofs

D.1 Folded covariance

Proof: [Theorem 1] Writing the covariance of the SPDE solutions on the interval $\Omega = [0, L] \subset \mathbb{R}$ in terms of the spectral representation gives an infinite series,

$$\mathsf{Cov}(x(u), x(v)) = \lambda_0 + \sum_{k=1}^{\infty} \cos(u\pi k/L)\cos(v\pi k/L)\lambda_k, \tag{50}$$

where $\lambda_0 = (\kappa^{2\alpha}L)^{-1}$ and $\lambda_k = 2L^{-1}(\kappa^2 + (\pi k/L)^2)^{-\alpha}$ are the variances of the weights for the basis functions $\cos(u\pi k/L)$, $k = 0, 1, 2, \ldots$.

We use the spectral representation of the Matérn covariance in the theorem statement, and show that the resulting expression is equal to the spectral representation of the covariance for the solutions to the given SPDE. First, let $\widetilde{r}(u,v)$ denote the folded covariance in the theorem statement, and note that on $\mathbb{R}$, the Matérn covariance (with the given variance) can be written as

$$r_M(u,v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (\kappa^2+\omega^2)^{-\alpha}\cos((v-u)\omega)\,\mathrm{d}\omega.$$

Thus,

$$\begin{aligned}
\widetilde{r}(u,v) &= \sum_{k=-\infty}^{\infty} \left(r_M(u,v-2kL) + r_M(u,2kL-v)\right) \\
&= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int_{-\infty}^{\infty} (\kappa^2+\omega^2)^{-\alpha}\left(\cos((v-u-2kL)\omega) + \cos((v+u-2kL)\omega)\right)\,\mathrm{d}\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} (\kappa^2+\omega^2)^{-\alpha}\sum_{k=-\infty}^{\infty}\left(\cos((v-u-2kL)\omega) + \cos((v+u-2kL)\omega)\right)\,\mathrm{d}\omega
\end{aligned}$$

Rewriting the cosines via Euler's formulas, we obtain

$$\begin{aligned}
&\sum_{k=-\infty}^{\infty} \left(\cos((v-u-2kL)\omega) + \cos((v+u-2kL)\omega)\right) \\
&\quad= \frac{1}{2}\sum_{k=-\infty}^{\infty}\left(\mathrm{e}^{i(v-u-2kL)\omega} + \mathrm{e}^{-i(v-u-2kL)\omega} + \mathrm{e}^{i(v+u-2kL)\omega} + \mathrm{e}^{-i(v+u-2kL)\omega}\right) \\
&\quad= \frac{1}{2}\sum_{k=-\infty}^{\infty}\left(\mathrm{e}^{iu\omega} + \mathrm{e}^{-iu\omega}\right)\left(\mathrm{e}^{i(v-2kL)\omega} + \mathrm{e}^{-i(v-2kL)\omega}\right) \\
&\quad= \cos(u\omega)\left(\mathrm{e}^{iv\omega}\sum_{k=-\infty}^{\infty}\mathrm{e}^{-2ikL\omega} + \mathrm{e}^{-iv\omega}\sum_{k=-\infty}^{\infty}\mathrm{e}^{2ikL\omega}\right) \\
&\quad= 2\pi\cos(u\omega)\left(\mathrm{e}^{iv\omega} + \mathrm{e}^{-iv\omega}\right)\sum_{k=-\infty}^{\infty}\delta(2L\omega - 2\pi k) \\
&\quad= \frac{2\pi}{L}\cos(u\omega)\cos(v\omega)\sum_{k=-\infty}^{\infty}\delta(\omega - \pi k/L)
\end{aligned}$$

where we used the Dirac-measure representation $\sum_{k=-\infty}^{\infty} e^{iks} = 2\pi\sum_{k=-\infty}^{\infty}\delta(s-2\pi k)$. Finally, combining the results yields

$$\begin{aligned}
\widetilde{r}(u,v) &= \frac{1}{L}\int_{-\infty}^{\infty}(\kappa^2+\omega^2)^{-\alpha}\cos(u\omega)\cos(v\omega)\sum_{k=-\infty}^{\infty}\delta(\omega-\pi k/L)\,\mathrm{d}\omega \\
&= \frac{1}{L}\sum_{k=-\infty}^{\infty}(\kappa^2+(\pi k/L)^2)^{-\alpha}\cos(u\pi k/L)\cos(v\pi k/L) \\
&= \frac{1}{\kappa^{2\alpha}L} + \frac{2}{L}\sum_{k=1}^{\infty}(\kappa^2+(\pi k/L)^2)^{-\alpha}\cos(u\pi k/L)\cos(v\pi k/L),
\end{aligned}$$

which is precisely the sought expression in (50). □

D.2 Modified half-Laplacian equivalence

Proof: [Lemma 1] We begin with the case $\Omega = \mathbb{R}^d$. First, we note that $\nabla \mathrm{e}^{i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}} = i\boldsymbol{k}\mathrm{e}^{i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}}$. We use the Fourier representations of ϕ and ψ and change the order of integration in the Hilbert space inner product:

$$\begin{aligned}
\langle \phi, \psi \rangle_{\mathcal{H}^1(\mathbb{R}^d,\kappa)} &= \left\langle \left\langle \widehat{\phi}(\boldsymbol{k}), \mathrm{e}^{i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k})}, \left\langle \widehat{\psi}(\boldsymbol{k'}), \mathrm{e}^{i\boldsymbol{k'}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k'})} \right\rangle_{\mathcal{H}^1(\mathbb{R}^d(\mathrm{d}\boldsymbol{u}),\kappa)} \\
&= \left\langle \widehat{\phi}(\boldsymbol{k})\widehat{\psi}(\boldsymbol{k'}), \left\langle \mathrm{e}^{i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}}, \mathrm{e}^{i\boldsymbol{k'}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathcal{H}^1(\mathbb{R}^d(\mathrm{d}\boldsymbol{u}),\kappa)} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k})\times\mathbb{R}^d(\mathrm{d}\boldsymbol{k'})} \\
&= \left\langle \widehat{\phi}(\boldsymbol{k})\widehat{\psi}(\boldsymbol{k'}), (\kappa^2 - \boldsymbol{k}^\mathsf{T}\boldsymbol{k'}) \left\langle \mathrm{e}^{i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}}, \mathrm{e}^{i\boldsymbol{k'}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{u})} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k})\times\mathbb{R}^d(\mathrm{d}\boldsymbol{k'})} \\
&= \left\langle \widehat{\phi}(\boldsymbol{k})\widehat{\psi}(-\boldsymbol{k}), \kappa^2 + \boldsymbol{k}^\mathsf{T}\boldsymbol{k} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k})}
\end{aligned}$$

where we used the identity $\left\langle \mathrm{e}^{i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}}, \mathrm{e}^{i\boldsymbol{k'}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{u})} = \delta(\boldsymbol{k}+\boldsymbol{k'})$, where $\delta(\cdot)$ is a Dirac delta measure. Similarly, the Fourier definition can be rewritten as

$$\begin{aligned}
&\left\langle (\kappa^2 - \Delta)^{1/2}\phi, (\kappa^2 - \Delta)^{1/2}\psi \right\rangle_{\mathbb{R}^d} \\
&= \left\langle \left\langle (\kappa^2 + \boldsymbol{k}^\mathsf{T}\boldsymbol{k})^{1/2}\widehat{\phi}(\boldsymbol{k}), \mathrm{e}^{i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k})}, \left\langle (\kappa^2 + \boldsymbol{k'}^\mathsf{T}\boldsymbol{k'})^{1/2}\widehat{\psi}(\boldsymbol{k'}), \mathrm{e}^{i\boldsymbol{k'}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k'})} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{u})} \\
&= \left\langle (\kappa^2 + \boldsymbol{k}^\mathsf{T}\boldsymbol{k})^{1/2}(\kappa^2 + \boldsymbol{k'}^\mathsf{T}\boldsymbol{k'})^{1/2}\widehat{\phi}(\boldsymbol{k})\widehat{\psi}(\boldsymbol{k'}), \left\langle \mathrm{e}^{i\boldsymbol{k}^\mathsf{T}\boldsymbol{u}}, \mathrm{e}^{i\boldsymbol{k'}^\mathsf{T}\boldsymbol{u}} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{u})} \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k})\times\mathbb{R}^d(\mathrm{d}\boldsymbol{k'})} \\
&= \left\langle (\kappa^2 + \boldsymbol{k}^\mathsf{T}\boldsymbol{k})\widehat{\phi}(\boldsymbol{k})\widehat{\psi}(-\boldsymbol{k}), 1 \right\rangle_{\mathbb{R}^d(\mathrm{d}\boldsymbol{k})},
\end{aligned}$$

so the two definitions are equivalent.

The proof for compact manifolds is similar. Let $\lambda_k \geq 0$ be the eigenvalue corresponding to eigenfunction E_k in Definition 4. Then, the modified half-Laplacian is defined by $(\kappa^2 - \Delta)^{1/2}\phi = \mathcal{F}^{-1}((\kappa^2 + \lambda_k)^{1/2}\mathcal{F}\phi)$, and we obtain

$$\left\langle (\kappa^2 - \Delta)^{1/2}\phi, (\kappa^2 - \Delta)^{1/2}\psi \right\rangle_\Omega = \left\langle \sum_{k=0}^{\infty} (\kappa^2 + \lambda_k)^{1/2}\widehat{\phi}(k)E_k, \sum_{k'=0}^{\infty} (\kappa^2 + \lambda_{k'})^{1/2}\widehat{\phi}(k')E_{k'} \right\rangle_\Omega,$$

and, since $\phi, \psi \in \mathcal{H}^1(\Omega,\kappa)$, we can change the order of integration and summation,

$$\begin{aligned}
&= \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} (\kappa^2 + \lambda_k)^{1/2}(\kappa^2 + \lambda_{k'})^{1/2}\widehat{\phi}(k)\widehat{\phi}(k') \left\langle E_k, E_{k'} \right\rangle_\Omega. \\
&= \sum_{k=0}^{\infty} (\kappa^2 + \lambda_k)\widehat{\phi}(k)\widehat{\phi}(k),
\end{aligned}$$

since the eigenfunctions E_k and $E_{k'}$ are orthonormal.

Now, starting from the Hilbert space inner product,

$$\begin{aligned}
\langle \phi, \psi \rangle_{\mathcal{H}^1(\Omega,\kappa)} &= \kappa^2 \langle \phi, \psi \rangle_\Omega + \langle \nabla\phi, \nabla\psi \rangle_\Omega \\
&= \kappa^2 \left\langle \sum_{k=0}^{\infty} \widehat{\phi}(k)E_k, \sum_{k'=0}^{\infty} \widehat{\psi}(k')E_{k'} \right\rangle_\Omega + \left\langle \nabla \sum_{k=0}^{\infty} \widehat{\phi}(k)E_k, \nabla \sum_{k'=0}^{\infty} \widehat{\psi}(k')E_{k'} \right\rangle_\Omega
\end{aligned}$$

and, since $\phi, \psi \in \mathcal{H}^1(\Omega,\kappa)$ and $E_k, E_{k'} \in \mathrm{L}^2(\Omega)$, we can change the order of differentiation and summation,

$$= \kappa^2 \left\langle \sum_{k=0}^{\infty} \widehat{\phi}(k)E_k, \sum_{k'=0}^{\infty} \widehat{\psi}(k')E_{k'} \right\rangle_\Omega + \left\langle \sum_{k=0}^{\infty} \widehat{\phi}(k)\nabla E_k, \sum_{k'=0}^{\infty} \widehat{\psi}(k')\nabla E_{k'} \right\rangle_\Omega$$

and, since in addition $\nabla E_k, \nabla E_{k'} \in \mathrm{L}^2(\Omega)$, we can change the order of summation and integration,

$$= \kappa^2 \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} \widehat{\phi}(k)\widehat{\psi}(k') \left\langle E_k, E_{k'}\right\rangle_{\Omega} + \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} \widehat{\phi}(k)\widehat{\psi}(k') \left\langle \nabla E_k, \nabla E_{k'}\right\rangle_{\Omega}.$$

Further, Green's identity for $\langle \nabla E_k, \nabla E_{k'}\rangle_{\Omega}$ yields

$$\begin{aligned}\langle \nabla E_k, \nabla E_{k'}\rangle_{\Omega} &= \langle E_k, -\Delta E_{k'}\rangle_{\Omega} + \langle E_k, \partial_{\boldsymbol{n}} E_{k'}\rangle_{\partial\Omega} \\ &= \lambda_{k'} \left\langle E_k, E_{k'}\right\rangle_{\Omega} + \langle E_k, \partial_{\boldsymbol{n}} E_{k'}\rangle_{\partial\Omega}.\end{aligned}$$

Since $\nabla\phi, \nabla\psi \in \mathrm{L}^2(\Omega)$ we can change the order of summation, integration and differentiation for the boundary integrals,

$$\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} \widehat{\phi}(k)\widehat{\psi}(k') \left\langle E_k, \partial_{\boldsymbol{n}} E_{k'}\right\rangle_{\partial\Omega} = \langle \phi, \partial_{\boldsymbol{n}}\psi\rangle_{\partial\Omega}.$$

By the boundary requirements in the Lemma, whenever Green's identity holds, the boundary integral vanishes, either because the boundary is empty, or the integrand is zero, so collecting all the terms, we obtain

$$\begin{aligned}\langle \phi, \psi\rangle_{\mathcal{H}^1(\Omega,\kappa)} &= \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} (\kappa^2 + \lambda'_k)\widehat{\phi}(k)\widehat{\psi}(k') \left\langle E_k, E_{k'}\right\rangle_{\Omega} + 0 \\ &= \sum_{k=0}^{\infty} (\kappa^2 + \lambda_k)\widehat{\phi}(k)\widehat{\psi}(k),\end{aligned}$$

and the proof is complete. □

D.3 Hilbert space convergence

Proof: [Theorem 2] The proofs are straightforward applications of the definitions. Let $\boldsymbol{w}_f$ and $\boldsymbol{w}_g$ be the Hilbert space coordinates of two test functions $f_n, g_n \in \mathcal{H}^1_n(\Omega, \kappa)$.

a) When $\mathcal{L} = \kappa^2 - \Delta$ and $\mathcal{E} = \mathcal{W}$,

$$\langle f_n, \mathcal{L}x_n\rangle_{\Omega} = \sum_{i,j} w_{f,i} \left\langle \psi_i, \mathcal{L}\psi_j\right\rangle_{\Omega} w_j = \sum_{i,j} w_{f,i}(\kappa^2 C_{i,j} + G_{i,j})w_j = \boldsymbol{w}_f^{\mathsf{T}}\boldsymbol{K}\boldsymbol{w}$$

due to Green's identity, so that

$$\mathsf{Cov}(\left\langle f_n, \mathcal{L}x_n\right\rangle_{\Omega}, \langle g_n, \mathcal{L}x_n\rangle_{\Omega}) = \boldsymbol{w}_f^{\mathsf{T}}\boldsymbol{K}\,\mathsf{Cov}(\boldsymbol{w}, \boldsymbol{w})\boldsymbol{K}^{\mathsf{T}}\boldsymbol{w}_g.$$

This covariance is equal to

$$\mathsf{Cov}(\left\langle f_n, \mathcal{W}\right\rangle_{\Omega}, \langle g_n, \mathcal{W}\rangle_{\Omega} = \langle f_n, g_n\rangle_{\Omega} = \sum_{i,j} w_{f,i} \left\langle \psi_i, \psi_j\right\rangle_{\Omega} w_{g,j} = \sum_{i,j} w_{f,i} C_{i,j} w_{g,j} = \boldsymbol{w}_f^{\mathsf{T}}\boldsymbol{C}\boldsymbol{w}_g$$

for every pair of test functions f_n, g_n when

$$\begin{aligned}\mathsf{Cov}(\boldsymbol{w}, \boldsymbol{w}) = \boldsymbol{K}^{-1}\boldsymbol{C}\boldsymbol{K}^{-\mathsf{T}}, \quad &\text{i.e. when} \\ \boldsymbol{Q} = \boldsymbol{K}^{\mathsf{T}}\boldsymbol{C}^{-1}\boldsymbol{K}.&\end{aligned}$$

b) Now, $\mathcal{L} = (\kappa^2 - \Delta)^{1/2}$ and $\mathcal{E} = \mathcal{W}$. Using the same technique as in a), but with Lemma 1 instead of Green's identity,

$$\langle \mathcal{L}f_n, \mathcal{L}x_n \rangle_\Omega = \boldsymbol{w}_f^\mathsf{T} \boldsymbol{K} \boldsymbol{w}$$

and

$$\mathsf{Cov}(\langle \mathcal{L}f_n, \mathcal{W} \rangle_\Omega, \langle \mathcal{L}g_n, \mathcal{W} \rangle_\Omega = \langle \mathcal{L}f_n, \mathcal{L}g_n \rangle_\Omega = \boldsymbol{w}_f^\mathsf{T} \boldsymbol{K} \boldsymbol{w}_g$$

so that

$$\begin{aligned} \mathsf{Cov}(\boldsymbol{w}, \boldsymbol{w}) = \boldsymbol{K}^{-1} \boldsymbol{K} \boldsymbol{K}^{-\mathsf{T}}, \quad \text{i.e.} \\ \boldsymbol{Q} = \boldsymbol{K}, \end{aligned}$$

noting that $\boldsymbol{K}$ is a symmetric matrix since both $\boldsymbol{C}$ and $\boldsymbol{G}$ are symmetric.

c) Now, $\mathcal{L} = \kappa^2 - \Delta$ and $\mathcal{E} = \mathcal{E}_n$ is a Gaussian measure on $\mathcal{H}_n^1(\Omega, \kappa)$ with precision $\boldsymbol{Q}_{\mathcal{E},n}$. Using the same technique as in a),

$$\mathsf{Cov}(\langle f_n, \mathcal{L}x_n \rangle_\Omega, \langle g_n, \mathcal{L}x_n \rangle_\Omega) = \boldsymbol{w}_f^\mathsf{T} \boldsymbol{K} \mathsf{Cov}(\boldsymbol{w}, \boldsymbol{w}) \boldsymbol{K}^\mathsf{T} \boldsymbol{w}_g$$

and

$$\mathsf{Cov}(\langle f_n, \mathcal{E}_n \rangle_\Omega, \langle g_n, \mathcal{E}_n \rangle_\Omega = \boldsymbol{w}_f^\mathsf{T} \boldsymbol{C} \boldsymbol{Q}_{\mathcal{E},n}^{-1} \boldsymbol{C} \boldsymbol{w}_g.$$

Requiring equality for all pairs of test functions yields

$$\begin{aligned} \mathsf{Cov}(\boldsymbol{w}, \boldsymbol{w}) = \boldsymbol{K}^{-1} \boldsymbol{C} \boldsymbol{Q}_{\mathcal{E},n}^{-1} \boldsymbol{C} \boldsymbol{K}^{-\mathsf{T}}, \quad \text{i.e.} \\ \boldsymbol{Q} = \boldsymbol{K}^\mathsf{T} \boldsymbol{C}^{-1} \boldsymbol{Q}_{\mathcal{E},n} \boldsymbol{C}^{-1} \boldsymbol{K}. \end{aligned}$$

□

Proof: [Theorem 3] First, we show that part (I) follows from part (II).

Let f and g be functions in $\mathcal{H}^1(\Omega, \kappa)$, and let f' be a solution to the PDE

$$\begin{cases} (\kappa^2 - \Delta) f'(\boldsymbol{u}) = f(\boldsymbol{u}), & \boldsymbol{u} \in \Omega, \\ \partial_{\boldsymbol{n}} f'(\boldsymbol{u}) = 0, & \boldsymbol{u} \in \partial\Omega, \end{cases}$$

and equivalently for g. Then f' and g' are also in $\mathcal{H}^1(\Omega, \kappa)$, and fulfil the requirements of Lemma 1, so that

$$\begin{aligned} \langle f, x_n \rangle_\Omega &= \left\langle (\kappa^2 - \Delta) f', x_n \right\rangle_\Omega \\ &= \left\langle (\kappa^2 - \Delta)^{1/2} f', (\kappa^2 - \Delta)^{1/2} x_n \right\rangle_\Omega \\ &= \left\langle f', (\kappa^2 - \Delta) x_n \right\rangle_\Omega \end{aligned}$$

and

$$\langle f, x \rangle_\Omega = \left\langle f', (\kappa^2 - \Delta) x \right\rangle_\Omega.$$

The convergence of x_n to x follows from part (II). In the Galerkin case, we have

$$\begin{aligned} \mathsf{Cov}(\langle f, x_n \rangle_\Omega, \langle g, x_n \rangle_\Omega) &= \mathsf{Cov}(\left\langle f', \mathcal{L}x_n \right\rangle_\Omega, \left\langle g', \mathcal{L}x_n \right\rangle_\Omega) \\ &\to \mathsf{Cov}(\left\langle f', \mathcal{L}x \right\rangle_\Omega, \left\langle g', \mathcal{L}x \right\rangle_\Omega) \\ &= \mathsf{Cov}(\langle f, x \rangle_\Omega, \langle g, x \rangle_\Omega), \end{aligned}$$

and similarly for the Least Squares case.

Part (II):

Let $f_n = \sum_k \psi_k w_{f,k}$ and $g_n = \sum_k \psi_k w_{g,k}$ be the projections of f and g onto $\mathcal{H}^1_n(\Omega, \kappa)$. In case (a), then

$$\begin{aligned}\langle f, \mathcal{L}x_n\rangle_\Omega &= \kappa^2 \langle f, x_n\rangle_\Omega + \langle \nabla f, \nabla x_n\rangle_\Omega - \langle f, \partial_{\boldsymbol{n}} x_n\rangle_{\partial\Omega} \\ &= \langle f, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)} - 0 \\ &= \langle f - f_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)} + \langle f_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)} \\ &= \langle f_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)}\,,\end{aligned}$$

and

$$\begin{aligned}\mathsf{Cov}(\langle f, \mathcal{L}x_n\rangle_\Omega, \langle g, \mathcal{L}x_n\rangle_\Omega) &= \mathsf{Cov}(\langle f_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)}, \langle g_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)}) \\ &= \mathsf{Cov}(\langle f_n, \mathcal{W}\rangle_\Omega, \langle g_n, \mathcal{W}\rangle_\Omega) \\ &= \langle f_n, g_n\rangle_\Omega \\ &\to \langle f, g\rangle_\Omega \\ &= \mathsf{Cov}(\langle f, \mathcal{W}\rangle_\Omega, \langle g, \mathcal{W}\rangle_\Omega)\end{aligned}$$

as $n \to \infty$. Similarly in case (b), for any $f \in \mathcal{H}^1(\Omega, \kappa)$ fulfilling the requirements of Lemma 1,

$$\begin{aligned}\langle \mathcal{L}f, \mathcal{L}x_n\rangle_\Omega &= \langle f, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)} \\ &= \langle f_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)}\,,\end{aligned}$$

and

$$\begin{aligned}\mathsf{Cov}(\langle \mathcal{L}f, \mathcal{L}x_n\rangle_\Omega, \langle \mathcal{L}g, \mathcal{L}x_n\rangle_\Omega) &= \mathsf{Cov}(\langle f_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)}, \langle g_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)}) \\ &= \mathsf{Cov}(\langle \mathcal{L}f_n, \mathcal{W}\rangle_\Omega, \langle \mathcal{L}g_n, \mathcal{W}\rangle_\Omega) \\ &= \langle f_n, g_n\rangle_{\mathcal{H}^1(\Omega,\kappa)} \\ &\to \langle f, g\rangle_{\mathcal{H}^1(\Omega,\kappa)} \\ &= \langle \mathcal{L}f, \mathcal{L}g\rangle_\Omega \\ &= \mathsf{Cov}(\langle \mathcal{L}f, \mathcal{W}\rangle_\Omega, \langle \mathcal{L}g, \mathcal{W}\rangle_\Omega)\end{aligned}$$

as $n \to \infty$. □

Proof: [Theorem 4] First, we show that part (II) follows from part (III).

Let f' and g' be defined as in the proof of Theorem 3. Then, since $\mathcal{L}_x = \kappa^2 - \Delta$,

$$\begin{aligned}\langle f, x_n\rangle_\Omega &= \left\langle f', \mathcal{L}_x x_n\right\rangle_\Omega \quad \text{and} \\ \langle f, x\rangle_\Omega &= \left\langle f', \mathcal{L}_x x\right\rangle_\Omega\,,\end{aligned}$$

and the convergence of x_n to x follows from part (III).

As in the proof of Theorem 3,

$$\langle f, \mathcal{L}_x x_n\rangle_\Omega = \langle f_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)}\,,$$

and

$$\begin{aligned}\mathsf{Cov}(\langle f, \mathcal{L}_x x_n\rangle_\Omega, \langle g, \mathcal{L}_x x_n\rangle_\Omega) &= \mathsf{Cov}(\langle f_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)}, \langle g_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)}) \\ &= \mathsf{Cov}(\langle f_n, y_n\rangle_\Omega, \langle g_n, y_n\rangle_\Omega) \\ &= \mathsf{Cov}(\langle f, y_n\rangle_\Omega, \langle g, y_n\rangle_\Omega) \\ &\to \mathsf{Cov}(\langle f, y\rangle_\Omega, \langle g, y\rangle_\Omega) \\ &= \mathsf{Cov}(\langle f, \mathcal{L}_x x\rangle_\Omega, \langle g, \mathcal{L}_x x\rangle_\Omega)\end{aligned}$$

as $n \to \infty$, due to requirement (I). □

E Spherical harmonics

In $\mathbb{R}^2$, the harmonic functions, sine and cosine, play an important role as basis functions in spectral representations of functions and random fields. On the sphere, this role is instead taken by the *spherical harmonics*. This section introduces these functions and their most important properties. The functions are used for the global climate parameters in the example in Section 4.

Definition 11 *The* spherical harmonic $Y_{k,m}(\boldsymbol{u})$, $\boldsymbol{u} = \begin{bmatrix} u_1, u_2, u_3 \end{bmatrix}^\mathsf{T} \in \mathbb{S}^2 \subset \mathbb{R}^3$, *of* order $k = 0, 1, 2, \ldots$ *and* mode $m = -k, \ldots, k$ *is defined by*

$$Y_{k,m}(\boldsymbol{u}) = \sqrt{(2k+1) \cdot \frac{(k-|m|)!}{(k+|m|)!}} \cdot \begin{cases} \sqrt{2}\sin(m\phi)P_{k,-m}(\cos\theta) & -k \leq m < 0, \\ P_{k,0}(\cos\theta) & m = 0, \\ \sqrt{2}\cos(m\phi)P_{k,m}(\cos\theta) & 0 < m \leq k, \end{cases}$$

where ϕ *is the longitude and* $\theta = \arccos(u_3)$ *is the colatitude, and* $P_{k,|m|}(u_3)$ *are* associated Legendre functions *(*$P_{k,0}(u_3)$ *are Legendre polynomials). Note that* $\sin\phi = u_2/\sqrt{u_1^2 + u_2^2}$, $\cos\phi = u_1/\sqrt{u_1^2 + u_2^2}$, *and* $\cos\theta = u_3$.

Theorem 5 *(Mostly from Wahba (1981)) Some properties of the spherical harmonics:*

1. *The spherical harmonics form an orthogonal basis for functions on the unit sphere,* $\mathbb{S}^2$*:*

$$\left\langle Y_{k,m}, Y_{k',m'} \right\rangle_{\mathbb{S}^2} = \begin{cases} 4\pi, & k' = k, m' = m, \\ 0, & \textit{otherwise.} \end{cases}$$

2. *The addition formula for spherical harmonics is*

$$\sum_{m=-k}^{k} Y_{k,m}(\boldsymbol{u})Y_{k,m}(\boldsymbol{v}) = (2k+1)P_{k,0}(\boldsymbol{u}^\mathsf{T}\boldsymbol{v}).$$

3. *The spherical harmonics are eigenfunctions to the Laplacian on* $\mathbb{S}^2$*,*

$$\Delta Y_{k,m}(\boldsymbol{u}) = -k(k+1)Y_{k,m}(\boldsymbol{u}).$$

4. *Let* $\phi(\boldsymbol{u})$ *be a square-integrable function on* $\mathbb{S}^2$*. Then* $\phi(\boldsymbol{u})$ *has series expansion*

$$\phi(\boldsymbol{u}) = (\mathcal{F}^{-1}\widehat{\phi})(\boldsymbol{u}) = \sum_{k=0}^{\infty}\sum_{m=-k}^{k} \widehat{\phi}(k,m)Y_{k,m}(\boldsymbol{u}),$$

with Fourier Bessel *coefficients*

$$\widehat{\phi}(k,m) = (\mathcal{F}\phi)(k,m) = \frac{1}{4\pi}\left\langle \phi(\boldsymbol{u}), Y_{k,m}(\boldsymbol{u}) \right\rangle_{\mathbb{S}^2(\mathrm{d}\boldsymbol{u})}.$$

Also,

$$\langle \phi, 1 \rangle_{\mathbb{S}^2} = 4\pi\widehat{\phi}(0,0) \quad \textit{and} \quad \langle \phi, \phi \rangle_{\mathbb{S}^2} = 4\pi\sum_{k,m}\widehat{\phi}(k,m)^2.$$

References

Allcroft, D. J. and Glasbey, C. A. (2003). A latent Gaussian Markov random field model for spatio-temporal rainfall disaggregation. *Journal of the Royal Statistical Society, Series C*, 52:487–498.

Arjas, E. and Gasbarra, D. (1996). Bayesian inference of survival probabilities, under stochastic ordering constraints. *Journal of the American Statistical Association*, 91(435):1101–1109.

Auslander, L. and MacKenzie, R. E. (1977). *Introduction to Differentiable Manifolds*. Dover Publications, Inc., New York.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, volume 101 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

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

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36(2):192–225.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24(3):179–195.

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

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746.

Besag, J. and Mondal, D. (2005). First-order intrinsic autoregressions and the de Wijs process. *Biometrika*, 92(4):909–920.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43(1):1–59.

Bolin, D. and Lindgren, F. (2009a). Spatial stochastic models generated by nested stochastic partial differential equations. *Preprints in Mathematical Sciences, Lund University*, (2009:14):submitted.

Bolin, D. and Lindgren, F. (2009b). Wavelet Markov models as efficient alternatives to tapering and convolution fields. *Preprints in Mathematical Sciences, Lund University*, (2009:13):submitted.

Brenner, S. C. and Scott, R. (2007). *The Mathematical Theory of Finite Element Methods*. Springer, 3rd edition.

Brohan, P., Kennedy, J., Harris, I., Tett, S., and Jones, P. (2006). Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *Journal of Geophysical Research*, 111.

Chilés, J. P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Cressie, N. and Huang, H. C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1340.

Cressie, N. and Verzelen, N. (2008). Conditional-mean least-squares fitting of Gaussian Markov random fields to Gaussian fields. *Computational Statistics & Data Analysis*, 52(5):2794–2807.

Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

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

Dahlhaus, R. and Künsch, H. R. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, 74(4):877–882.

Davis, T. A. (2006). *Direct methods for sparse linear systems*. SIAM Book Series on the Fundamentals of Algortihms. SIAM, Philadelphia.

Diggle, P. J. and Ribeiro, P. J. (2006). *Model-based Geostatistics*. Springer Series in Statistics. Springer.

Duff, I. S., Erisman, A. M., and Reid, J. K. (1989). *Direct Methods for Sparse Matrices*. Monographs on Numerical Analysis. The Clarendon Press Oxford University Press, New York, 2nd edition. Oxford Science Publications.

Edelsbrunner, H. (2001). *Geometry and Topology for Mesh Generation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.

Federer, H. (1951). Hausdorff measure and Lebesgue area. *Proceedingsof the National Academy of Sciences of the United States of America*, 37(2):90–94.

Federer, H. (1978). Colloquium lectures on geometric measure theory. *Bulletin of the American Mathematical Society*, 84(3):291–338.

Fuentes, M. (2001). High frequency kriging for nonstationary environmental processes. *Environmetrics*, 12(5):469–483.

Fuentes, M. (2008). Approximate likelihood for large irregular spaced spatial data. *Journal of the American Statistical Association*, 102(477):321–331.

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

George, A. and Liu, J. W. H. (1981). *Computer solution of large sparse positive definite systems*. Prentice-Hall Inc., Englewood Cliffs, N.J. Prentice-Hall Series in Computational Mathematics.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97:590–600.

Gschlößl, S. and Czado, C. (2007). Modelling count data with overdispersion and spatial effects. *Statistical papers*. http://dx.doi.org/10.1007/s00362-006-0031-6.

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

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimentional lattice. *Biometrika*, 69(1):95–105.

Hansen, J., Ruedy, R., Glascoe, J., and Sato, M. (1999). Giss analysis of surface temperature change. *Journal of Geophysical Research*, 104:30997–31022.

Hansen, J., Ruedy, R., Sato, M., Imhoff, M., Lawrence, W., Easterling, D., Peterson, T., and Karl, T. (2001). A closer look at united states and global surface temperature change. *Journal of Geophysical Research*, 106:23947–23963.

Hartman, L. and Hössjer, O. (2008). Fast kriging of large data sets with gaussian markov random fields. *Computational Statistics & Data Analysis*, 52(5):2331–2349.

Henderson, R., Shimakura, S., and Gorst, D. (2002). Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association*, 97(460):965–972.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, 5(2):173–190.

Higdon, D., Swall, J., and Kern, J. (1999). Non-stationary spatial modeling. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics, 6*, pages 761–768. Oxford University Press, New York.

Hjelle, Ø. and Dæhlen, M. (2006). *Triangulations and Applications*. Springer.

Hrafnkelsson, B. and Cressie, N. A. C. (2003). Hierarchical modeling of count data with application to nuclear fall-out. *Environmental and Ecological Statistics*, 10:179–200.

Hughes-Oliver, J. M., Gonzalez-Farias, G., Lu, J. C., and Chen, D. (1998). Parametric nonstationary correlation models. *Statistics and Probability Letters*, 40(15):267–278.

Jun, M. and Stein, M. L. (2008). Nonstationary covariancs models for global data. *The Annals of Applied Statistics*, 2(4):1271–1289.

Kleoden, P. E. and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*. Springer, 3rd edition.

Kneib, T. and Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, 34(1):207–228.

Kotelenez, P. (1999). *Numerical Solution of Stochastic Differential Equations*. Springer, 3rd edition.

Krantz, S. G. and Parks, H. R. (2008). *Geometric Integration Theory*. Birkhäuser.

Lindgren, F. and Rue, H. (2008). A note on the second order random walk model for irregular locations. *Scandinavian Journal of Statistics*, 35(4):691–700.

Paciorek, C. and Schervish, M. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17:483–506.

Peterson, T. and Vose, R. (1997). An overview of the Global Historical Climatology Network temperature database. *Bulletin of the American Meteorological Society*, 78(12):2837–2849.

Pettitt, A. N., Weir, I. S., and Hart, A. G. (2002). A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing*, 12(4):353–367.

Quarteroni, A. M. and Valli, A. (2008). *Numerical Approximation of Partial Differential Equations*. Springer, 2nd edition.

Rozanov, A. (1982). *Markov Random Fields*. Springer Verlag, New York.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B*, 63(2):325–338.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.

Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, 29(1):31–50.

Samko, S. G., Kilbas, A. A., and Maricev, O. I. (1992). *Fractional integrals and derivatives: theory and applications*. Gordon and Breach Science Publishers, Yverdon.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119.

Smith, T. (1934). Change of variables in Laplace's and other second-order differential equations. *Proceedings of the Physical Society*, 46(3):344–349.

Song, H., Fuentes, M., and Gosh, S. (2008). A compariative study of Gaussian geostatistical models and Gaussian Markov random field models. *Journal of Multivariate Analysis*, 99:1681–1697.

Stein, M. (2005). Space-time covariance functions. *Journal of the American Statistical Association*, 100:310–321.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, New York.

Stein, M. L., Chi, Z., and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society, Series B*, 66(2):275–296.

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

Wahba, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM Journal of Scientific and Statistical Computing*, 2(1):5–16.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2):311–324.

Weir, I. S. and Pettitt, A. N. (2000). Binary probability maps using a hidden conditional autoregressive Gaussian process with an application to Finnish common toad data. *Journal of the Royal Statistical Society, Series C*, 49(4):473–484.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(3/4):434–449.

Whittle, P. (1963). Stochastic processes in several dimensions. *Bull. Inst. Internat. Statist.*, 40:974–994.