An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach

An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach

Finn Lindgren and Håvard Rue†

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

and Johan Lindström

Mathematical Statistics, Centre for Mathematical Sciences
Lund University, Lund, Sweden

Summary. Continuously indexed Gaussian fields (GFs) is the most important ingredient in spatial statistical modelling and geostatistics. The specification through the covariance function gives an intuitive interpretation of the field 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 at an all-time-high, this fact seems still to be a computational bottleneck in many 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, formulated as a basis function representation. 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.

Keywords: Approximate Bayesian inference, Covariance functions, Gaussian Markov random fields, Gaussian fields, Latent Gaussian models, Sparse matrices, Stochastic partial differential equations.

1. Introduction

Gaussian fields (GFs) have a dominant role in spatial statistics and especially in the traditional field of geostatistics (Cressie, 1993; Stein, 1999; Chilés and Delfiner, 1999; Diggle and Ribeiro, 2006), and form an important building block in modern hierarchical spatial models (Banerjee et al., 2004). GFs

†*Address for Correspondence:* Håvard Rue, Department of Mathematical Sciences, The Norwegian University of Science and Technology, N-7491 Trondheim, Norway.
Email: hrue@math.ntnu.no
WWW-address: http://www.math.ntnu.no/∼hrue

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 $\boldsymbol{s} \in \mathcal{D}$, $x(\boldsymbol{s})$ is 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 relative position of two locations, in which case it is said to be stationary, and it is isotropic if the covariance functions only depends on the Euclidean distance between the locations. Since a regular covariance matrix is positive definite, the covariance function must be a positive definite function. This restriction makes it difficult to "invent" covariance functions stated as closed form expressions. Bochner's theorem can be used in this context, as it characterises all continuous positive definite functions in $\mathbb{R}^d$.

Although GFs are convenient from both an analytical and a practical point of view, the computational issues have 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 situation is informally referred to as "the big n problem".

There are several approaches trying to overcome or avoid "the big n problem". The spectral representation approach for the likelihood (Whittle, 1954) makes it possible to estimate the (power-)spectrum (using discrete Fourier transforms calculations) and compute the log-likelihood from it (Guyon, 1982; Dahlhaus and Künsch, 1987; Fuentes, 2008) but this is only possible for directly observed stationary GFs on a (near-)regular lattice. Vecchia (1988) and Stein et al. (2004) propose to use an approximate likelihood constructed through a sequential representation and then simplify the conditioning set, and similar ideas also apply when computing conditional expectations (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; Eidsvik et al., 2010). 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 $\pi(x_i \mid \boldsymbol{x}_{-i})$, $i = 1, \ldots, n$, 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; $Q_{ij} \neq 0 \iff i \in \partial j \cup j$, 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 (George and Liu, 1981; Duff et al., 1989; Davis, 2006) at a typical cost of $\mathcal{O}(n^{3/2})$ for two-dimensional GMRFs; see (Rue and Held, 2005) for detailed algorithms. GMRFs have very good computational properties, which is of major importance in Bayesian inferential methods. This is further enhanced by the link to nested integrated Laplace approximations (INLA) (Rue et al., 2009), which allows for fast and accurate Bayesian inference for latent Gaussian field models.

Although GMRFs have very good computational properties, there are reasons for why current statistical models based on GMRFs are relatively simple, in particular when applied to area data from regions or counties. First, there 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, the reason for this is that one must construct a positive definite precision matrix in order to obtain a positive definite covariance matrix as its inverse, so the conditions for proper covariance matrices are replaced by essentially equivalent conditions for sparse precision matrices. Therefore, often simplistic approaches are taken, like letting Q_{ij} be related to the reciprocal distance between sites i and j (Besag et al., 1991; Arjas and Gasbarra, 1996; Weir and Pettitt, 2000; Pettitt et al., 2002; Gschlößl and Czado, 2007), 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. The 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 closely approximate most of the commonly used covariance functions in geostatistics, and proposed to use them as computational replacements for GFs for computational reasons like doing Kriging (Hartman and Hössjer, 2008). 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), using a time-consuming numerical optimisation. Despite these 'proof-of-concept' results, several authors have followed up this idea without any notable progress in the methodology (Hrafnkelsson and Cressie, 2003; Song et al., 2008; Cressie and Verzelen, 2008), but the approach itself has shown useful even for spatio-temporal models (Allcroft and Glasbey, 2003).

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

(a) Do the modelling using a GF on a set of locations $\{\boldsymbol{s}_i\}$, to construct a discretised GF with covariance matrix $\boldsymbol{\Sigma}$
(b) 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 word "represents" instead of approximates.)
(c) 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 accurately in order to maintain 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 Stein (1999, p.14) which concluded a detailed theoretical analysis with "*Use the Matérn model*". The GMRF representation can be constructed explicitly using a certain stochastic partial differential equation (SPDE) which has GFs with Matérn covariance function as the solution when driven by Gaussian white noise. The result is a basis function representation with piecewise linear basis functions, and Gaussian weights with Markov dependencies determined by a general triangulation of the domain.

Rather surprisingly, extending this basic result seems to open new doors and opportunities, and provide rather simple answers to rather difficult modelling problems. In particular, we will show how this approach extends to Matérn fields on manifolds, non-stationary fields and fields with oscillating covariance functions. Further, we will discuss the link to the deformation method by Sampson and Guttorp (1992) for non-stationary covariances for non-isotropic models, and how our approach naturally extends to non-separable space-time models. 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 avail-

able explicitly. An important observation is that the resulting modelling strategy does not involve having to construct explicit formulae for the covariance functions, which are instead only defined implicitly through the SPDE specifications.

The plan of rest of this paper is as follows. In Sec. 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 precision matrices for GMRFs based on this relationship. In Sec. 3, the results are extended to fields on triangulated manifolds, non-stationary and oscillating models, and non-separable space-time models. The extensions are illustrated with a non-stationary analysis of global temperature data in Sec. 4, and we conclude the main part of the paper with a discussion in Sec. 5. Thereafter follows four technical appendices, with explicit representation results (A), theory for random fields on manifolds (B), the Hilbert space representation details (C), and proofs of the technical details (D).

2. Preliminaries and main results

This section will introduce the Matérn covariance model and discuss its representation through a 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

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

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. Lacking a simple relationship, we will throughout this paper use the empirically derived 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 a Gaussian field $x(\boldsymbol{u})$ with the Matérn covariance is a solution to the linear fractional stochastic partial differential equation (SPDE)

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

where $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo-differential operator that we will define later in (4) through its spectral properties (Whittle, 1954, 1963). The innovation process $\mathcal{W}$ is spatial Gaussian white noise with unit variance, Δ is the *Laplacian*

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

and the marginal variance is

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

We will name any solution to (2) *a Matérn field* in the following. However, the limiting solutions to the SPDE (2) as $\kappa \to 0$ or $\nu \to 0$ 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 App. C.3.

Further, there is an implicit assumption of appropriate boundary conditions for the SPDE, as for $\alpha \geq 2$ the null-space of the differential operator is non-trivial, containing, for example, the functions $\exp(\kappa \boldsymbol{e}^{\mathsf{T}}\boldsymbol{u})$, for all $\|\boldsymbol{e}\| = 1$. The Matérn fields are the only *stationary* solutions to the SPDE.

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

$$R(\boldsymbol{k}) = (2\pi)^{-d}\left(\kappa^2 + \|\boldsymbol{k}\|^2\right)^{-\alpha}, \tag{3}$$

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

$$\left\{\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2}\phi\right\}(\boldsymbol{k}) = (\kappa^2 + \|\boldsymbol{k}\|^2)^{\alpha/2}(\mathcal{F}\phi)(\boldsymbol{k}), \tag{4}$$

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 ourselves 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 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

$$\mathsf{E}(x_{ij} \mid \boldsymbol{x}_{-ij}) = \frac{1}{a}\left(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}\right), \qquad \mathsf{Var}(x_{ij} \mid \boldsymbol{x}_{-ij}) = 1/a \tag{5}$$

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

$$\begin{array}{|ll} \hline -1 & \\ a & -1 \\ \hline \end{array} \tag{6}$$

which displays the elements of the precision matrix related to a single location (Chapter 3.4.2 in Rue and Held, 2005, uses a related graphical notation). Due to symmetry, we only display the upper right quadrant, with "a" as the central element. The approximate result (Besag, 1981, Eq. (14)) is that

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

where l is the Euclidean distance between ij and $i'j'$. Evaluated for continuous distances, this is a generalised covariance function, obtained from (1) in the limit $\nu \to 0$, with $\kappa^2 = a - 4$ and $\sigma^2 = a/(4\pi)$, even though (1) requires $\nu > 0$. Informally, this means that the discrete model defined by (5) generates approximate solutions to the SPDE in (2) on a unit-distance regular grid, with $\nu = 0$.

Solving (2) for $\alpha = 1$ gives a generalised random field with spectrum

$$R_1(\boldsymbol{k}) \propto \left\{(a-4) + \|\boldsymbol{k}\|^2\right\}^{-1},$$

meaning that (some discretised version of) the SPDE acts like a linear filter with squared transfer-function equal to R_1. If we replace the noise term on the right hand side of (2) by Gaussian noise with spectrum

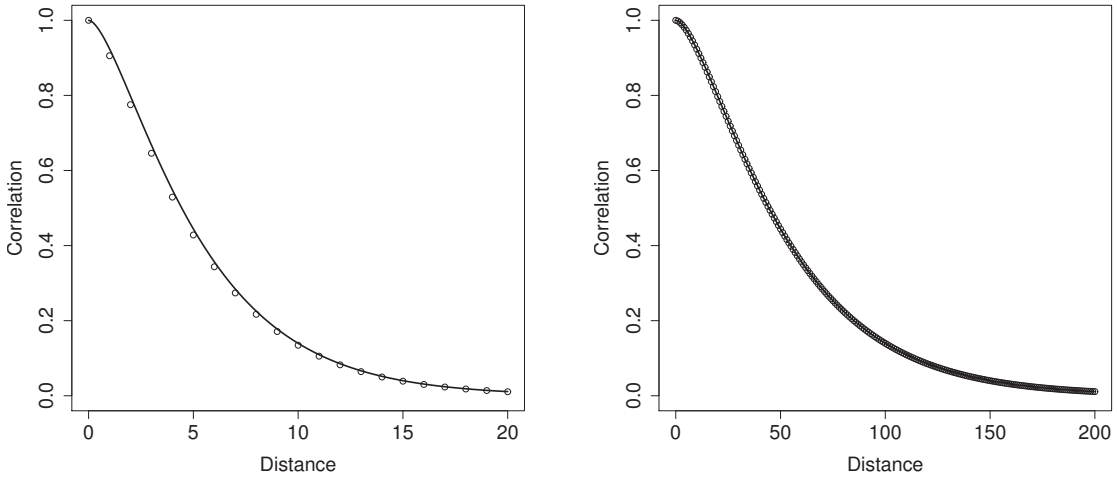


Fig. 1. The Matérn correlations (solid line) for range 10 (left) and 100 (right), and the correlations for the GMRF representation (circles).

R_1, the resulting solution has 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 (6),

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

The marginal variance is $1/(4\pi\nu(a-4)^\nu)$. Fig. 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 results 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 in the marginal variance is 4% for range 10 and negligible for range 100.

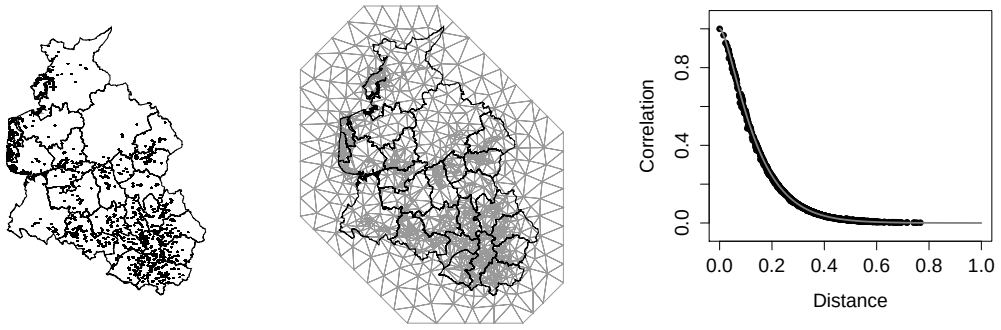
Our first result confirms the above heuristics.

MAIN RESULT 1. *The coefficients in the GMRF representation of (2) on a regular unit-distance two-dimensional infinite lattice for $\nu = 1, 2, \ldots$, is found by convolving (6) by itself ν times.*

Simple extensions of this result includes anisotropy along the main axes, as presented in App. A. A rigorous formulation of the result is derived in the subsequent Appendices, showing that the basic result is a special case of a more general link between SPDEs and GMRFs. The first such generalisation, based on irregular grids, is the next main result.

2.3. Main result 2

Although Main Result 1 is useful in itself, it 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, such as maximally allowed edge length, and minimally allowed angles. This



(a) (b) (c)

Fig. 2. Panel (a) displays the locations of leukaemia survival observations, panel (b) the triangulation using 3446 triangles, and panel (c) displays a stationary correlation function (line) and the corresponding GMRF approximation (dots) for $\nu = 1$ and approximate range 0.26.

is a standard problem in engineering for solving partial differential equations using finite element methods (Ciarlet, 1978; Brenner and Scott, 2007; Quarteroni and Valli, 2008), where the quality of the solutions depends on the triangulation properties. Typically, the triangulation is chosen to maximise the minimum interior triangle angle, so called Delaunay triangulations, which helps ensure that the transitions between small and large triangles are smooth. The extra vertices are added heuristically to try to minimise the total number of triangles needed to fulfil the size and shape constraints. See for example Edelsbrunner (2001); Hjelle and Dæhlen (2006) for algorithm details. Our implementation in the `R-inla` package (`www.r-inla.org`) is based on Hjelle and Dæhlen (2006).

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. Fig. 2(a) displays the locations of 1043 cases of acute myeloid leukaemia in adults who have been diagnosed between 1982 and 1998 in Northwest England. In the analysis, the spatial scale has been normalised so that the height of the study region is equal to one. 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 1749 and the number of triangles is 3446.

In order to construct a GMRF representation of the Matérn field on the triangulated lattice, we start with a stochastic weak formulation of the SPDE (2). Define the inner product

$$\langle f, g \rangle = \int f(\boldsymbol{u}) g(\boldsymbol{u}) d\boldsymbol{u} \tag{7}$$

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

$$\left\{ \langle \phi_j, (\kappa^2 - \Delta)^{\alpha/2} x \rangle,\, j = 1, \ldots, m \right\} \stackrel{d}{=} \{ \langle \phi_j, \mathcal{W} \rangle,\, j = 1, \ldots, m \} \tag{8}$$

for every appropriate finite set of *test functions* $\{\phi_j(\boldsymbol{u}), j = 1, \ldots, m\}$, where "$\stackrel{d}{=}$" denotes equality in distribution.

The next step is to construct a finite element *representation* of the solution to the SPDE (Brenner and

Scott, 2007) as

$$x(\boldsymbol{u}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{u}) w_k \tag{9}$$

for some chosen basis functions $\{\psi_k\}$ and Gaussian distributed weights $\{w_k\}$. Here, n is the number of vertices in the triangulation. We choose 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. An interpretation of the representation (9) with this choice of basis functions is that the weights determine the values of the field at the vertices, and the values in the interior of the triangles are determined by linear interpolation. The full distribution of the continuously indexed solution is determined by the joint distribution of the weights.

The finite dimensional solution is obtained by finding the distribution for the representation weights in (9) that fulfils the stochastic weak SPDE formulation (8) for only a *specific* set of test functions, with $m = n$. The choice of test functions, in relation to the basis functions, governs the approximation properties of the resulting model representation. We choose $\phi_k = (\kappa^2 - \Delta)^{1/2}\psi_k$ for $\alpha = 1$ and $\phi_k = \psi_k$ for $\alpha = 2$. These two approximations are denoted the *least squares* and the *Galerkin* solution, respectively. For $\alpha \geq 3$, we let $\alpha = 2$ in the left-hand side of (2), and replace the right-hand side with a field generated by $\alpha - 2$, and let $\phi_k = \psi_k$. In essence, this generates a recursive Galerkin formulation, terminating in either $\alpha = 1$ or $\alpha = 2$; see App. C for details.

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

$$C_{ij} = \langle \psi_i, \psi_j \rangle, \qquad G_{ij} = \langle \nabla\psi_i, \nabla\psi_j \rangle, \qquad (\boldsymbol{K}_{\kappa^2})_{ij} = \kappa^2 C_{ij} + G_{ij}.$$

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 (9) for $\alpha = 1, 2, \ldots$, as a function of κ^2. Then the finite dimensional representations of the solutions to* (2) *have precisions*

$$\begin{cases} \boldsymbol{Q}_{1,\kappa^2} = \boldsymbol{K}_{\kappa^2}, \\ \boldsymbol{Q}_{2,\kappa^2} = \boldsymbol{K}_{\kappa^2}\boldsymbol{C}^{-1}\boldsymbol{K}_{\kappa^2}, \\ \boldsymbol{Q}_{\alpha,\kappa^2} = \boldsymbol{K}_{\kappa^2}\boldsymbol{C}^{-1}\boldsymbol{Q}_{\alpha-2,\kappa^2}\boldsymbol{C}^{-1}\boldsymbol{K}_{\kappa^2}, \quad \textit{for } \alpha = 3, 4, \ldots. \end{cases} \tag{10}$$

Some remarks concerning this result:

(a) 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. Explicit formulae are given in App. A.

(b) The matrix $\boldsymbol{C}^{-1}$ is dense, which makes the precision matrix dense as well. In App. C.5, we show that $\boldsymbol{C}$ can be replaced by the diagonal matrix $\widetilde{\boldsymbol{C}}$, where $\widetilde{C}_{ii} = \langle \psi_i, 1 \rangle$, which makes the precision matrices sparse, and hence we obtain GMRF models.

(c) A consequence of the previous remarks is that we have an explicit mapping from the parameters of the GF model to the elements of a GMRF precision matrix, with computational cost of $\mathcal{O}(n)$ for any triangulation.

(d) For the special case where all the vertices are points on a regular lattice, using a regular triangularisation reduces Main Result 2 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.

(e) 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 $\mathbb{R}^2$.

(f) 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 (4). 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. For our SPDE (2) this corresponds to $\alpha = 1, 2, 3, \ldots$; see (3). This result indicates that a different approach may be needed to provide representation results when α is not an integer, such as approximating the spectrum itself. Given approximations for general $0 \leq \alpha \leq 2$, the recursive approach could then be used for general $\alpha > 2$.

Although the 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 possible test functions. However, for a given triangulation, it is the best possible approximation in the sense made explicit in App. C, where we also show weak convergence to the full SPDE solutions. Using standard results from the finite element literature (Brenner and Scott, 2007), it is also possible to derive rates of convergence results, like for $\alpha = 2$,

$$\sup_{f \in \mathcal{H}^1; \|f\|_{\mathcal{H}^1} \leq 1} \mathsf{E}\left(\langle f, x_n - x\rangle_{\mathcal{H}^1}^2\right) \leq ch^2. \tag{11}$$

Here, x_n is the GMRF representation of the SPDE solution x, h is the diameter of the largest circle that can be inscribed in a triangle in the triangulation, c is some constant. The Hilbert space scalar product and norm include the values and the gradients of the field, and are defined in Definition 2 in App. B. The result holds for general $d \geq 1$, with h proportional to the edge lengths between the vertices, when the minimal mesh angles are bounded away from zero.

To see how well we are able to approximate the Matérn covariance, Fig. 2(c) displays the empirical correlation function (dots) and the theoretical one for $\nu = 1$ with approximate range 0.26, using the triangulation in Fig. 2(b). The match is quite good. Some dots shows 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 Fig. 2(b) we chose to use a fine resolution in the study-area and a reduced resolution outside. A minor drawback using these GMRFs in place of given stationary covariance models, is the boundary effects due to the boundary conditions of the SPDE. In Main Result 2 we used Neumann conditions that inflates the variance near the boundary (see App. A.4 for details) but other choices are also possible (see Rue and Held, 2005, Chapter 5).

2.4. Leukaemia example

We will now return to the Henderson et al. (2002) example from the beginning of Sec. 2.3 which models spatial variation in leukaemia survival data in Northwest England. The specification, in (pseudo) Wilkinson-Rogers notation (McCullagh and Nelder, 1989, Sec. 3.4) is

$$\text{survival(time, censoring)} \;\sim\; \text{intercept} + \text{sex} + \text{age} + \text{wbc} + \text{tpi} + \text{spatial(location)}$$

using a Weibull-likelihood for the survival times, and where "wbc" is the white blood-cell count at diagnosis, "tpi" is the Townsend deprivation index (which is a measure of economic 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 are the marginal variance and range for the spatial component, and the shape-parameter in the Weibull distribution.

Kneib and Fahrmeir (2007) reanalysed the same data-set using a Cox proportional hazards model but, for computational reasons, used a low-rank approximation for the spatial component. With our GMRF representation we easily work with a sparse 1749×1749 precision matrix for the spatial component. We

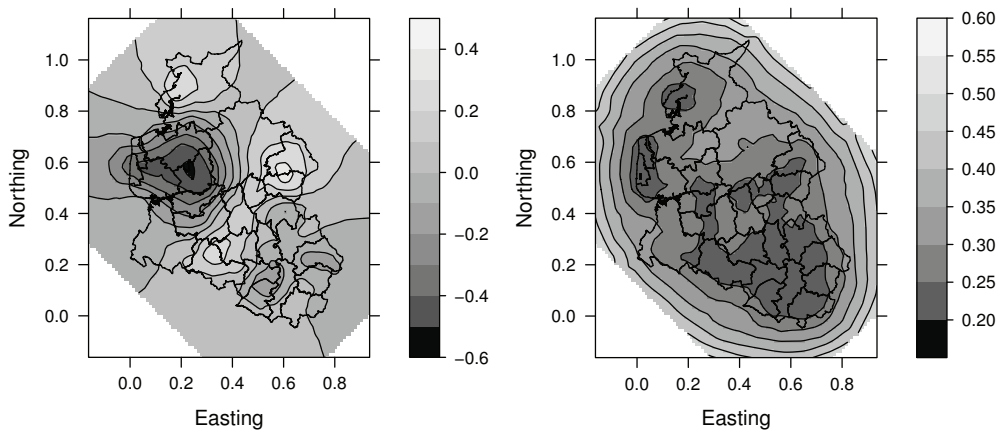


Fig. 3. The posterior mean (left) and standard deviation (right) of the spatial effect on survival using the GMRF representation.

ran the model in `R-inla` (`www.r-inla.org`) using integrated nested Laplace approximations to do the full Bayesian analysis (Rue et al., 2009). Fig. 3 displays the posterior mean and standard deviation of the spatial component. A full Bayesian analysis took about 16 seconds on a quad-core laptop, and factorising the 2797×2797 (total) precision matrix took about 0.016 seconds on average.

3. Extensions: Beyond classical Matérn models

In this section we will discuss five extensions to the SPDE, widening the usefulness of the GMRF construction results in various ways. The first extension is to consider solutions to the SPDE on a manifold, which allows us to *define* Matérn fields on domains such as a 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 (2) which makes it possible to construct oscillating fields. The fourth extension generalises the non-stationary SPDE to a more general class of non-isotropic fields. Finally, the fifth extension shows how the SPDE generalises to non-separable space-time models.

An important feature in our approach, is that all these extensions still give explicit GMRF representations similar to (9) and (10), 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 beyond the geometric properties of the triangulation. In Sec. 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 manifolds. Gaussian fields on manifolds is a well-studied subject with important application to excursion sets in brain mapping (Adler and Taylor, 2007; Bansal et al., 2007; Adler, 2009). Our main objective is to construct Matérn fields on the sphere, which is important for the analysis of global spatial and spatio-temporal models. To simplify the current discussion we will therefore restrict the construction of Matérn fields to a unit radius sphere $\mathbb{S}^2$ in three dimensions, leaving the general case for the appendices.

Just as for $\mathbb{R}^d$, models on a sphere can be constructed via a spectral approach (Jones, 1963). A more direct 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. Using the great circle distances in the original covariance function would not work in general, since for differentiable fields this does not yield a valid positive definite covariance function (this follows from Gneiting, 1998, Theorem 2). 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 of $\mathbb{R}^d$, 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 field $W(\cdot)$ 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 (7) to a surface integral on $\mathbb{S}^2$. The theory in App. B through App. D covers the general manifold setting.

To illustrate the continuous index definition and the Markov representation of Matérn fields on a sphere, Fig. 4 shows the locations of 7280 meteorological measurement stations on the globe, together with an irregular triangulation. The triangulation was constrained to have minimal angles 21° and maximum edge lengths corresponding to 500 km based on an average Earth radius of 6370 km. The triangulation includes all the stations more than 10 km apart, requiring a total of 15182 vertices and 30360 triangles. The resulting Gaussian field model for $\alpha = 2$ is illustrated in Fig. 5, for $\kappa^2 = 16$, corresponding to an approximate correlation range 0.7 on a unit radius globe. Numerically calculating the covariances between a point on the equator and all other points shows, in Fig. 5(a), that despite the highly irregular triangulation, the deviations from the theoretical covariances determined by the SPDE (calculated via a spherical Fourier series) are practically non-detectable for distances larger than the local edge length (≤ 0.08), and nearly un-detectable even for shorter distances. A random realisation from the model is shown in Fig. 5(b), resampled to a longitude/latitude grid with an area preserving cylindrical projection. The number of Markov neighbours of each node ranges from 10 to 34, with an average of 19. The resulting structure of the precision matrix is shown in Fig. 6(a), with the corresponding ordering of the nodes shown visually in Fig. 6(b) by mapping the node indices to grey-scales. The ordering uses the Markov graph structure to recursively divide the graph into conditionally independent sets (Karypis and Kumar, 1999), which helps make the Cholesky factor of the precision matrix sparse.

3.2. Non-stationary fields

From a traditional point of view, 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 (Sampson and Guttorp, 1992; Higdon, 1998; Hughes-Oliver et al., 1998; Cressie and Huang, 1999; Higdon et al., 1999; Fuentes, 2001; Gneiting, 2002; Stein, 2005; Paciorek and Schervish, 2006; Jun and Stein, 2008; Yue and Speckman, 2010). The SPDE approach has the additional huge advantage that the resulting (non-stationary) Gaussian field is a GMRF, which allows for swift computations and can additionally be defined on a manifold.

In the SPDE defined in (2), 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

$$(\kappa^2(\boldsymbol{u}) - \Delta)^{\alpha/2}(\tau(\boldsymbol{u})x(\boldsymbol{u})) = \mathcal{W}(\boldsymbol{u}). \tag{12}$$

For simplicity, we choose to keep the variance for the innovation constant and instead scale the resulting

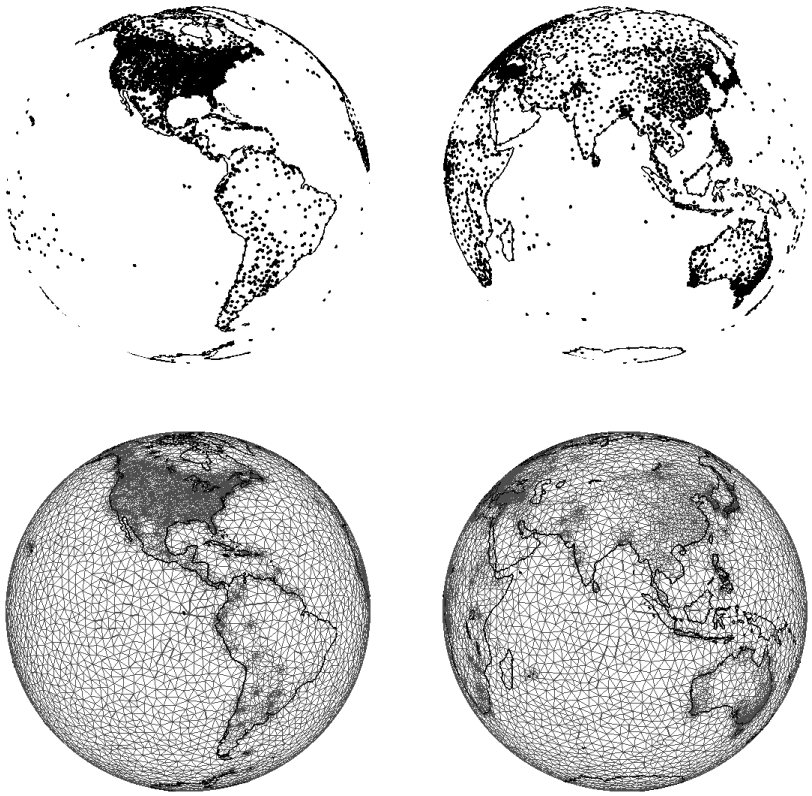
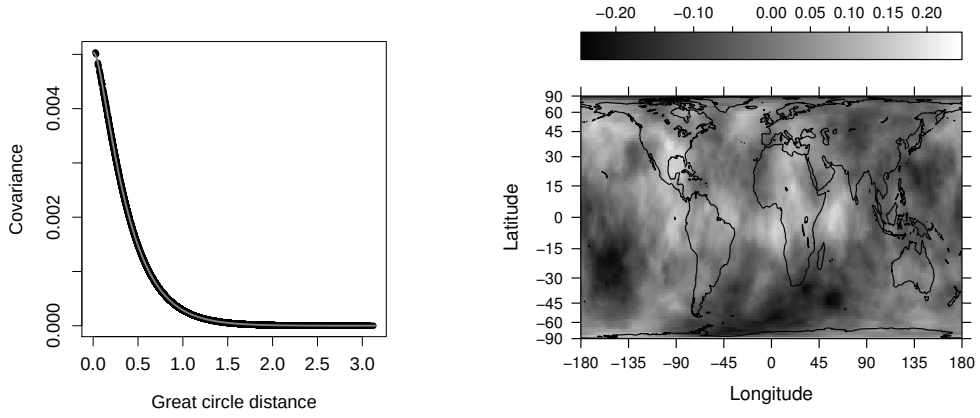
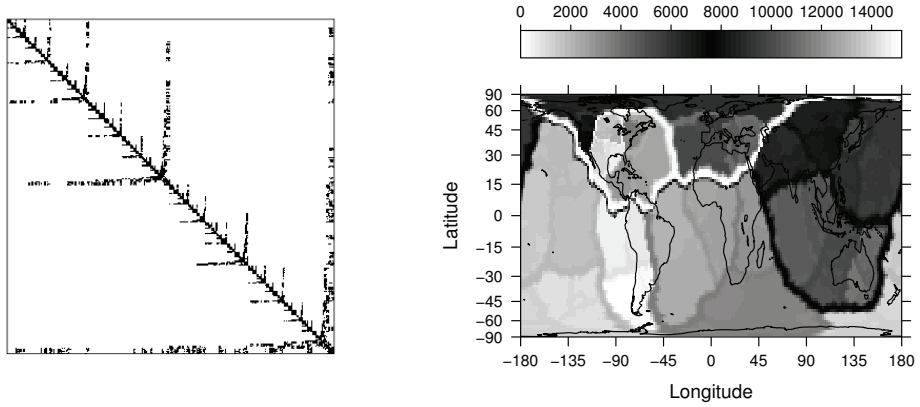


Fig. 4. Data locations (top row) and triangulation (bottom row) for the global temperature dataset analysed in Sec. 4, with a coastline map superimposed.



(a) (b)

Fig. 5. Covariances (a) and a random sample (b) from the stationary SPDE model (2) on the unit sphere, with $\nu = 1$ and $\kappa^2 = 16$. In (a), the black dots show the numerical result for the GMRF approximation, and the superimposed grey curve is the theoretical covariance function.



(a) (b)

Fig. 6. The structure of the (reordered) 15182×15182 precision matrix (a) and a visual representation of the reordering (b). The indices of each triangulation node has been mapped to grey-scales, showing the governing principle of the reordering algorithm, recursively dividing the graph into conditionally independent sets.

process $x(\boldsymbol{u})$ with a scaling parameter $\tau(\boldsymbol{u})$. Non-stationarity is achieved when one or both parameters are non-constant. Of particular interest is the case where they vary slowly with $\boldsymbol{u}$, for example in a low-dimensional representation like

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

where the basis functions $\{B_i^{(\cdot)}(\cdot)\}$ are smooth over the domain of interest. With slowly varying parameters $\kappa^2(\boldsymbol{u})$ and $\tau(\boldsymbol{u})$, the appealing *local* interpretation of (12) 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.

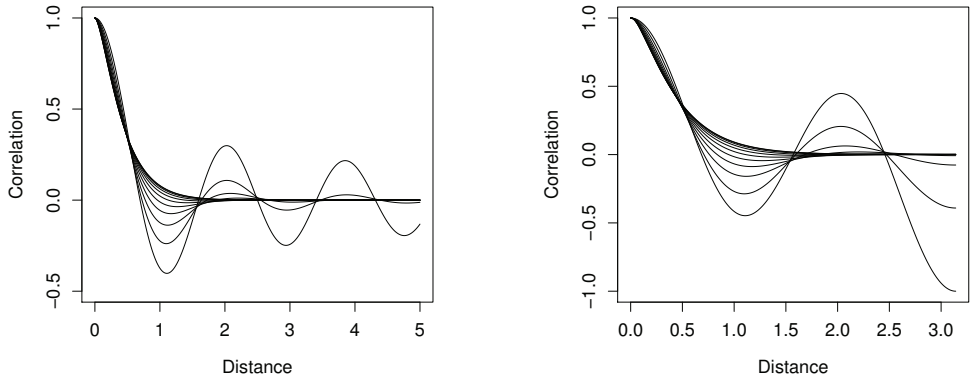
The GMRF representation of (12) is found using the same approach as for the stationary case, with minor changes. For convenience, we assume that both κ^2 and τ can be considered as constant within the support of the basis functions $\{\psi_k\}$, and hence

$$\langle \psi_i, \kappa^2 \psi_j \rangle = \int \psi_i(\boldsymbol{u}) \psi_j(\boldsymbol{u}) \kappa^2(\boldsymbol{u}) \, d\boldsymbol{u} \approx C_{ij} \kappa^2(\boldsymbol{u}_j^*) \tag{13}$$

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 (10) at no additional cost, see App. A.3. If we improve the integral approximation (13) 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

Another extension is to consider a complex version of the basic equation (2). For simplicity, we only consider the case $\alpha = 2$. With innovation processes $\mathcal{W}_1$ and $\mathcal{W}_2$ as two independent white noise fields,



(a) (b)

Fig. 7. Correlation functions from oscillating SPDE models, for $\theta = 0, 0.1, \ldots, 1$, on (a) $\mathbb{R}^2$ and on (b) $\mathbb{S}^2$, with $\kappa^2 = 12$, $\nu = 1$.

and an *oscillation* parameter θ, the complex version becomes

$$(\kappa^2 \mathrm{e}^{i\pi\theta} - \Delta)(x_1(\boldsymbol{u}) + i x_2(\boldsymbol{u})) = \mathcal{W}_1(\boldsymbol{u}) + i\mathcal{W}_2(\boldsymbol{u}), \quad 0 \leq \theta < 1. \tag{14}$$

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

$$R(\boldsymbol{k}) = (2\pi)^{-d} \left(\kappa^4 + 2\cos(\pi\theta)\kappa^2\|\boldsymbol{k}\|^2 + \|\boldsymbol{k}\|^4\right)$$

on $\mathbb{R}^d$. The corresponding covariance functions for $\mathbb{R}$ and $\mathbb{R}^2$ are given in App. A. For general manifolds, no closed form expression can be found. In Fig. 7, 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 App. A. The details of the construction, given in App. C.4, also reveal the possibility of multivariate fields, similar to Gneiting et al. (2010).

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, on $\mathbb{R}^d$ invariant to addition of cosine functions of arbitrary direction, with wave number κ.

3.4. Non-isotropic models and spatial deformations

The non-stationary model defined in Sec. 3.2, has locally isotropic correlations, despite having globally non-stationary correlations. This can be relaxed by widening the class of considered SPDEs, allowing a non-isotropic Laplacian, and also by including a directional derivative term. This also provides a link 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. Using the link to SPDE models, the resulting model can interpreted as a non-stationary SPDE in the original domain.

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 $\widetilde{\Omega}$,

$$(\kappa^2 - \widetilde{\nabla} \cdot \widetilde{\nabla})\widetilde{x}(\widetilde{\boldsymbol{u}}) = \widetilde{\mathcal{W}}(\widetilde{\boldsymbol{u}}), \tag{15}$$

generating a stationary Matérn field. A change of variables onto the undeformed space Ω yields (Smith, 1934)

$$\frac{1}{\det\{\boldsymbol{F}(\boldsymbol{u})\}}\left[\kappa^2 - \det\{\boldsymbol{F}(\boldsymbol{u})\}\nabla \cdot \frac{\boldsymbol{F}(\boldsymbol{u})\boldsymbol{F}(\boldsymbol{u})^{\mathsf{T}}}{\det\{\boldsymbol{F}(\boldsymbol{u})\}}\nabla\right] x(\boldsymbol{u}) = \frac{1}{\det\{\boldsymbol{F}(\boldsymbol{u})\}^{1/2}}\mathcal{W}(\boldsymbol{u}), \tag{16}$$

where $\boldsymbol{F}(\boldsymbol{u})$ is the Jacobian of the deformation function $\boldsymbol{f}$. This non-stationary SPDE exactly reproduces the deformation method with Matérn covariances (Sampson and Guttorp, 1992). A sparse GMRF approximation can be constructed using the same principles as for the simpler non-stationary model in Sec. 3.2.

An important remark is that the parameters of the resulting SPDE do not depend directly on the deformation function itself, but only its Jacobian. A possible option for parametrising 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, given either from covariate information or as a weighted sum of vector basis functions. Addition of subtraction of a directional derivative term further generalises the model. Allowing all parameters, including the variance of the white nose, to vary across the domain, results in a very general non-stationary model that includes both the deformation method and the model in Sec. 3.2. The model class can be interpreted as changes of metric in Riemannian manifolds, which is a natural generalisation of deformation between domains embedded in Euclidean spaces. A full analysis is beyond the scope of this paper, but the technical appendices cover much of the necessary theory.

3.5. Non-separable space-time models

A separable space-time covariance function can be characterised as having a spectrum that can be written as a product or sum of spectra in only space or time. In contrast, a non-separable model can have interaction between the space and time dependency structures. While it is difficult to explicitly construct non-separable non-stationary covariance functions, non-separable SPDE models can be obtained with relative ease, using locally specified parameters. Arguably, the most simple non-separable SPDE that can be applied to the GMRF method is the transport and diffusion equation,

$$\left\{\frac{\partial}{\partial t} + (\kappa^2 + \boldsymbol{m} \cdot \nabla - \nabla \cdot \boldsymbol{H}\nabla)\right\} x(\boldsymbol{u}, t) = \mathcal{E}(\boldsymbol{u}, t), \tag{17}$$

where $\boldsymbol{m}$ is a transport direction vector, $\boldsymbol{H}$ is a positive definite diffusion matrix (for general manifolds strictly a *tensor*), and $\mathcal{E}(\boldsymbol{u}, t)$ is a stochastic space-time noise field. It is clear that even this stationary formulation yields non-separable fields, since the spatio-temporal power spectrum of the solution is

$$R_x(\boldsymbol{k}, \omega) = R_{\mathcal{E}}(\boldsymbol{k}, \omega)\left\{(\omega + \boldsymbol{m} \cdot \boldsymbol{k})^2 + (\kappa^2 + \boldsymbol{k} \cdot \boldsymbol{H}\boldsymbol{k})^2\right\}^{-1}, \tag{18}$$

which is strictly non-separable even with $\boldsymbol{m} = \boldsymbol{0}$ and $\boldsymbol{H} = \boldsymbol{I}$. The driving noise is an important part of the specification, and may require an additional layer in the model. In order to ensure a desired regularity of the solutions, the noise process can be chosen to be white in time but with spatial dependence, such as a solution to $(\kappa^2 - \nabla \cdot \nabla)^{\alpha/2}\mathcal{E}(\boldsymbol{u}, t) = \mathcal{W}(\boldsymbol{u}, t)$, for some $\alpha \geq 1$, where $\mathcal{W}(\boldsymbol{u}, t)$ is space-time white noise. A GMRF representation can be obtained by first applying the ordinary spatial method, and then discretising the resulting system of coupled temporal SDEs with, for example, an Euler method. Allowing all the parameters to vary with location in space (and possibly in time) generates a large class of non-separable non-stationary models. The stationary models evaluated by Heine (1955) can be obtained as special cases.

4. Example: Global temperature reconstruction

4.1. Problem background

When analysing past observed weather and climate, the Global Historical Climatology Network (GHCN) data set‡ (Peterson and Vose, 1997) is commonly used. On 2010-08-08, the data contained meteorological observations from 7280 stations spread across continents, where each of the 597373 rows of observations contains the monthly mean temperatures from a specific station and year. The data spans the period 1702 through 2010, 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 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 sea surface temperature measurements. These analyses process the data in different ways to reduce the influence of station specific effects (a procedure knows as *homogenisation*), 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 griding 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 to reconstruct the entire spatio-temporal yearly (or even monthly) 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 covariance model, which reduces the computational burden. However, we will allow for regional deviations from rotational symmetry in the expectations. The model separates weather from climate by assuming that the climate can be parametrised by non-stationary 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 (12), given the climate. Using the triangulation from Fig. 4 with piecewise linear basis function, the GMRF representation given in Sec. A.3 will be used, with $\boldsymbol{x}_t$ denoting the discretised field at time t. To avoid complications due to temporal dependence between monthly values, we aggregate the measurements into yearly means, and model only the yearly average temperature at each location. A full analysis needs to take local station dependent effects into account. Here, we include only the effect of elevation. To incorporate a completely integrated station homogenisation procedure into the model would go far beyond the scope of this paper, and we therefore use the "adjusted" GHCN data set, which includes some outlier quality control and relative station calibrations.

4.2. Model summary

The climate and observation model is governed by a parameter vector $\boldsymbol{\theta} = \{\boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_\tau, \boldsymbol{\theta}_s, \theta_\epsilon\}$, and we denote the yearly temperature fields $\boldsymbol{x} = \{\boldsymbol{x}_t\}$ and the yearly observations $\boldsymbol{y} = \{\boldsymbol{y}_t\}$, with $t = 1970, \ldots, 1989$. Using basis function matrices $\boldsymbol{B}_\mu$ (all 49 spherical harmonics up to and including order 6, see Wahba, 1981), $\boldsymbol{B}_\kappa$ and $\boldsymbol{B}_\tau$ (B-splines of order 2 in $\sin(\text{latitude})$, shown in Fig. 8), the expectation field is given by $\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{\theta}} = \boldsymbol{B}_\mu \boldsymbol{\theta}_\mu$, the local spatial dependence $\kappa(\boldsymbol{u})$ is defined through $\log \boldsymbol{\kappa}^2 = \boldsymbol{B}_\kappa \boldsymbol{\theta}_\kappa$,

‡http://www.ncdc.noaa.gov/ghcn/ghcn.html

and the local variance scaling $\tau(\boldsymbol{u})$ is defined through $\log \boldsymbol{\tau} = \boldsymbol{B}_\tau \boldsymbol{\theta}_\tau$. The prior distribution for the climate field is chosen as approximate solutions to the SPDE $\Delta\mu(\boldsymbol{u}) = \sigma_\mu \mathcal{W}(\boldsymbol{u})$, where $\sigma_\mu \gg 0$, which provides natural relative prior weights for the spherical harmonic basis functions.

The yearly temperature fields $\boldsymbol{x}_t$ are defined conditionally on the climate as

$$(\boldsymbol{x}_t|\boldsymbol{\theta}) \sim \mathsf{N}(\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{\theta}}, \boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{\theta}}^{-1}),$$

where $\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{\theta}}$ is the GMRF precision corresponding to the model (12) with parameters determined by $(\boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_\tau)$. Introducing *observation matrices* $\boldsymbol{A}_t$, that extract the nodes from $\boldsymbol{x}_t$ for each observation, the observed yearly weather is modelled as

$$(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}) \sim \mathsf{N}(\boldsymbol{A}_t\boldsymbol{x}_t + \boldsymbol{S}_t\boldsymbol{\theta}_s, \boldsymbol{Q}_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}}^{-1}),$$

where $\boldsymbol{S}_t\boldsymbol{\theta}_s$ are station specific effects and $\boldsymbol{Q}_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}} = \boldsymbol{I}\exp(\theta_\epsilon)$ is the observation precision. Since we only use the data for illustrative purposes here, we will ignore all station specific effects except for elevation. We also ignore any remaining residual dependencies between consecutive years, analysing only the marginal distribution properties of each year.

The Bayesian analysis draws all its conclusions from the properties of the posterior distributions of $(\boldsymbol{\theta}|\boldsymbol{y})$ and $(\boldsymbol{x}|\boldsymbol{y})$, so that all uncertainty about the weather $\boldsymbol{x}_t$ is included in the distribution for the model parameters $\boldsymbol{\theta}$, and conversely for $\boldsymbol{\theta}$ and $\boldsymbol{x}_t$. One of the most important steps is how to determine the conditional distribution for the weather given observations and model parameters,

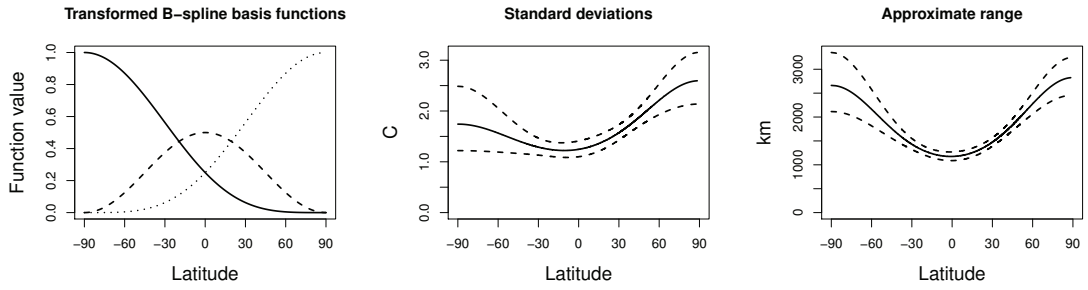
$$(\boldsymbol{x}_t|\boldsymbol{y}_t, \boldsymbol{\theta}) \sim \mathsf{N}\left\{\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{\theta}} + \boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}}^{-1}\boldsymbol{A}_t^{\mathsf{T}}\boldsymbol{Q}_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}}\left(\boldsymbol{y}_t - \boldsymbol{A}_t\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{\theta}} - \boldsymbol{S}_t\boldsymbol{\theta}_s\right), \boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}}^{-1}\right\},$$

where $\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}} = \boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{\theta}} + \boldsymbol{A}_t^{\mathsf{T}}\boldsymbol{Q}_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}}\boldsymbol{A}_t$ is the conditional precision, and the expectation is the Kriging estimator of $\boldsymbol{x}_t$. Due to the compact support of the basis functions, determined by the triangulation, each observation depends on at most three neighbouring nodes in $\boldsymbol{x}_t$, which makes the conditional precision have the same sparsity structure as the field precisions $\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{\theta}}$. The computational cost of the Kriging estimates is $\mathcal{O}(n)$ in the number of observations, and approximately $\mathcal{O}(n^{3/2})$ in the number of basis functions. If basis functions with non-compact support had been used, such as a Fourier basis, the posterior precisions would have been fully dense matrices, with computational cost $\mathcal{O}(n^3)$ in the number of basis functions, regardless of the sparsity of the prior precisions. This shows that when constructing computationally efficient models it is not enough to consider the theoretical properties of the prior model, but instead the whole sequence of computations needs to be taken into account.

4.3. Results

We implemented the model using `R-inla`. Since $(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ is Gaussian, the results are only approximate with regards to the numerical integration of the covariance parameters $(\boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_\tau, \boldsymbol{\theta}_\epsilon)$. Due to the large size of the data set, this initial analysis is based on data only from the period 1970 through 1989, requiring 336960 nodes in a joint model for the yearly temperature fields, measurements, and linear covariate parameters, with 15182 nodes in each field, and the number of observations in each year ranging between approximately 1300 and 1900, for each year including all stations with no missing monthly values. The full Bayesian analysis took about one hour to compute on a 12 core computer, with a peak memory use of about 50GB during the parallel numerical integration phase. This is a notable improvement over earlier work by Das (2000) where partial estimation of the parameters in a deformation based covariance model of the type in Sec. 3.4 took more than a week on a super-computer.

The 95% credible interval for the measurement standard deviation, including local unmodelled effects, was calculated to $(0.628, 0.650)$ °C, with posterior expectation 0.634 °C. The spatial covariance parameters are harder to interpret individually, but we instead show the resulting spatially varying field



(a) (b) (c)

Fig. 8. Three transformed B-spline basis functions of order 2 (a), and approximate 95% credible intervals for (b) standard deviation and (c) correlation range of the yearly weather, as functions of latitude.

standard deviations and correlation ranges in Fig. 8, including pointwise 95% credible intervals. Both curves show a clear dependence on latitude, with both larger variance and correlation range near the poles, compared with the equator. The standard deviations range between 1.2 and 2.6 °C, and the correlation ranges vary between 1175 and 2825 km. There is an asymmetric north/south pole effect for the variances, but a symmetric curve is admissible in the credible intervals.

Evaluating the estimated climate and weather for a period of only 20 years is difficult, since "climate" is typically defined as averages over periods of 30 years. Also, the spherical harmonics used for the climate model are not of high enough order to capture all regional effects. To alleviate these problems, we base the presentation on what can reasonably be called the *empirical* climate and weather anomalies for the period 1970 through 1989, in effect using the period average as reference. Thus, instead of evaluating the distributions of $(\boldsymbol{\mu}|\boldsymbol{y})$ and $(\boldsymbol{x}_t - \boldsymbol{\mu}|\boldsymbol{y})$, we instead consider $(\bar{\boldsymbol{x}}|\boldsymbol{y})$ and $(\boldsymbol{x}_t - \bar{\boldsymbol{x}}|\boldsymbol{y})$, where $\bar{\boldsymbol{x}} = \sum_{t=1970}^{1989} \boldsymbol{x}_t/20$. In Fig. 9(a) and (c), the posterior expectation of the empirical climate, $\mathsf{E}(\bar{\boldsymbol{x}}|\boldsymbol{y})$, is shown (including the estimated effect of elevation), together with the posterior expectation of the temperature anomaly for 1980, $\mathsf{E}(\boldsymbol{x}_{1980} - \bar{\boldsymbol{x}}|\boldsymbol{y})$. The corresponding standard deviations are shown in Fig. 9(b) and (d). As expected, the temperatures are low near the poles and high near the equator, and some of the relative warming effect of the thermohaline circulation on the Alaska and northern European climates can also be seen. There is a clear effect of regional topography, showing cold areas for high elevations such as in the Himalayas, Andes, and Rocky Mountains, as indicated by an estimated cooling effect of 5.2 °C per km of increased elevation. It is clear from Fig. 9(b) and (d) that including ocean based measurements is vital for analysis of regional ocean climate and weather, in particular for the south-east Pacific Ocean.

With this in mind, one might expect that the analysis period and data coverage is too restricted to allow detection of global trends, especially since the simple model we use a priori assumes a constant climate. However, the present analysis, including the effects of all parameter uncertainties, still yields a 95% Bayesian prediction interval $(0.87, 2.18)$ °C per century (expectation 1.52 °C) for the global average temperature trend over the analysed 20 year period. The posterior standard deviation for each global average temperature anomaly was calculated to about 0.09 °C. Comparing the values with the corresponding estimates in the GISS series, which has an observed trend of 1.48 °C per century for this period, yields a standard deviation for the differences between the series of only 0.04 °C. Thus, the results here are similar to the GISS results, even without the use of ocean data.

The estimated trend has less than a 2% probability of occurring in a random sample from the temporally stationary model used in the analysis. From a purely statistical point of view, this could indicate either that there is a large amount of unmodelled temporal correlation in the yearly weather averages, or that the expectation is non-stationary, i.e. that the climate was changing. Since it is impossible to distin-

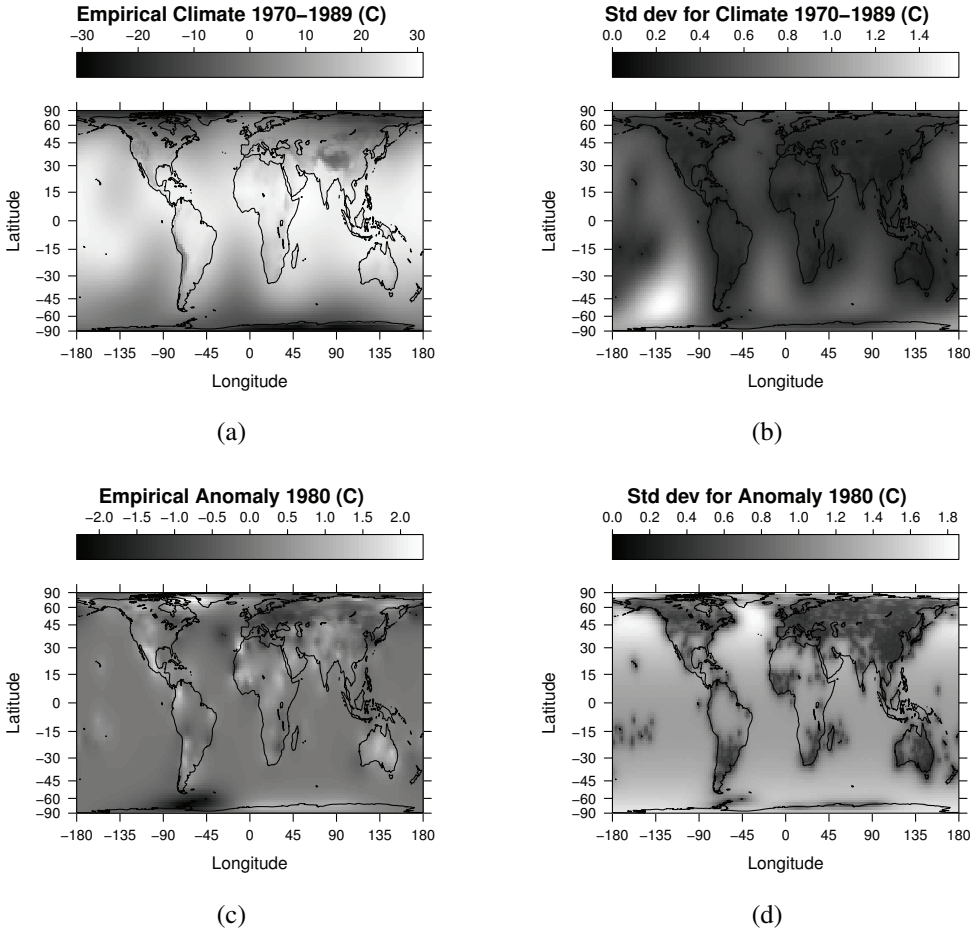


Fig. 9. Posterior means for the empirical 1970–1989 climate (a) and for the empirical mean anomaly 1980 (c), with the corresponding posterior standard deviations in (b) and (d). The climate includes the estimated effect of elevation. An area-preserving cylindrical projection is used.

guish between these two cases using only statistical methods on the single realisation of the actual climate and weather system that is available, a full analysis should incorporate knowledge from climate system physics to properly balance the change in climate and short-term dependence in the weather in the model.

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 (2011). 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 requires 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. The SPDE approach allows for easy construction of non-stationary models, defined in a natural way that provides good local interpretation, via spatially varying parameters, and is computationally very efficient, as we still obtain GMRF representations. 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 paper, such an approach can also provide a huge computational benefit.

On the negative side, the 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.

Acknowledgement

This paper is dedicated to the memory of Julian E. Besag (1945–2010), whose work on Markov random fields from 1974 onwards inspired us to investigate the link to Gaussian random field models commonly used in spatial statistics.

The authors would like to thank the Research Section Committee and the anonymous reviewers for their very helpful comments and suggestions. We are also grateful to Peter Guttorp for encouraging us to address the global temperature problem, Daniel Simpson for providing the convergence rate result (11), the key reference in App. C.5 and for invaluable discussions and comments, and to Georg Lindgren for numerous comments on the manuscript.

A. Explicit results

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

A.1. Regular lattices

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

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

where $\boldsymbol{H}$ is a diagonal d-dimensional matrix with positive diagonal elements (compare with Sec. 3.4).

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 $s_i = (\gamma_i + \delta_i)/2$. Since $d = 1$, we can write $\boldsymbol{H} = H \geq 0$, and the elements on row i, around the diagonal, of the precision are given by

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

where $a_i = H/(\gamma_i s_i)$, $b_i = H/(\delta_i s_i)$, and $c_i = \kappa^2 + a_i + b_i$. If the spacing is regular, $s = \delta = \gamma$, and $a = a_i = b_i \equiv H/\delta^2$ and $c = c_i \equiv \kappa^2 + 2a$. 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 $s = \gamma\delta$, $a = H_{11}/\gamma^2$, $b = H_{22}/\delta^2$, and $c = \kappa^2 + 2a + 2b$. The precision elements are then given by

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

If the grid distances are proportional to the square root of the corresponding diagonal elements of $\boldsymbol{H}$ (such as in the isotropic case $\gamma = \delta$ and $H_{11} = H_{22}$), the expressions simplify to $s = \gamma\delta$, $a = b = H_{11}/\gamma^2 = H_{22}/\delta^2$, and $c = \kappa^2 + 4a$.

A.2. 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. For implementation of the theory in App. C, we need to calculate

$$\widetilde{C}_{ii} = \langle \psi_i, 1 \rangle_\Omega \,, \qquad C_{ij} = \langle \psi_i, \psi_j \rangle_\Omega \,, \qquad G_{ij} = \langle \nabla\psi_i, \nabla\psi_j \rangle_\Omega \,, \qquad B_{ij} = \langle \psi_i, \partial_{\boldsymbol{n}} \psi_j \rangle_{\partial\Omega} \,. \tag{19}$$

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 (19), 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 $|T| = \|\boldsymbol{e}_0 \times \boldsymbol{e}_1\|/2$, 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

$$\left[\widetilde{C}_{i,i}(T)\right]_{i=0,1,2} = \frac{|T|}{3}\begin{bmatrix}1 & 1 & 1\end{bmatrix},$$
$$\left[C_{i,j}(T)\right]_{i,j=0,1,2} = \frac{|T|}{12}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & 1\\ 1 & 1 & 2\end{bmatrix}$$

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

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

For the boundary integrals in (19), the contribution from the triangle is

$$\left[B_{i,j}(T)\right]_{i,j=0,1,2} = \frac{-1}{4|T|}\begin{bmatrix}\boldsymbol{0} & \boldsymbol{e}_0 & \boldsymbol{e}_0\\ \boldsymbol{e}_1 & \boldsymbol{0} & \boldsymbol{e}_1\\ \boldsymbol{e}_2 & \boldsymbol{e}_2 & \boldsymbol{0}\end{bmatrix}^{\mathsf{T}}\begin{bmatrix}b_0\boldsymbol{I}\\ b_1\boldsymbol{I}\\ b_2\boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{e}_0 & \boldsymbol{e}_1 & \boldsymbol{e}_2\end{bmatrix},$$

where $b_k = \mathbb{I}(\text{Edge } k \text{ in } T \text{ lies on } \partial\Omega)$. 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, parametrised as in Sec. A.1 and Sec. C.4, the modified $\boldsymbol{G}$ matrix elements are given by

$$\left[G_{i,j}(T)\right]_{i,j=0,1,2} = \frac{1}{4|T|}\begin{bmatrix}\boldsymbol{e}_0 & \boldsymbol{e}_1 & \boldsymbol{e}_2\end{bmatrix}^{\mathsf{T}}\operatorname{adj}(\boldsymbol{H})\begin{bmatrix}\boldsymbol{e}_0 & \boldsymbol{e}_1 & \boldsymbol{e}_2\end{bmatrix}, \tag{20}$$

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

A.3. Non-stationary and oscillating models

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

$$\boldsymbol{Q}_2(\kappa^2, \theta) = \kappa^4\boldsymbol{C} + 2\kappa^2\cos(\pi\theta)\boldsymbol{G} + \boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G}, \tag{21}$$

where $\theta = 0$ corresponds to the regular Matérn case and $0 < \theta < 1$ are oscillating models. Using the approximation from (13), the non-stationary model (12) with $\alpha = 2$ has precision matrix given by

$$\boldsymbol{Q}_2(\kappa^2(\cdot), \tau(\cdot)) = \boldsymbol{\tau}\left(\boldsymbol{\kappa}^2\boldsymbol{C}\boldsymbol{\kappa}^2 + \boldsymbol{\kappa}^2\boldsymbol{G} + \boldsymbol{G}\boldsymbol{\kappa}^2 + \boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G}\right)\boldsymbol{\tau} \tag{22}$$

where $\boldsymbol{\kappa}^2$ and $\boldsymbol{\tau}$ are diagonal matrices, with $\kappa^2_{ii} = \kappa(\boldsymbol{u}_i)^2$ and $\tau_{ii} = \tau(\boldsymbol{u}_i)$. As shown in Sec. C.5, all the $\boldsymbol{C}$ should be replaced by $\widetilde{\boldsymbol{C}}$ to obtain a Markov model.

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

$$\begin{cases} (\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}), & \boldsymbol{u} \in \Omega \\ \partial_{\boldsymbol{n}} (\kappa^2 - \Delta)^j x(\boldsymbol{u}) = 0, & \boldsymbol{u} \in \partial\Omega,\ j = 0, 1, \ldots, \lfloor (\alpha - 1)/2 \rfloor, \end{cases} \tag{23}$$

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* (23) *for $\Omega = [0, L]$ and a positive integer α, then*

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

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 App. D.1. In practice, when the effective range is small compared to L, only the three main terms need to be included for a very close approximation:

$$\begin{aligned} \mathsf{Cov}(x(u), x(v)) &\approx r_M(u, v) + r_M(u, -v) + r_M(u, 2L - v) && (24) \\ &= r_M(0, v - u) + r_M(0, v + u) + r_M(0, 2L - (v + u)). && (25) \end{aligned}$$

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 covariance functions

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

$$r(u, v) = \frac{1}{2 \sin(\pi\theta) \kappa^3} e^{-\kappa \cos(\pi\theta/2)|v-u|} \sin\left\{ \pi\theta/2 + \kappa \sin(\pi\theta/2)|v - u| \right\}, \tag{26}$$

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

$$r(\boldsymbol{u}, \boldsymbol{v}) = \frac{1}{4\pi \sin(\pi\theta)\kappa^2 i} \left\{ K_0\left(\kappa \|\boldsymbol{v} - \boldsymbol{u}\| e^{-i\pi\theta/2} \right) - K_0\left(\kappa \|\boldsymbol{v} - \boldsymbol{u}\| e^{i\pi\theta/2} \right) \right\} \tag{27}$$

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

B. Manifolds, random fields, and operator identities

B.1. Manifold calculus

In order to state concisely the necessary theory for constructing solutions to stochastic partial differential equations on more general spaces than $\mathbb{R}^d$, we need to introduce some concepts from differential geometry and manifolds. A main point is that, loosely speaking, for statisticians familiar with measure theory and stochastic calculus on $\mathbb{R}^d$, many of the familiar rules for calculus for random processes and fields still apply, as long as all expressions are defined in coordinate-free manners. Here, we give a brief overview of the concepts uses in the subsequent appendices. 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 piecewise smooth $(d-1)$-manifold. We also require the manifolds to be *metric* manifolds, so that distances between points and angles between vectors are well-defined.

A *bounded* manifold has a finite maximal distance between points. If such a manifold is *complete* in the set-sense, it is called *compact*. Finally, if the manifold is compact but has no boundary, it is *closed*. The most common metric manifolds are subsets of $\mathbb{R}^d$ equipped with the Euclidean metric. The prime example of a closed manifold is the unit sphere $\mathbb{S}^2$ embedded in $\mathbb{R}^3$. In Fourier analysis for images, the *flat torus* commonly appears, when considering periodic continuations of a rectangular region. Topologically, this is equivalent to a torus, but with a different metric compared with a torus embedded in $\mathbb{R}^3$. The d-dimensional hypercube $[0,1]^d$ is a compact manifold with a closed boundary.

From the metric associated with the manifold it is possible to define differential operators. Let ϕ denote a function $\phi : \Omega \mapsto \mathbb{R}$. The *gradient* of ϕ at $\boldsymbol{u}$ is a vector $\nabla\phi(\boldsymbol{u})$ defined indirectly via directional derivatives. In $\mathbb{R}^d$ with Euclidean metric, the gradient operator ∇ is formally given by the column vector $\left[\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_d}\right]^{\mathsf{T}}$. The *Laplacian* Δ of ϕ at $\boldsymbol{u}$ (or the Laplace-Beltrami operator) can be defined as the sum of the second order directional derivatives, with respect to a local orthonormal basis, and is denoted $\Delta\phi(\boldsymbol{u}) = \nabla \cdot \nabla\phi(\boldsymbol{u})$. In Euclidean metric on $\mathbb{R}^d$, we can write $\Delta = \frac{\partial^2}{\partial u_1^2} + \cdots + \frac{\partial^2}{\partial u_d^2}$. At the boundary of Ω, the vector $\boldsymbol{n}_\partial(\boldsymbol{u})$ denotes the unit length outward normal vector at the point $\boldsymbol{u}$ on the boundary $\partial\Omega$. The *normal derivative* of a function ϕ is the directional derivative $\partial_{\boldsymbol{n}}\phi(\boldsymbol{u}) = \boldsymbol{n}_\partial(\boldsymbol{u}) \cdot \nabla\phi(\boldsymbol{u})$.

An alternative to defining integration on general manifolds through mapping subsets into $\mathbb{R}^d$, is to replace Lebesgue integration with integrals defined through normalised Hausdorff measures (Federer, 1951, 1978), here denoted $H^d_\Omega(\cdot)$. This leads to a natural generalisation of Lebesgue measure and integration, that coincides with the regular theory on $\mathbb{R}^d$. We write the area of a d-dimensional Hausdorff measurable subset $A \subset \Omega$ as $|A|_\Omega = H^d_\Omega(1_A)$, and the Hausdorff integral of a (measurable) function ϕ as $H^d_\Omega(\phi)$. An inner product between scalar or vector valued functions ϕ and ψ is defined through

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

A function $\phi : \Omega \mapsto \mathbb{R}^m$, $m \geq 1$, is said to be square integrable if and only if $\|\phi\|^2_\Omega = \langle \phi, \phi \rangle_\Omega < \infty$, denoted $\phi \in \mathrm{L}^2(\Omega)$.

A fundamental relation, that corresponds to integration by parts for functions on $\mathbb{R}$, is *Green's first identity*,

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

Typical statements of the identity require $\phi \in C^1(\Omega)$ and $\psi \in C^2(\Omega)$, but we will relax these requirements considerably in Lemma 1.

We also need to define Fourier transforms on general manifolds, where the usual cosine and sine functions do not exist.

DEFINITION 1 (GENERALISED FOURIER REPRESENTATION). *The Fourier transform pair for a function $\phi \in \mathrm{L}^2 : \mathbb{R}^d \mapsto \mathbb{R}$ 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 a function $\phi \in \mathrm{L}^2 : \Omega \mapsto \mathbb{R}$ is given by

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

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 2. *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 $\langle \phi, 1 \rangle_\Omega = 0$.

Note that for $\kappa > 0$, the norms are equivalent, and that the Hilbert space $\mathcal{H}^1$ is a quintessential Sobolev space.

B.2. Generalised Gaussian random fields

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

In classical theory for Gaussian fields, the following definition can be used:

DEFINITION 3 (GAUSSIAN FIELD). *A random function $x : \Omega \mapsto \mathbb{R}$ on a manifold Ω is a Gaussian field (GF) if $\{x(\boldsymbol{u}_k), k = 1, \ldots, n\}$ are jointly Gaussian random vectors for every finite set of points $\{u_k \in \Omega, k = 1, \ldots, n\}$. If there is a constant $b \geq 0$ such that $\mathsf{E}(x(\boldsymbol{u})^2) \leq b$ for all $\boldsymbol{u} \in \Omega$, the random field has bounded second moments.*

The complicating issue in dealing with the fractional SPDEs considered in this paper is that for some parameter values, the solutions themselves are discontinuous everywhere, although still more regular than white noise. Thus, since the solutions do not necessarily have well-defined pointwise meaning, the above definition is not applicable, and the driving white noise itself is also not a regular random field. Inspired by Adler and Taylor (2007), we solve this by using a generalised definition based on generalised functions.

DEFINITION 4 (GENERALISED FUNCTION). *For a given function space $\mathcal{F}$, an $\mathcal{F}$-*generalised function *$x : \Omega \mapsto \mathbb{R}$ with an associated generating additive measure $x^* : \mathcal{F} \mapsto \mathbb{R}$, is an equivalence class of objects identified through the collection of integration properties defined by $\langle \phi, x \rangle_\Omega = x^*(\phi)$, for all x^*-measurable functions $\phi \in \mathcal{F}$.*

When x^* is absolutely continuous with respect to the Hausdorff measure on Ω, x is a set of regular functions, at most differing on sets with Hausdorff measure zero. The definition allows many of the

regular integration rules to be used for generalised functions, without any need to introduce a heavy theoretical notational machinery, and provides a straightforward way of generalising Def. 3 to the kind of entities we need for the subsequent analysis.

DEFINITION 5 (GENERALISED GAUSSIAN FIELD). *A* generalised GF x *on* Ω *is a random* $\mathrm{L}^2(\Omega)$*-generalised function such that for every finite set of* test functions $\{\phi_i \in \mathrm{L}^2(\Omega),\, i = 1, \ldots, n\}$*, the inner products* $\langle \phi_i, x \rangle_\Omega$*,* $i = 1, \ldots, n$*, are jointly Gaussian. If there is a constant* $b \geq 0$ *such that* $\mathsf{E}(\langle \phi, x \rangle_\Omega^2) \leq b \, \|\phi\|_\Omega^2$ *for every* $\phi \in \mathrm{L}^2(\Omega)$*, the generalised field* x *has* $\mathrm{L}^2(\Omega)$*-bounded second moments, abbreviated as* is $\mathrm{L}^2(\Omega)$-bounded*.*

Of particular importance is the fact that white noise can be defined directly as a generalised GF.

DEFINITION 6 (GAUSSIAN WHITE NOISE). Gaussian white noise $\mathcal{W}$ *on a manifold* Ω *is an* $\mathrm{L}^2(\Omega)$*-bounded generalised GF such that for any set of test functions* $\{\phi_i \in \mathrm{L}^2(\Omega),\, i = 1, \ldots, n\}$*, the integrals* $\langle \phi_i, \mathcal{W} \rangle_\Omega$*,* $i = 1, \ldots, n$*, are jointly Gaussian, with expectation and covariance measures given by*

$$\mathsf{E}(\langle \phi_i, \mathcal{W} \rangle_\Omega) = 0 \quad \textit{and} \quad \mathsf{Cov}\big(\langle \phi_i, \mathcal{W} \rangle_\Omega \,,\, \langle \phi_j, \mathcal{W} \rangle_\Omega\big) = \langle \phi_i, \phi_j \rangle_\Omega \,.$$

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 equal to the area of the region.*

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 (Adler and Taylor, 2007), 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

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 generalised functions and random fields. Here, we present the two fundamental identities needed for the subsequent SPDE analysis; Green's first identity and a scalar product characterisation of the half-Laplacian.

B.3.1. Stochastic Green's first identity

We here state a generalisation of Green's first identity, showing that the identity applies to generalised fields, as opposed to only differentiable functions.

LEMMA 1. *If* $\nabla f \in \mathrm{L}^2(\Omega)$ *and* Δx *is* $\mathrm{L}^2(\Omega)$*-bounded, then (with probability* 1*)*

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

If ∇x *is* $\mathrm{L}^2(\Omega)$*-bounded and* $\Delta f \in \mathrm{L}^2(\Omega)$ *, then (with probability* 1*)*

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

For brevity, we include only a sketch of the proof:

PROOF. The requirements imply that each integrand can be approximated arbitrarily closely in the L^2 senses using C^q functions $\widetilde{f}$ and $\widetilde{x}$, where q in each case is large enough for the regular Green's identity to hold for $\widetilde{f}$ and $\widetilde{x}$. Using the triangle inequality, it follows that the expectation of the squared difference between the left and right hand sides of the identity can be bounded by an arbitrarily small positive constant. Hence, the difference is zero in quadratic mean, and the identity holds with probability 1. □

B.3.2. The half-Laplacian

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. 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:

$$\left\{\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2}\phi\right\}(\boldsymbol{k}) = (\kappa^2 + \|\boldsymbol{k}\|^2)^{\alpha/2}(\mathcal{F}\phi)(\boldsymbol{k}), \quad \text{on } \mathbb{R}^d$$
$$\left\{\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2}\phi\right\}(k) = (\kappa^2 + \lambda_k)^{\alpha/2}(\mathcal{F}\phi)(k), \quad \text{on compact } \Omega,$$

where λ_k, $k = 0, 1, 2, \ldots$, are the eigenvalues of $-\Delta$. The formal definition is mostly of theoretical interest, since in practice, the generalised Fourier basis and eigenvalues for the Laplacian are unknown. In addition, even if the functions are known, working directly in the Fourier basis is computationally expensive for general observation models, since the basis functions do not have compact support, which leads to dense covariance and precision matrices. The following Lemma provides an integration identity that allows practical calculations involving the half-Laplacian.

LEMMA 2. *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 (a) $\Omega = \mathbb{R}^d$, (b) Ω is closed, or (c) Ω is compact and $\langle \phi, \partial_{\boldsymbol{n}}\psi \rangle_{\partial\Omega} = \langle \partial_{\boldsymbol{n}}\phi, \psi \rangle_{\partial\Omega} = 0$.

For proof, see App. D.2. The Lemma shows that for functions ψ fulfilling the requirements, we can use the Hilbert space inner product as a definition of the half-Laplacian. This also generalises in a natural way to random fields x with $\mathrm{L}^2(\Omega)$-bounded ∇x, as well as to suitably well-behaved unbounded manifolds.

It would be tempting to eliminate the qualifiers in part (c) of the Lemma by subtracting the average of the two boundary integrals to the relation, and extend the Lemma to a complete equivalence relation. However, the motivation may be problematic, since the half-Laplacian is defined for a wider class of functions than the Laplacian, and it is unclear whether such a generalisation necessarily yields the same half-Laplacian as the Fourier definition for functions that are not of the class $\Delta\phi \in \mathrm{L}^2(\Omega)$. See Ilić et al. (2008) for a partial result.

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 App. C.1, we state the needed convergence and stochastic finite element method definitions. The main result for Matérn covariance models is stated in App. C.2, followed by generalisations to intrinsic and oscillating fields in App. C.3 and App. C.4. Finally, the full finite element constructions are modified to Markov models in App. C.5.

C.1. Weak convergence and stochastic FEM

We start by stating formal definitions of convergence of Hilbert spaces and of random fields in such spaces (Def. 7 and 8) as well as the definition of the finite element constructions that will be used (Def. 9).

DEFINITION 7 (DENSE SUBSPACE SEQUENCES). *A finite subspace $\mathcal{H}^1_n(\Omega, \kappa) \subset \mathcal{H}^1(\Omega, \kappa)$ is spanned by a finite set of basis functions $\Psi_n = \{\psi_1, \ldots, \psi_n\}$. We say that a sequence of subspaces $\{\mathcal{H}^1_n\}$ is dense in $\mathcal{H}^1$ if for every $f \in \mathcal{H}^1$ there is a sequence $\{f_n\}$, $f_n \in \mathcal{H}^1_n$, such that $\lim_{n\to\infty} \|f - f_n\|_{\mathcal{H}^1(\Omega,\kappa)} = 0$.*

If the subspace sequence is *nested*, there is a monotonely convergent sequence $\{f_n\}$, but that is not a requirement here. For given $\mathcal{H}^1_n$, we can choose the projection of $f \in \mathcal{H}^1$ onto $\mathcal{H}^1_n$, i.e. the f_n that minimises $\|f - f_n\|_{\mathcal{H}^1}$. The error $f - f_n$ is orthogonal to $\mathcal{H}^1_n$, and the basis coordinates can be determined via the system of equations $\langle \psi_k, f_n \rangle_{\mathcal{H}^1(\Omega,\kappa)} = \langle \psi_k, f \rangle_{\mathcal{H}^1(\Omega,\kappa)}$, for all $k = 1, \ldots, n$.

DEFINITION 8 (WEAK CONVERGENCE). *A sequence of* $\mathrm{L}^2(\Omega)$*-bounded generalised Gaussian fields* $\{x_n\}$ *is said to* converge weakly *to an* $\mathrm{L}^2(\Omega)$*-bounded generalised Gaussian field* 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 by* $x_n \xrightarrow{D(\mathrm{L}^2(\Omega))} x$*.*

DEFINITION 9 (FINITE ELEMENT APPROXIMATIONS). *Let* $\mathcal{L}$ *be a second order elliptic differential operator, and let* $\mathcal{E}$ *be a generalised Gaussian field 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 solutions *are given by Gaussian* $\boldsymbol{w} = \{w_1, \ldots, w_n\}$ *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 every pair of test functions $f_n, g_n \in \mathcal{H}^1_n(\Omega, \kappa)$*.*

b) The weak least squares solutions *are given by Gaussian* $\boldsymbol{w} = \{w_1, \ldots, w_n\}$ *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 every pair of test functions $f_n, g_n \in \mathcal{H}^1_n(\Omega, \kappa)$*.*

C.2. The basic Matérn-like cases

In the remainder of the appendices, we let $\mathcal{L} = (\kappa^2 - \Delta)$. In the classic Matérn case, the SPDE $\mathcal{L}^{\alpha/2} x = \mathcal{W}$ can, for integer α values, be unravelled into an iterative formulation

$$\mathcal{L}^{1/2} y_1 = \mathcal{W}, \quad \mathcal{L} y_2 = \mathcal{W}, \quad \mathcal{L} y_k = y_{k-2}, \quad k = 3, 4, \ldots, \alpha.$$

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 will 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.

We now formulate the three main theorems of the paper, that show what the precision matrices should look like for given basis functions (Thm. 2), that the finite Hilbert representations converge to the true distributions for $\alpha = 1$ and $\alpha = 2$ and dense Hilbert space sequences (Thm. 3), and finally that the iterative constructions for $\alpha \geq 3$ also converge (Thm. 4). Note that a sequence $\mathcal{H}^1_n(\Omega, \kappa)$ of piecewise linear Hilbert spaces defined on non-degenerate triangulations of Ω is a dense sequence in $\mathcal{H}^1(\Omega, \kappa)$ if the maximal edge length decreases to zero. 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 (FINITE ELEMENT PRECISIONS). *Define matrices $\boldsymbol{C}$, $\boldsymbol{G}$, and $\boldsymbol{K}$ through*

$$C_{i,j} = \langle \psi_i, \psi_j \rangle_\Omega, \quad G_{i,j} = \langle \nabla\psi_i, \nabla\psi_j \rangle_\Omega, \quad \boldsymbol{K} = \kappa^2 \boldsymbol{C} + \boldsymbol{G}$$

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 $\mathcal{L}^{\alpha/2} x = \mathcal{E}$, $\mathcal{L} = (\kappa^2 - \Delta)$, with Neumann boundaries, and $\partial_{\boldsymbol{n}} \psi_k = 0$ on $\partial\Omega$.

a) When $\alpha = 2$ and $\mathcal{E} = \mathcal{W}$, the weak Galerkin solution is obtained for $\boldsymbol{Q} = \boldsymbol{K}^\mathsf{T} \boldsymbol{C}^{-1} \boldsymbol{K}$.

b) When $\alpha = 1$ and $\mathcal{E} = \mathcal{W}$, the weak least squares solution is obtained for $\boldsymbol{Q} = \boldsymbol{K}$.

c) When $\alpha = 2$ and $\mathcal{E}$ is an $\mathrm{L}^2(\Omega)$-bounded Gaussian field in $\mathcal{H}^1_n(\Omega, \kappa)$ with mean zero and precision $\boldsymbol{Q}_{\mathcal{E},n}$, the weak Galerkin solution is obtained for $\boldsymbol{Q} = \boldsymbol{K}^\mathsf{T} \boldsymbol{C}^{-1} \boldsymbol{Q}_{\mathcal{E},n} \boldsymbol{C}^{-1} \boldsymbol{K}$.

THEOREM 3 (CONVERGENCE). *Let x be a weak solution to the SPDE $\mathcal{L}^{\alpha/2} x = \mathcal{W}$, $\mathcal{L} = (\kappa^2 - \Delta)$, with Neumann boundaries on a manifold Ω, and let x_n be a weak $\mathcal{H}^1_n(\Omega, \kappa)$ approximation, when $\mathcal{W}$ is Gaussian white noise. Then,*

$$x_n \xrightarrow{D(\mathrm{L}^2(\Omega))} x, \tag{I}$$

$$\mathcal{L}^{\alpha/2} x_n \xrightarrow{D(\mathrm{L}^2(\Omega))} \mathcal{L}^{\alpha/2} x, \tag{II}$$

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

(a) $\alpha = 2$, and x_n is the Galerkin solution, or
(b) $\alpha = 1$ and x_n is the least squares solution.

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

$$y_n \xrightarrow{D(\mathrm{L}^2(\Omega))} y, \tag{I}$$

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

$$x_n \xrightarrow{D(\mathrm{L}^2(\Omega))} x, \tag{II}$$

$$\mathcal{L} x_n \xrightarrow{D(\mathrm{L}^2(\Omega))} \mathcal{L} x. \tag{III}$$

For proofs of the three theorems, see App. D.3.

C.3. The intrinsic cases

When $\kappa = 0$, the Hilbert space from Def. 2 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 partially 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. How to construct models with more fine-tuned control of the null-space is a subject for further research.

To approximate intrinsic fields with $\alpha \geq 2$ and free boundaries, the matrix $\boldsymbol{K}$ in Thm. 2 should be replaced by $\boldsymbol{G} - \boldsymbol{B}$ (due to 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 Thm. 3 and Thm. 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 $\mathbb{R}^2$ converges to the *de Wij* process, which is an intrinsic generalised Gaussian random field. As can be seen in App. A.1, for $\alpha = 1$ and $\kappa = 0$, the $\boldsymbol{Q}$-matrix (equal to $\boldsymbol{G}$) 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 Thm. 3 and Def. 8. 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 Sec. 3.3, as well as non-isotropic versions, we introduce a coupled system of SPDEs for $\alpha = 2$,

$$\begin{bmatrix} h_1 - \nabla \cdot \boldsymbol{H}_1 \nabla & -h_2 + \nabla \cdot \boldsymbol{H}_2 \nabla \\ h_2 - \nabla \cdot \boldsymbol{H}_2 \nabla & h_1 - \nabla \cdot \boldsymbol{H}_1 \nabla \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \end{bmatrix} \tag{28}$$

which is equivalent to the complex SPDE

$$(h_1 + i h_2 - \nabla \cdot (\boldsymbol{H}_1 + i\boldsymbol{H}_2)\nabla)(x_1(\boldsymbol{u}) + i x_2(\boldsymbol{u})) = \mathcal{E}_1(\boldsymbol{u}) + i\mathcal{E}_2(\boldsymbol{u}). \tag{29}$$

The model in Sec. 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 (28) 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:

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

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 $\boldsymbol{Q} = \boldsymbol{Q}(h_1, \boldsymbol{H}_1) + \boldsymbol{Q}(h_2, \boldsymbol{H}_2)$, and $\boldsymbol{Q}(\cdot, \cdot)$ is the precision generated for the regular iterated model with 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. 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, following standard practice in FEM, $\boldsymbol{C}$ can be approximated with a diagonal matrix as follows. Let $\widetilde{\boldsymbol{C}}$ be a diagonal matrix, with $\widetilde{C}_{ii} = \sum_j C_{ij} = \langle \psi_i, 1 \rangle_\Omega$, and note that this preserves the interpretation of the matrix as an integration matrix. Substituting $\boldsymbol{C}^{-1}$ with $\widetilde{\boldsymbol{C}}^{-1}$ yields a Markov approximation to the FEM solution.

The convergence rate for the Markov approximation is the same as for the full FEM model, which can be shown by adapting the details of the convergence proofs. 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 Markov $(\widetilde{x}_n)$ and the full FEM solution (x_n) for $\alpha = 2$ yields the error

$$\mathsf{Cov}(\langle f, \mathcal{L}\widetilde{x}_n \rangle_\Omega, \langle g, \mathcal{L}\widetilde{x}_n \rangle_\Omega) - \mathsf{Cov}(\langle f, \mathcal{L}x_n \rangle_\Omega, \langle g, \mathcal{L}x_n \rangle_\Omega) = \boldsymbol{w}_f(\widetilde{\boldsymbol{C}} - \boldsymbol{C})\boldsymbol{w}_g.$$

Requiring $\|f\|_{\mathcal{H}^1(\Omega,\kappa)}, \|g\|_{\mathcal{H}^1(\Omega,\kappa)} \leq 1$, it follows from Lemma 1 in Chen and Thomée (1985) that the covariance error is bounded by ch^2, where c is some constant, and h is the diameter of the largest circle that can be inscribed in a triangle of the triangulation. This shows that the convergence rate from (11) will not be affected by the Markov approximation. In practice, the $\boldsymbol{C}$ matrix in $\boldsymbol{K}$ should also be replaced by $\widetilde{\boldsymbol{C}}$. This improves the approximation when either h or κ is large, with numerical comparisons showing a covariance error reduction of as much as a factor 3. See Bolin and Lindgren (2009) 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 (THM. 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,

$$\mathsf{Cov}(x(u), x(v)) = \lambda_0 + \sum_{k=1}^{\infty} \cos(u\pi k/L)\cos(v\pi k/L)\lambda_k, \tag{30}$$

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. The Matérn covariance on $\mathbb{R}$ (with variance given by the SPDE) 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, with $\widetilde{r}(u, v)$ denoting the folded covariance in the theorem statement,

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

Rewriting the cosines via Euler's formulas, we obtain

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

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

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

which is precisely the sought expression in (30). □

D.2. Modified half-Laplacian equivalence

PROOF (LEMMA 2). For the sake of brevity, we only present the proof for compact manifolds, as the the proof for $\Omega = \mathbb{R}^d$ follows the same principle but without the boundary complications. The main difference is that the Fourier representation is discrete for compact manifolds and continuous for $\mathbb{R}^d$.

Let $\lambda_k \geq 0$, $k = 0, 1, 2, \ldots$, be the eigenvalue corresponding to eigenfunction E_k of $-\Delta$ (Definition 1). Then, with $\widehat{\phi}(k) = (\mathcal{F}\phi)(k)$, the modified half-Laplacian from App. B.3.2 is defined through $\mathcal{F}\{(\kappa^2-\Delta)^{1/2}\phi\}(k) = (\kappa^2+\lambda_k)^{1/2}\widehat{\phi}(k)$, and we obtain

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

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

$$= \sum_{k=0}^{\infty}(\kappa^2+\lambda_k)\widehat{\phi}(k)\widehat{\psi}(k),$$

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,

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

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

$$\langle \nabla E_k, \nabla E_{k'} \rangle_\Omega = \langle E_k, -\Delta E_{k'} \rangle_\Omega + \langle E_k, \partial_{\boldsymbol{n}} E_{k'} \rangle_{\partial\Omega} = \lambda_{k'} \langle E_k, E_{k'} \rangle_\Omega + \langle E_k, \partial_{\boldsymbol{n}} E_{k'} \rangle_{\partial\Omega} .$$

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

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

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

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

and the proof is complete. □

D.3. Hilbert space convergence

PROOF (THM. 2, FINITE ELEMENT PRECISIONS). 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 $f_n, g_n \in \mathcal{H}^1_n(\Omega, \kappa)$, and let $\mathcal{L} = (\kappa^2 - \Delta)$.

For case a), $\alpha = 2$ and $\mathcal{E} = \mathcal{W}$, so that

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

due to Green's identity, and

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

This covariance is equal to

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

for every pair of test functions f_n, g_n when $\boldsymbol{Q} = \mathsf{Cov}(\boldsymbol{w}, \boldsymbol{w})^{-1} = \boldsymbol{K}^\mathsf{T} \boldsymbol{C}^{-1} \boldsymbol{K}$.

For case b), $\alpha = 1$ and $\mathcal{E} = \mathcal{W}$. Using the same technique as in a), but with Lemma 2 instead of Green's identity, $\left\langle \mathcal{L}^{1/2} f_n, \mathcal{L}^{1/2} x_n \right\rangle_\Omega = \langle f_n, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)} = \boldsymbol{w}_f^\mathsf{T} \boldsymbol{K} \boldsymbol{w}$ and

$$\mathsf{Cov}\left(\left\langle \mathcal{L}^{1/2} f_n, \mathcal{W}\right\rangle_\Omega, \left\langle \mathcal{L}^{1/2} g_n, \mathcal{W}\right\rangle_\Omega\right) = \left\langle \mathcal{L}^{1/2} f_n, \mathcal{L}^{1/2} g_n\right\rangle_\Omega = \langle f_n, g_n\rangle_{\mathcal{H}^1(\Omega,\kappa)} = \boldsymbol{w}_f^\mathsf{T} \boldsymbol{K} \boldsymbol{w}_g$$

so that $\boldsymbol{Q} = \boldsymbol{K}^\mathsf{T} \boldsymbol{K}^{-1} \boldsymbol{K} = \boldsymbol{K}$, noting that $\boldsymbol{K}$ is a symmetric matrix since both $\boldsymbol{C}$ and $\boldsymbol{G}$ are symmetric.

Finally, for case c), $\alpha = 2$ and $\mathcal{E} = \mathcal{E}_n$ is a Gaussian field on $\mathcal{H}^1_n(\Omega, \kappa)$ 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 the finite basis representation of the noise $\mathcal{E}_n$ gives $\mathsf{Cov}(\langle f_n, \mathcal{E}_n\rangle_\Omega, \langle g_n, \mathcal{E}_n\rangle_\Omega) = \boldsymbol{w}_f^\mathsf{T} \boldsymbol{C} \boldsymbol{Q}_{\mathcal{E},n}^{-1} \boldsymbol{C} \boldsymbol{w}_g$. Requiring equality for all pairs of test functions yields $\boldsymbol{Q} = \boldsymbol{K}^\mathsf{T} \boldsymbol{C}^{-1} \boldsymbol{Q}_{\mathcal{E},n} \boldsymbol{C}^{-1} \boldsymbol{K}$. Here, keeping the transposes allows the proof to apply also to the intrinsic free boundary cases. □

PROOF (THM. 3, CONVERGENCE). First, we show that part (I) follows from part (II). Let $\mathcal{L} = (\kappa^2 - \Delta)$, let f and g be functions in $\mathcal{H}^1(\Omega, \kappa)$, and let $\tilde{f}$ be the solution to the PDE

$$\begin{cases} \mathcal{L}\tilde{f}(\boldsymbol{u}) = f(\boldsymbol{u}), & \boldsymbol{u} \in \Omega, \\ \partial_{\boldsymbol{n}} \tilde{f}(\boldsymbol{u}) = 0, & \boldsymbol{u} \in \partial\Omega, \end{cases}$$

and correspondingly for g. Then $\tilde{f}$ and $\tilde{g}$ are in $\mathcal{H}^1(\Omega, \kappa)$, and further fulfil the requirements of Lemma 1 and Lemma 2. Therefore,

$$\begin{aligned} \langle f, x_n\rangle_\Omega = \langle \mathcal{L}\tilde{f}, x_n\rangle_\Omega = \langle \tilde{f}, x_n\rangle_{\mathcal{H}^1(\Omega,\kappa)} = \langle \tilde{f}, \mathcal{L} x_n\rangle_\Omega, \quad \text{and} \\ \langle f, x\rangle_\Omega = \langle \mathcal{L}\tilde{f}, x\rangle_\Omega = \langle \tilde{f}, x\rangle_{\mathcal{H}^1(\Omega,\kappa)} = \langle \tilde{f}, \mathcal{L} x\rangle_\Omega, \end{aligned}$$

where the last equality holds when $\alpha = 2$, since $\mathcal{W}$ is $\mathrm{L}^2(\Omega)$-bounded. The convergence of x_n to x follows from part (II). In the Galerkin case (a), we have

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

and similarly for the Least Squares case (b).

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

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

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 2,

$$\left\langle \mathcal{L}^{1/2} f, \mathcal{L}^{1/2} x_n \right\rangle_\Omega = \langle f, x_n \rangle_{\mathcal{H}^1(\Omega,\kappa)} = \langle f_n, x_n \rangle_{\mathcal{H}^1(\Omega,\kappa)} ,$$

and

$$\begin{aligned}
&\mathsf{Cov}\left(\left\langle \mathcal{L}^{1/2} f, \mathcal{L}^{1/2} x_n \right\rangle_\Omega, \left\langle \mathcal{L}^{1/2} g, \mathcal{L}^{1/2} x_n \right\rangle_\Omega\right) = \mathsf{Cov}(\langle f_n, x_n \rangle_{\mathcal{H}^1(\Omega,\kappa)}, \langle g_n, x_n \rangle_{\mathcal{H}^1(\Omega,\kappa)}) \\
&\quad = \mathsf{Cov}\left(\left\langle \mathcal{L}^{1/2} f_n, \mathcal{W} \right\rangle_\Omega, \left\langle \mathcal{L}^{1/2} g_n, \mathcal{W} \right\rangle_\Omega\right) = \langle f_n, g_n \rangle_{\mathcal{H}^1(\Omega,\kappa)} \\
&\quad \to \langle f, g \rangle_{\mathcal{H}^1(\Omega,\kappa)} = \left\langle \mathcal{L}^{1/2} f, \mathcal{L}^{1/2} g \right\rangle_\Omega = \mathsf{Cov}\left(\left\langle \mathcal{L}^{1/2} f, \mathcal{W} \right\rangle_\Omega, \left\langle \mathcal{L}^{1/2} g, \mathcal{W} \right\rangle_\Omega\right)
\end{aligned}$$

as $n \to \infty$. □

PROOF (THM. 4, ITERATIVE CONVERGENCE). First, we show that part (II) follows from part (III). Let $\tilde{f}$ and $\tilde{g}$ be defined as in the proof of Thm. 3. Then, since $\mathcal{L} = \kappa^2 - \Delta$,

$$\langle f, x_n \rangle_\Omega = \langle \tilde{f}, \mathcal{L} x_n \rangle_\Omega \quad \text{and} \quad \langle f, x \rangle_\Omega = \langle \tilde{f}, \mathcal{L} x \rangle_\Omega ,$$

and the convergence of x_n to x follows from part (III). For part (III), as in the proof of Thm. 3, $\langle f, \mathcal{L} x_n \rangle_\Omega = \langle f_n, x_n \rangle_{\mathcal{H}^1(\Omega,\kappa)}$, and

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

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

References

Adler, R. J. (2009). *The Geometry of Random Fields*. In Applied Mathematics, Volume 62 of Classics in applied mathematics. Society for Industrial & Applied Mathematics.

Adler, R. J. and J. Taylor (2007). *Random Fields and Geometry*. Springer Monographs in Mathematics. Springer.

Allcroft, D. J. and C. A. Glasbey (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 D. Gasbarra (1996). Bayesian inference of survival probabilities, under stochastic ordering constraints. *Journal of the American Statistical Association 91*(435), 1101–1109.

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

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

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

Bansal, R., L. H. Staib, D. Xu, H. Zhu, and B. S. Peterson (2007). Statistical analyses of brain surfaces using Gaussian random fields on 2-D manifolds. *IEEE Transaction on Medical Imaging 26*(1), 46–57.

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-dimensional recurrence equations. *Journal of the Royal Statistical Society, Series B 43*(3), 302–309.

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

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

Besag, J., J. York, and A. Mollié (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 F. Lindgren (2009). Wavelet Markov models as efficient alternatives to tapering and convolution fields. *Preprints in Mathematical Sciences, Lund University* (2009:13), submitted.

Bolin, D. and F. Lindgren (2011). Spatial models generated by nested stochastic partial differential equations. *Annals of Applied Statistics*, to appear.

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

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

Chen, C. M. and V. Thomée (1985). The lumped mass finite element method for a parabolic problem. *Journal of the Australian Mathematical Society, Series B 26*, 329–354.

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

Ciarlet, P. G. (1978). *The finite element method for elliptic problems*. North-Holland Pub. Co.

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

Cressie, N. and N. Verzelen (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. New York: John Wiley & Sons Inc. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

Cressie, N. A. C. and G. Johannesson (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 H. R. Künsch (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika 74*(4), 877–882.

Das, B. (2000). *Global covariance modeling: a deformation approach to anisotropy*. Ph. D. thesis, University of Washington, Department of Statistics.

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

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

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

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

Eidsvik, J., A. O. Finley, S. Banerjee, and H. Rue (2010). Approximate bayesian inference for large spatial datasets using predictive process models. Technical report 9, Department of mathematical sciences, Norwegian University of Science and Technology.

Federer, H. (1951). Hausdorff measure and Lebesgue area. *Proceedings of 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., M. G. Genton, and D. Nychka (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics 15*(3), 502–523.

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

Gneiting, T. (1998). Simple tests for the validity of correlation function models on the circle. *Statistics and Probability Letters 39*(2), 119–122.

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

Gneiting, T., W. Kleiber, and M. Schlather (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association 105*(491), 1167–1177.

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

Guttorp, P. and T. Gneiting (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., R. Ruedy, J. Glascoe, and M. Sato (1999). GISS analysis of surface temperature change. *Journal of Geophysical Research 104*, 30997–31022.

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

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

Heine, V. (1955). Models for two-dimensional stationary stochastic processes. *Biometrika 42*(1), 170–178.

Henderson, R., S. Shimakura, and D. Gorst (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., J. Swall, and J. Kern (1999). Non-stationary spatial modeling. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics, 6*, pp. 761–768. New York: Oxford University Press.

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

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

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

Ilić, M., I. W. Turner, and V. Anh (2008). A numerical solution using an adaptively preconditioned Lanczos method for a class of linear systems related with the fractional Poisson equation. *Journal of Applied Mathematics and Stochastic Analysis 2008*(Article ID 104525).

Jones, R. H. (1963). Stochastic processes on a sphere. *Annals of Mathematical Statistics 34*(1), 213–218.

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

Karypis, G. and V. Kumar (1999). A fast and highly quality multilevel scheme for partitioning irregular graphs. *SIAM Journal of Scientific Computing 20*(1), 359–392.

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

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

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

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*. London: Chapman and Hall.

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

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

Pettitt, A. N., I. S. Weir, and A. G. Hart (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 A. Valli (2008). *Numerical Approximation of Partial Differential Equations* (2nd ed.). Springer.

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

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

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

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

Sampson, P. D. and P. Guttorp (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., M. Fuentes, and S. Gosh (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*. New York: Springer-Verlag.

Stein, M. L., Z. Chi, and L. J. Welty (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 A. N. Pettitt (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.

Yue, Y. and P. Speckman (2010). Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics 19*(1), 96–116.

Authors response, F. Lindgren, H. Rue, J. Lindström

We are delighted by the deluge of insightful comments, the details of which we can only begin to answer here. We have grouped our responses into a few common themes, mentioning commenters' names only when referring to specific issues.

Triangulations

As pointed out by Cooley, Hoeting, and Brown, it is not necessary to place the triangulation vertices at observation points. Indeed, the *observation matrix* in the global temperature example Section 4.2 was introduced for this very reason. For a given triangulation, the matrix can be used to extract any observable linear combination of field values, allowing observations in arbitrary locations as well as regional averages. For point observations, each row of the matrix contains three non-zero elements, one for each corner of the triangle containing the point, and the sparsity structure of the posterior precision matrix is unaffected. There is also no requirement to use a regular grid for such models. The triangulation implementation in R-INLA has a parameter for a minimum required distance between data-located vertices. In the temperature example this was set to 10 km, allowing the vertex placement to follow the data density, without generating excessively small triangles. An example where the triangulation is chosen completely independently of the data locations is given by Bolin and Lindgren (2011). Crujeiras and Prieto add that the triangulation could be chosen adaptively based on local approximation error estimates, and we agree that this can potentially be useful for non-stationary anisotropic models.

Approximation properties

Ippoliti, Martin and Bhansali note that the SPDE/GMRF precision coefficients do not match those obtained by inverting the covariance matrix of a field sampled on a regular grid. However, the approximation is not aiming to approximate the field only at the vertices of the triangulation, but also at the intermediate points obtained via linear interpolation. If the same interpolation method is used with the sampled covariances, the overall field covariances is underestimated, whereas the SPDE/GMRF approach gives an overall closer approximation. The upper and lower envelopes of all the pairwise covariances for the two settings are shown in Figure 1, together with the target covariance function. Bolin and Lindgren (2009) investigates how this effect influences the Kriging results, comparing with tapering, kernel convolutions, as well as alternative choices of basis functions in the GMRF construction, including wavelets and B-splines.

Kernel methods

Kernel convolution methods, as mentioned by Furrer, Furrer, and Nychka, are useful theoretical tools, but can in practice be cumbersome and computationally intensive (Bolin and Lindgren, 2009). The kernel generated by the SPDE operator for a Matérn field takes the shape of another Matérn covariance with different shape parameter. The kernels are singular for $\alpha \leq d$, and non-differentiable for $\alpha \leq d + 2$, so the commonly used discrete kernel sums result in Kriging and

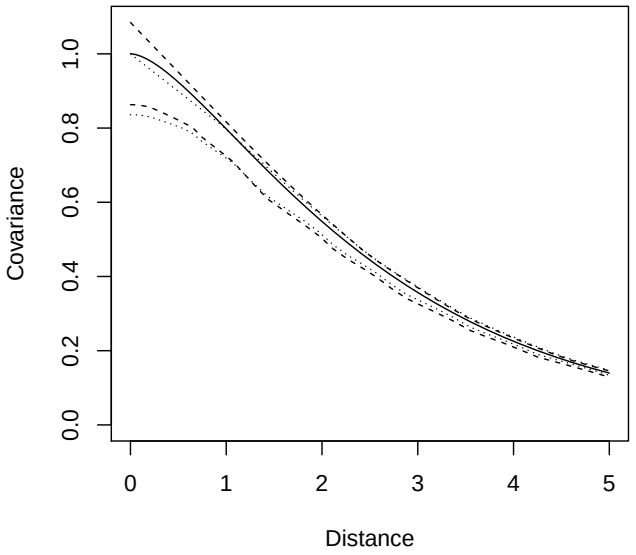


Figure 1: Target covariance (solid), linear interpolation covariance envelopes (dotted), and GMRF covariance envelopes (dashed). The envelopes show the minimum and maximum covariances for each pairwise point distance within the domain. The linear interpolation is accurate at grid nodes, but the GMRF approximation is closer to the target overall

parameter estimation artefacts, and do not yield either correct pointwise distribution or correct distributions for regional averages, unless the range is large compared to the kernel placement distances. Also, since the kernels have non-compact support, they yield dense matrices for the posterior precisions. Using compactly supported kernels as suggested by Mateu is similar to moving average processes in time-series analysis, and problematic unless the data are densely and evenly located on the domain of interest, whereas the GMRF models are counterparts to auto-regressive models, that are much more flexible tools for approximating general dependence structures.

Parameter estimation

Both when estimating parameters and calculating Kriging interpolations, the results are influenced by the choice of boundary conditions. The easiest method for avoiding these effects is to extend the triangulation beyond the study region by an amount large enough to cover the correlation range of the field, since this allows the boundary effects to drop off to virtually zero before reaching the data locations that influence the likelihood. As seen in Figure 2 (left), this eliminates the bias in the maximum likelihood estimates of the field variances σ^2. The bias for the rescaling parameter τ^{-2} is reduced when the triangulation resolution is increased, as seen in Figure 2 (right). The parameter $\sigma^2\kappa^{2\nu}$

used in the comparison by Lee and Kaufman is equivalent to τ^{-2} in the SPDE models used in the paper. We believe that the highly variable results in their comparison is explained by noting that the precision matrix was chosen by simply deleting rows and columns, which is equivalent to conditioning, in this case leading to approximate Dirichlet boundary conditions. In order to compensate for the resulting small variances near the boundary of such a model, the variance parameter needs to be greatly overestimated. The combined comparisons show that Neumann boundaries are safer, and that extending the boundary further reduces the bias.

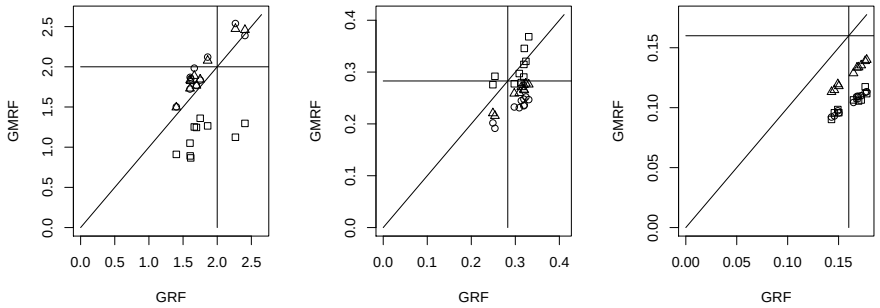


Figure 2: From left to right, estimated σ^2, κ, and $\tau^{-2} \propto \sigma^2\kappa^2$ for a $h = 1$ lattice GMRF (squares), a $h = 1$ with irregular extension (circles) and a $h = 1/2$ lattice with irregular extension, all with Neumann boundary conditions, plotted against the corresponding estimates based on a sampled covariance matrix. The estimations were computed on 10 samples from an $\nu = 1$ stationary Matérn field, with exact observations at a 20 by 20 lattice.

Boundaries

In the paper, we used Neumann boundary conditions for simplicity and ease of characterisation of the properties of the resulting models. For intrinsic models, these conditions are too restrictive, and need to be relaxed to achieve the desired field properties. The key lies in how the boundary conditions affect the null-space of the differential operator. The $\mathbf{B}$ matrix used in Appendix C.3 relaxes the constraints on the null-space, leading to intrinsic models. Normally, the rank deficiency of the precision matrix is used to determine the order of intrinsicness. However, since only some of the eigenfunctions of the Laplacian can be represented exactly in the piecewise linear basis, the rank deficiency does not tell the whole story. For a more complete picture, the continuous domain problem needs to be analysed more carefully.

A common alternative to Neumann conditions is to let $\partial_{\mathbf{n}}^2 = 0$, which is easily accomplished for regular grids by replacing the 2D grid Laplacian with the 1D Laplacian along the boundaries. For a unit square domain, the resulting null-space is spanned by the four functions 1, u_1, u_2, and u_1u_2, but the rank deficiency is only 3, since the last function is not piecewise linear. While this gives approximately the standard polynomially intrinsic models on $\mathbb{R}^2$, the con-

struction also hints at a more general method for more general SPDE models currently under investigation. The idea is to start with a fully intrinsic precision for the interior of the domain and add appropriate penalties generated by SPDE models within the boundary. For 1D models, this eliminates the boundary effects entirely, and the higher dimension cases show promising results.

Model checking

We have not yet investigated the model checking issue mentioned by Gelman and Møller, since this is a general problem for spatial models, and not specific to the SPDE/GMRF approach. However, there appears to be opportunities for using the GMRF *increments* in similar ways to residual analysis for time-series models, and also using the close link to the continuous space SPDE formulation itself when interpreting the results.

Priors

Choosing priors for the model parameters is a general issue for spatial models, but the handling of the boundary in the SPDE formulation may present further complications. When the correlation range is longer than the size of the domain, estimating κ becomes very difficult. In such situations, the intrinsic models ($\kappa = 0$) can be used, reducing the set of parameters to an overall scaling factor. This also handles applications with only a single realisation of the random field, where it is impossible to separate long correlation range from a fixed spatial trend, and the posterior distribution for κ typically becomes degenerate, requiring a careful choice of prior. A heuristic approach when not using intrinsic models is to specify a prior for κ that puts low probability for range longer than the diameter of the domain. In the temperature example, we used independent Gaussian priors of that type for the weights for the basis functions controlling $\log \tau$ and $\log \kappa^2$. We are currently extending the temperature example into a full analysis, where the prior for all the SPDE parameters is constructed jointly, giving more control over the behaviour.

Simultaneous auto-regressions

Kent astutely noted the connection to simultaneous auto-regressions (SAR) that, for even integer values of α, provides another direct link between Markov models and SPDEs, and the GMRF construction in the paper also takes this form. Using the notation from Theorem 2 in Appendix C.2, for $\alpha = 2$ we have $\sum_j K_{ij} x_j \sim N(0, \widetilde{C}_{ii})$, where $\widetilde{C}$ is the diagonal matrix from Appendix C.5. In our early experimentation, we approached the GMRF construction problem by various attempts at modifying the graph Laplacian mentioned by Mondal. In hindsight, the current approach that builds more directly on the continuous domain Laplacian feels more natural to us when the goal is to build spatially consistent Markov models. The results do resemble the graph Laplacian, but as the SPDE models are extended to non-stationary models and fields on manifolds, the graph becomes less useful as such, and is purely a computational device. This becomes particularly clear when extending the methods to fractional SPDEs.

Fractional operators

Although the results as presented in the paper are limited to integer values for α in the SPDE generating Matérn field, the GMRF construction can be extended into a more general class of continuous domain Markov models, which contains close approximators of Matérn models with fractional α. The result from Rozanov (1977) mentioned in Section 2.3 means that a stationary random field is a Markov field if and only if the spectral density is the reciprocal of a polynomial. In the isotropic case, such spectra take the form

$$\widetilde{S}(\mathbf{k}) = (2\pi)^{-d} \left(\sum_{i=0}^{p} b_i \|\mathbf{k}\|^{2i} \right)^{-1}$$

where p is the degree of the polynomial and b_i are coefficients in a non-negative polynomial, and the corresponding discretised precision matrix can be obtained as

$$\mathbf{Q} = \sum_{i=0}^{p} b_i \mathbf{C} \left(\mathbf{C}^{-1} \mathbf{G} \right)^i .$$

We need to find coefficients b_i so that the model defined by $\widetilde{S}(\mathbf{k})$ is an appropriate approximation of a model defined by the Matérn spectrum $S(\mathbf{k}) = (2\pi)^{-d} \left(\kappa^2 + \|\mathbf{k}\|^2\right)^{-\alpha}$. A sensible choice is to let $p = \lceil \alpha \rceil$, and we use a convenient weighting function $w(\mathbf{k})$ for the deviation between the spectra, such that

$$\int_{\mathbb{R}^d} \left(\widetilde{S}(\mathbf{k}) - S(\mathbf{k}) \right)^2 w(\mathbf{k}) \, \mathrm{d}\mathbf{k} \propto \int_{\kappa^2}^{\infty} \left(z^\alpha - \sum_{i=0}^{p} b_i (z - \kappa^2)^i \right)^2 z^{-2p-1-\lambda} \mathrm{d}z.$$

Taking derivatives with respect to all b_i and evaluating the integrals, we obtain a linear system of equations that can be solved easily. For $\lceil \alpha \rceil = 1$ and $\lceil \alpha \rceil = 2$, the coefficients are given through

$$\begin{bmatrix} b_0 \\ b_1 \kappa^2 \end{bmatrix} = \kappa^{2\alpha} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2+\lambda} & \frac{1}{1+\lambda} \\ \frac{1}{1+\lambda} & \frac{1}{\lambda} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{2+\lambda-\alpha} \\ \frac{1}{1+\lambda-\alpha} \end{bmatrix},$$

$$\begin{bmatrix} b_0 \\ b_1 \kappa^2 \\ b_2 \kappa^4 \end{bmatrix} = \kappa^{2\alpha} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{4+\lambda} & \frac{1}{3+\lambda} & \frac{1}{2+\lambda} \\ \frac{1}{3+\lambda} & \frac{1}{2+\lambda} & \frac{1}{1+\lambda} \\ \frac{1}{2+\lambda} & \frac{1}{1+\lambda} & \frac{1}{\lambda} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{4+\lambda-\alpha} \\ \frac{1}{3+\lambda-\alpha} \\ \frac{1}{2+\lambda-\alpha} \end{bmatrix},$$

respectively.

The limiting case $\lambda \longrightarrow \infty$ is equivalent to Taylor approximation at $\mathbf{k} = \mathbf{0}$, and gives $(b_0, b_1) = \kappa^{2\alpha-2}(\kappa^2, \alpha)$ for $\lceil \alpha \rceil = 1$ and $(b_0, b_1, b_2) = \kappa^{2\alpha-4}(\kappa^4, \alpha\kappa^2, \alpha(\alpha-1)/2)$ for $\lceil \alpha \rceil = 2$. These limiting approximations provide good agreement for integrals of the field over regions, but for better behaviour of the short-distance covariances, λ needs to be chosen more carefully. For a given measure of deviation between the desired and approximate covariance functions, the optimal λ can be determined numerically, as a function of α. For fractional α between 0 and 2 on $\mathbb{R}^2$, the parsimonious choice $\lambda = \alpha - \lfloor \alpha \rfloor$ approximately minimises the maximal absolute difference between the covariance functions. As noted

by Cooley and Hoeting, α is in practice often chosen only from the integers and half-integers, and we obtain $(b_0, b_1) = \kappa^{-1}(3\kappa^2/4, 3/8)$ for $\alpha = 1/2$ and $(b_0, b_1, b_2) = \kappa^{-1}(15\kappa^4/16, 15\kappa^2/8, 15/128)$ for $\alpha = 3/2$. Combined with the recursive construction for $\alpha > 2$, this provides GMRF approximations for all positive integers and half-integers. This includes the exponential covariance in $\mathbb{R}^2$, which corresponds to $\alpha = 3/2$. The resulting covariance is shown in Figure 3, together with the covariance from the spectral Taylor approximation. Further investigations are needed to determine how well the measurement noise model can incorporate the resulting deviation in small scale variation introduced by the approximation. Also shown in the figure is the covariance for a model with $\alpha = 2$, showing the same qualitative behaviour at zero, but different mid-range behaviour.

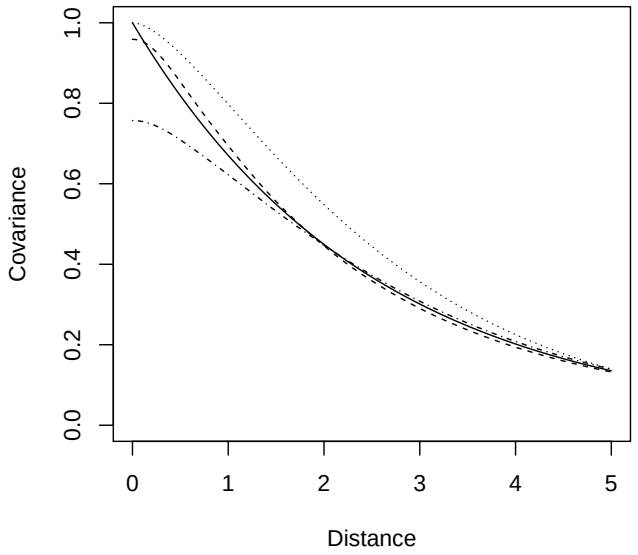


Figure 3: Covariances based spectral approximation: Desired exponential covariance ($d = 2$, $\alpha = 3/2$, range 5, solid line), Taylor approximation (dash-dotted), parsimonious weighting ($\lambda = 1/2$, dashed), and theoretical covariance for $\alpha = 2$ (dotted)

Long-range dependence

As discussed by Bhattacharya et al. (1983), apparent long-range dependence in data can not be distinguished from a non-stationary mean or trend. An alternative to constructing covariance functions with such behaviour is therefore to use a two-stage model, where the local behaviour is treated separately from the long-range behaviour. In practical situations, spatially varying basis functions are often used to capture large-scale variations, leaving the rest for a

spatial field component. This can easily be extended to allow the basis weights to differ between realisation of the field, in effect increasing the spectral density near zero. For identifiability reasons, intrinsic models can be preferable, but alternatives such as conditioning on zero integral for the field can also be used, and are implemented in R-INLA. As suggested by Fernhead, another even more general approach is to model the observed field as the sum of several latent fields with different range. Care has to be taken to handle the near non-identifiability of such models, noting that each individual latent component may be of less interest than their sum.

In some cases, these approaches can be motivated by considering the physical interpretation of the observed system, where long-range dependence may appear due to an unobserved latent physical process, e.g. deep-water processes with long-term memory affecting surface-water processes that interact more rapidly with external forces. The rational spectra generated by the nested SPDE approach due to Bolin and Lindgren (2011) allows more general models even with just a single latent process. While not Markov as such, they are Markov on an augmented state space, leading to almost the same computational efficiency as the pure Markov models.

Deformations

When using the non-stationary SPDE reparameterisation of the deformation method, all distances are interpreted within a fixed topology, and the issue of folding is transformed into requiring strictly positive definite diffusion tensors. By parameterising with scalar and vector fields, the estimated parameter fields can be used to interpret and understand the non-stationarity. A simple example is shown by Ottavi and Simpson. Furthermore, this yields a larger practical class of models, since the parameters need no longer correspond to a simple deformation. When using a Matérn process prior in a traditional deformation model, Schmidt rightly points out that fixing $\alpha = 2$ for the deformation field would give undesirable foldings due to insufficient differentiability. Increasing α should alleviate the problem to the same extent as any other choice of more differentiable deformation field model would do. Similarly the parameters in the non-stationary SPDE models can be constructed via general Gaussian fields, but direct comparison to the more general deformation methods is difficult, since the models would need to be constructed on the embedding space, whereas the SPDE as used in this paper is defined on the manifold itself, regardless of any embedding.

General extensions

It is important to note that the GMRF models can be combined in hierarchical modelling frameworks to allow highly non-Gaussian observation processes. Of these, of particular note are the log-Gaussian Cox-processes, mentioned by Diggle, Illian, Simpson, Møller, and Höhle. The likelihood can be rewritten on a form that allows the use of the INLA method for inference, and as Diggle notes one can choose freely between gridded count data and using the actual point data itself.

Functional data can be treated either directly in the general observation model, or by incorporating desired basis functions into the finite element basis

itself, leading to block matrices in the precisions. In a setting with a local set of temporal basis functions $\{\xi_k^1, \ldots, \xi_k^m\}$ for each spatial triangulation vertex k, the resulting elements of the joint $\mathbf{K}$-matrix take the form

$$\left\langle \psi_k \xi_k^i, \mathcal{L}\psi_{k'} \xi_{k'}^{i'} \right\rangle_{\Omega \times \mathbb{R}} = \langle \psi_k, \mathcal{L}\psi_{k'} \rangle_\Omega \left\langle \xi_k^i, \xi_{k'}^{i'} \right\rangle_{\mathbb{R}}$$

for a given spatial differential operator $\mathcal{L}$, and similarly for the other matrices in the precision construction.

For more general spatio-temporal SPDE models, we agree with Crujeiras and Prieto that a finite volume approach is preferable to finite elements, and Fuglstad presents an example of such a solution. The diagonal approximation to the $\mathbf{C}$ matrix is precisely what a simple finite volume method would produce in the purely spatial case, lending further weight to the appropriateness of the approximation.

References

Bhattacharya, R. N., V. K. Gupta, and E. Waymire (1983). The Hurst effect under trends. *Journal of Applied Probability 20*(3), pp. 649–662.

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

Bolin, D. and F. Lindgren (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Annals of Applied Statistics 5*(1), 523–550.

Rozanov, J. A. (1977). Markov random fields and stochastic partial differential equations. *Mathematics of the USSR-Sbornik 32*(4), 515.