J. R. Statist. Soc. B (2011)
73, *Part* 4, *pp.*

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

Finn Lindgren and Håvard Rue

Norwegian University of Science and Technology, Trondheim, Norway

and Johan Lindström

Lund University, Sweden

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, March 16th, 2011*, Professor D. M. Titterington *in the Chair*]

Summary. Continuously indexed Gaussian fields (GFs) are 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 factorizing dense matrices is cubic in the dimension. Although 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 precision matrix involved 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 parameterization. We show that, using an approximate stochastic weak solution to (linear) stochastic partial differential equations, 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 by using GFs but do the computations by using GMRFs. Perhaps more importantly, our approach generalizes 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 fields; Gaussian Markov random 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 are one of a few appropriate multivariate models with an explicit and computable normalizing constant and have good analytic properties otherwise. In a domain $\mathcal{D} \in \mathbb{R}^d$ with co-ordinate $\mathbf{s} \in \mathcal{D}$, $x(\mathbf{s})$ is a continuously indexed GF if all finite collections $\{x(\mathbf{s}_i)\}$ are jointly

Address for correspondence: Håvard Rue, Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.
E-mail: hrue@math.ntnu.no

 1369–7412/11/73000

Gaussian distributed. In most cases, the GF is specified by using a mean function $\mu(\cdot)$ and a covariance function $C(\cdot,\cdot)$, so the mean is $\boldsymbol{\mu} = (\mu(\mathbf{s}_i))$ and the covariance matrix is $\boldsymbol{\Sigma} = (C(\mathbf{s}_i, \mathbf{s}_j))$. 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 depends only 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 characterizes 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 factorize dense $n \times n$ (covariance) matrices. Although the computational power today is at an all time high, the tendency seems to be that the dimension n is always set, or we want to set it, a little 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), page 387), and the situation is informally referred to as 'the big n problem'.

There are several approaches to try 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 to 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) proposed to use an approximate likelihood constructed through a sequential representation and then to 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) applied covariance tapering to zero-out parts of the covariance matrix to gain a computational speed-up. 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), section 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), section 2.1, for a condensed review. A GMRF is a discretely indexed Gaussian field $\mathbf{x}$, where the full conditionals $\pi(x_i|\mathbf{x}_{-i})$, $i = 1, \ldots, n$, depend only 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 $\mathbf{Q}$ (the inverse covariance matrix) relates directly to the notion of neighbours; $Q_{ij} \neq 0 \Leftrightarrow i \in \partial j \cup j$; see, for example, Rue and Held (2005) section 2.2. Algorithms for Markov chain Monte Carlo sampling 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 Besag (1974, 1975). However, GMRFs also allow for fast direct numerical algorithms (Rue, 2001), as numerical factorization of the matrix $\mathbf{Q}$ can be done by 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 are of major importance in Bayesian inferential methods. This is further enhanced by the link to nested integrated Laplace approximations (Rue *et al.*, 2009), which allow fast and accurate Bayesian inference for latent GF models.

Although GMRFs have very good computational properties, there are reasons 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 parameterize the precision matrix of a GMRF to achieve 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 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 by 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 parameterization 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 they 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 optimization. Despite these 'proof-of-concept' results, several researchers 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 to be useful even for spatiotemporal models (Allcroft and Glasbey, 2003).

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

(a) Do the modelling by using a GF on a set of locations $\{\mathbf{s}_i\}$, to construct a discretized GF with covariance matrix $\mathbf{\Sigma}$.
(b) Find a GMRF with local neighbourhood and precision matrix $\mathbf{Q}$ that *represents* the GF in the best possible way, i.e. $\mathbf{Q}^{-1}$ is close to $\mathbf{\Sigma}$ in some norm. (We deliberately use the word 'represents' instead of approximates.)
(c) Do the computations using the GMRF representation by using numerical methods for sparse matrices.

Such an approach relies on several assumptions. First the GF must be of such a type that there is a GMRF with local neighbourhood that can represent it sufficiently accurately 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 speed-up compared with treating the GF directly.

The purpose of this paper is to demonstrate that these requirements can indeed be met for certain members of GFs 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 cover the most important and most used covariance model in spatial statistics; see Stein (1999), page 14., which concluded a detailed theoretical analysis with '*Use the Matérn model*'. The GMRF representation can be constructed explicitly by 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 dependences determined by a general triangulation of the domain.

Rather surprisingly, extending this basic result seems to open new doors and opportunities, and to provide quite simple answers to rather difficult modelling problems. In particular, we shall show how this approach extends to Matérn fields on manifolds, non-stationary fields and fields with oscillating covariance functions. Further, we shall discuss the link to the deformation method of 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 by using GFs and the computations by using the GMRF representation, still holds for these extensions as the GMRF representation is still available 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 the rest of this paper is as follows. In Section 2, we discuss the relationship between Matérn covariances and a specific SPDE, and we present the two main results for explicitly constructing the precision matrices for GMRFs based on this relationship. In Section 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 Section 4, and we conclude the main part of the paper with a discussion in Section 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 an SPDE. We shall 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. Matérn covariance model and its stochastic partial differential equation

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

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

Here, K_ν is the modified Bessel function of the 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 that are made by using such a model. However, ν is usually fixed since it is poorly identified in typical applications. A more natural interpretation of the scaling parameter κ is as a range parameter ρ; the Euclidean distance where $x(\mathbf{u})$ and $x(\mathbf{v})$ is almost independent. Lacking a simple relationship, we shall 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 various scientific fields (Guttorp and Gneiting, 2006), but the important relationship that we shall make use of is that a GF $x(\mathbf{u})$ with the Matérn covariance is a solution to the linear fractional SPDE

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

where $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo-differential operator that we shall define later in equation (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 shall name any solution to equation (2) *a Matérn field* in what follows. 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 shall return to this issue in Appendix C.3. Further, there is an implicit assumption of appropriate boundary conditions for the SPDE, as for $\alpha \geqslant 2$ the null space of the differential operator is non-trivial, containing, for example, the functions $\exp(\kappa \mathbf{e}^{\mathrm{T}}\mathbf{u})$, for all $\|\mathbf{e}\| = 1$. The Matérn fields are the only *stationary* solutions to the SPDE.

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

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

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

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

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 shall restrict ourselves to dimension $d = 2$ although our results are general.

2.2.1. Main result 1

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

$$\begin{aligned} \mathrm{E}(x_{ij}|\mathbf{x}_{-ij}) &= \frac{1}{a}(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}), \\ \mathrm{var}(x_{ij}|\mathbf{x}_{-ij}) &= 1/a \end{aligned} \tag{5}$$

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

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

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

equation 14)) is that

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

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

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

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

meaning that (some discretized 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 equation (2) by Gaussian noise with spectrum 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): $\nu = 1$,

$$\begin{array}{|ccc} & 1 & \\ -2a & & 2 \\ 4+a^2 & -2a & 1 \end{array}$$

$\nu = 2$,

$$\begin{array}{|cccc} & -1 & & \\ & 3a & & -3 \\ -3(a^2+3) & & 6a & -3 \\ a(a^2+12) & -3(a^2+3) & 3a & -1 \end{array}$$

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

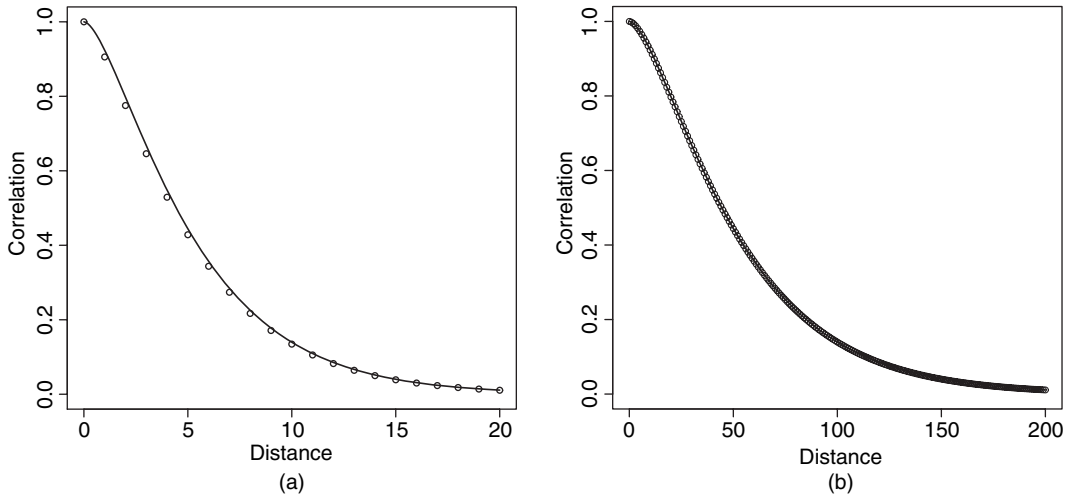

Fig. 1. Matérn correlations (———) for range (a) 10 and (b) 100, and the correlations for the GMRF representation (∘)

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 main result confirms the above heuristics.

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

Simple extensions of this result include anisotropy along the main axes, as presented in Appendix 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 generalization, which is 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 we do 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 is a standard problem in engineering for solving partial differential equations by using finite element methods (FEMs) (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 maximize the minimum interior triangle angle, so-called Delaunay triangulations, which helps to ensure that the transitions between small and large triangles are smooth. The extra vertices are added heuristically to try to minimize the total number of triangles that are needed to fulfil the size and shape constraints. See for example Edelsbrunner (2001), and 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 shall use an example from Henderson *et al.* (2002) which models spatial variation in leukaemia survival data in north-west England. Fig. 2(a) displays the locations of 1043 cases of acute myeloid leukaemia in adults who were diagnosed between 1982 and 1998 in north-west England. In the analysis, the spatial scale has been normalized so that the width of the study region is equal to 1. Fig. 2(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.

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

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

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

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

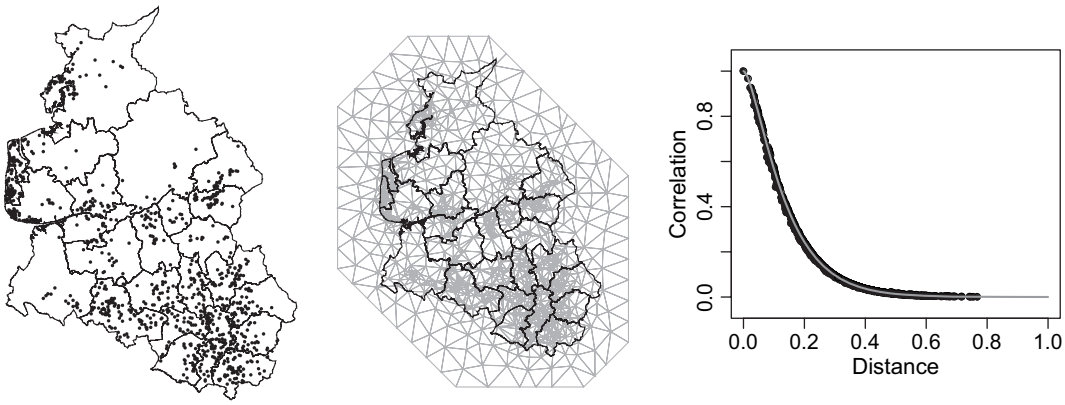

Fig. 2. (a) Locations of leukaemia survival observations, (b) triangulation using 3446 triangles and (c) a stationary correlation function (———) and the corresponding GMRF approximation (•) for $\nu = 1$ and approximate range 0.26

for every appropriate finite set of *test functions* $\{\phi_j(\mathbf{u}), j = 1, \ldots, m\}$, where '$\stackrel{\mathrm{d}}{=}$' 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

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

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 0 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 equation (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 $\alpha \geqslant 3$, we let $\alpha = 2$ on the left-hand side of equation (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 Appendix C for details.

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

$$\begin{aligned} C_{ij} &= \langle \psi_i \psi_j \rangle, \\ G_{ij} &= \langle \nabla \psi_i \nabla \psi_j \rangle, \\ (\mathbf{K}_{\kappa^2})_{ij} &= \kappa^2 C_{ij} + G_{ij}. \end{aligned}$$

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

Result 2. Let $\mathbf{Q}_{\alpha,\kappa^2}$ be the precision matrix for the Gaussian weights $\mathbf{w}$ as defined in equation (9) for $\alpha = 1, 2, \ldots$, as a function of κ^2. Then the finite dimensional representations of the solutions to equation (2) have precisions

$$\left.\begin{aligned} \mathbf{Q}_{1,\kappa^2} &= \mathbf{K}_{\kappa^2}, \\ \mathbf{Q}_{2,\kappa^2} &= \mathbf{K}_{\kappa^2}\mathbf{C}^{-1}\mathbf{K}_{\kappa^2}, \\ \mathbf{Q}_{\alpha,\kappa^2} &= \mathbf{K}_{\kappa^2}\mathbf{C}^{-1}\mathbf{Q}_{\alpha-2,\kappa^2}\mathbf{C}^{-1}\mathbf{K}_{\kappa^2}, \qquad \text{for } \alpha = 3, 4, \ldots. \end{aligned}\right\} \tag{10}$$

Some remarks concerning this result are as follows.

(a) The matrices $\mathbf{C}$ and $\mathbf{G}$ 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 Appendix A.

(b) The matrix $\mathbf{C}^{-1}$ is dense, which makes the precision matrix dense as well. In Appendix C.5, we show that $\mathbf{C}$ can be replaced by the diagonal matrix $\tilde{\mathbf{C}}$, where $\tilde{C}_{ii} = \langle \psi_i, 1 \rangle$, 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 $\mathcal{O}(n)$ for any triangulation.

(d) For the special case where all the vertices are points on a regular lattice, using a regular triangularization 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 by 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 equation (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 $0 \leqslant \alpha \leqslant 2$, 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 use only a subset of the possible test functions. However, for a given triangulation, it is the best possible approximation in the sense that is made explicit in Appendix 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 $\alpha = 2$,

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

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 and c is some constant. The Hilbert space scalar product and norm are defined in definition 2 in Appendix B, which also includes the values and the gradients of the field. The result holds for general $d \geqslant 1$, 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 can approximate the Matérn covariance, Fig. 2(c) displays the empirical correlation function (dots) and the theoretical function for $\nu = 1$ with approximate range 0.26, using the triangulation in Fig. 2(b). The match is quite good. Some dots show a discrepancy from the true correlations, but these can be identified to be due to the rather rough triangulation outside the area of interest which is 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 in 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 inflate the variance near the boundary (see Appendix A.4 for details) but other choices are also possible (see Rue and Held (2005), chapter 5).

2.4. Leukaemia example

We shall now return to the example from Henderson *et al.* (2002) at the beginning of Section 2.3 which models spatial variation in leukaemia survival data in north-west England. The specification, in (pseudo) Wilkinson–Rogers notation (McCullagh and Nelder (1989), section 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 hyperparameters 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 by 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 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

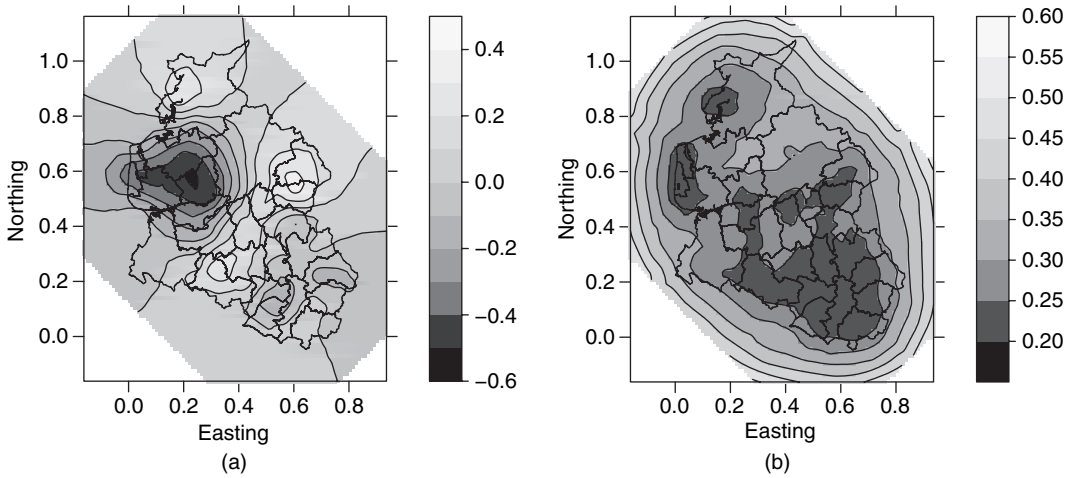

Fig. 3. (a) Posterior mean and (b) standard deviation of the spatial effect on survival by using the GMRF representation

took about 16 s on a quad-core laptop, and factorizing the 2797×2797 (total) precision matrix took about 0.016 s on average.

3. Extensions: beyond classical Matérn models

In this section we shall 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 GFs. The third extension is to study a complex version of equation (2) which makes it possible to construct oscillating fields. The fourth extension generalizes the non-stationary SPDE to a more general class of non-isotropic fields. Finally, the fifth extension shows how the SPDE generalizes to non-separable space–time models.

An important feature in our approach is that all these extensions still give explicit GMRF representations that are similar to expression (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 Section 4 we shall illustrate the use of these extensions with a non-stationary model for global temperatures.

3.1. Matérn fields on manifolds

We shall now move away from $\mathbb{R}^2$ and consider Matérn fields on manifolds. GFs on manifolds are 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 spatiotemporal models. To simplify the current discussion we shall 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 $\mathbb{R}^d$ cannot 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 GF $W(\cdot)$ 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 Appendices B–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

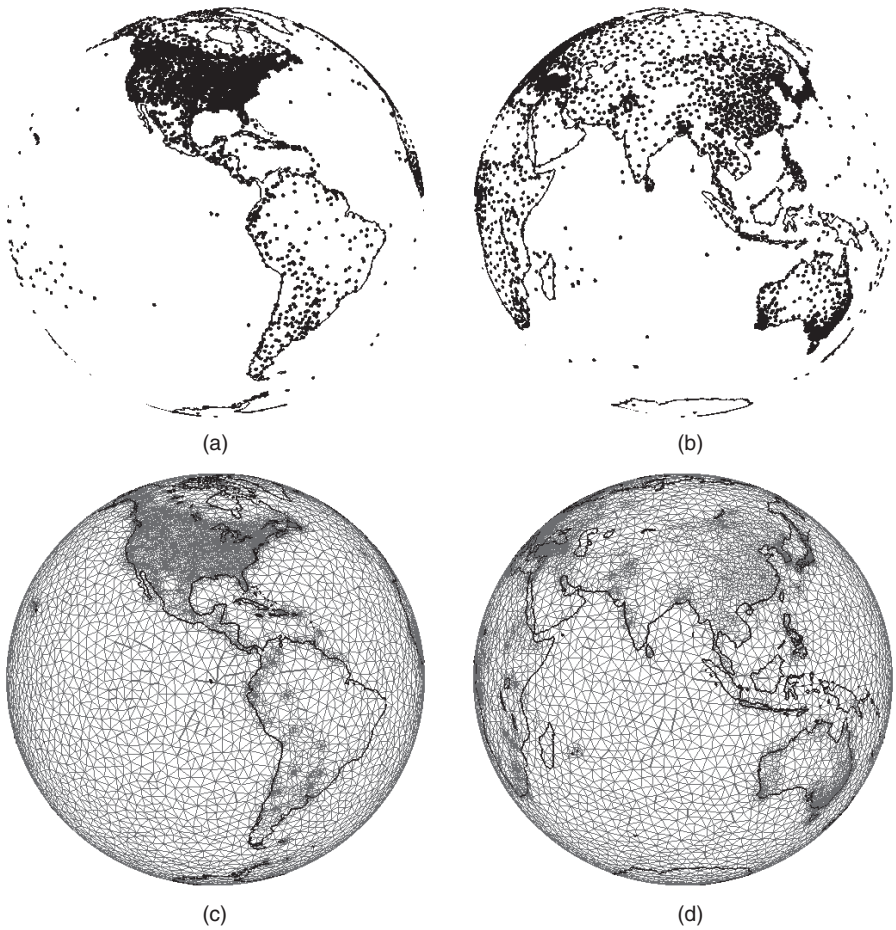

Fig. 4. (a), (b) Data locations and (c), (d) triangulation for the global temperature data set analysed in Section 4, with a coastline map superimposed

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 15 182 vertices and 30 360 triangles. The resulting GF 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 that are larger than the local edge length (0.08 or less), and nearly undetectable even for shorter distances. A random realization 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 divide the graph recursively into conditionally independent sets (Karypis and Kumar, 1998), which helps to make the Cholesky factor of the precision matrix sparse.

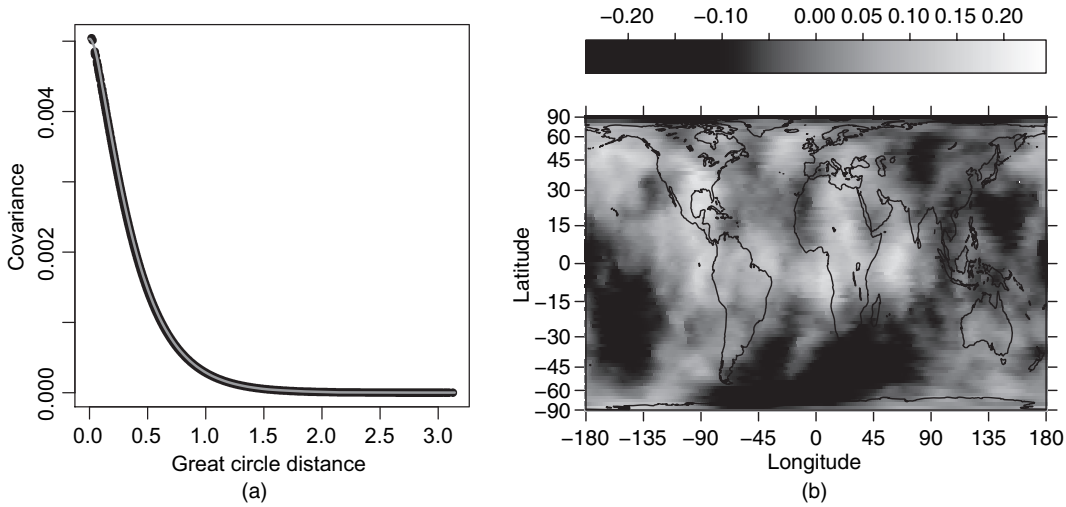

Fig. 5. (a) Covariances (•, numerical result for the GMRF approximation; ——, theoretical covariance function) and (b) a random sample from the stationary SPDE model (2) on the unit sphere, with $\nu = 1$ and $\kappa^2 = 16$

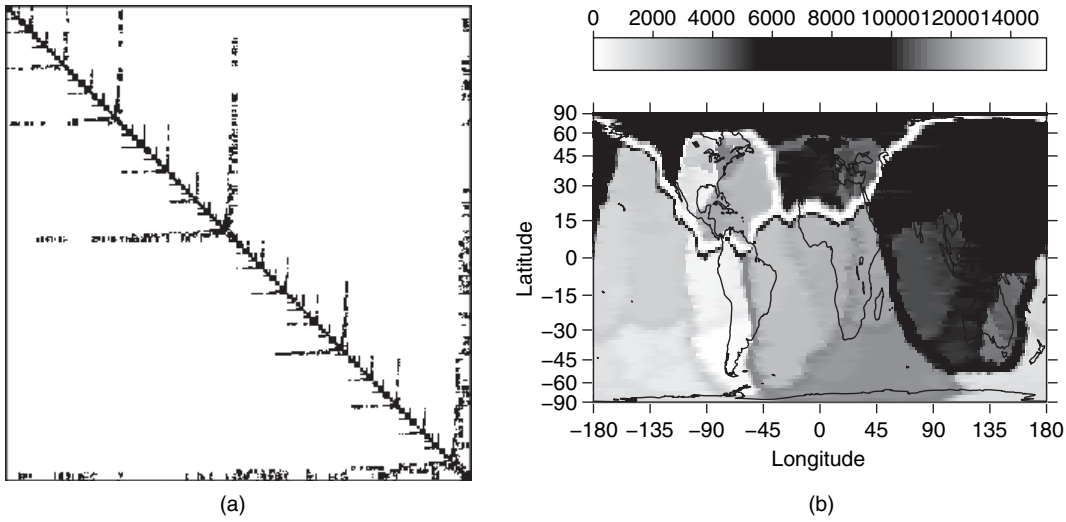

Fig. 6. (a) Structure of the (reordered) 15182 × 15182 precision matrix and (b) a visual representation of the reordering: the indices of each triangulation node have been mapped to grey scales showing the governing principle of the reordering algorithm, recursively dividing the graph into conditionally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

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 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) GF is a GMRF, which allows for swift computations and can additionally be defined on a manifold.

In the SPDE defined in equation (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 $\mathbf{u}$, and we write

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

For simplicity, we choose to keep the variance for the innovation constant and instead scale the resulting process $x(\mathbf{u})$ with a scaling parameter $\tau(\mathbf{u})$. Non-stationarity is achieved when one or both parameters are non-constant. Of particular interest is the case where they vary slowly with $\mathbf{u}$, e.g. in a low dimensional representation like

$$\log\{\kappa^2(\mathbf{u})\} = \sum_i \beta_i^{(\kappa^2)} B_i^{(\kappa^2)}(\mathbf{u})$$

and

$$\log\{\tau(\mathbf{u})\} = \sum_i \beta_i^{(\tau)} B_i^{(\tau)}(\mathbf{u})$$

where the basis functions $\{B_i^{(\cdot)}(\cdot)\}$ are smooth over the domain of interest. With slowly varying parameters $\kappa^2(\mathbf{u})$ and $\tau(\mathbf{u})$, the appealing *local* interpretation of equation (12) as a Matérn field remains unchanged, whereas the actual form of the non-stationary correlation function achieved 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 equation (12) is found by 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

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

for a naturally defined $\mathbf{u}_j^*$ in the support of ψ_i and ψ_j. The consequence is a simple scaling of the matrices in expression (10) at no additional cost; see Appendix A.3. If we improve the integral approximation (13) from considering $\kappa^2(\mathbf{u})$ locally constant to locally planar, the computational preprocessing cost increases but is still $\mathcal{O}(1)$ for each element in the precision matrix $\mathbf{Q}_\alpha$.

3.3. Oscillating covariance functions

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

$$\{\kappa^2 \exp \mathrm{i}\pi\theta - \Delta\}\{x_1(\mathbf{u}) + \mathrm{i}x_2(\mathbf{u})\} = \mathcal{W}_1(\mathbf{u}) + \mathrm{i}\mathcal{W}_2(\mathbf{u}), \qquad 0 \leqslant \theta < 1. \tag{14}$$

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

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

on $\mathbb{R}^d$. The corresponding covariance functions for $\mathbb{R}$ and $\mathbb{R}^2$ are given in Appendix A. For general manifolds, no closed form expression can be found. In 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 Appendix A. The details

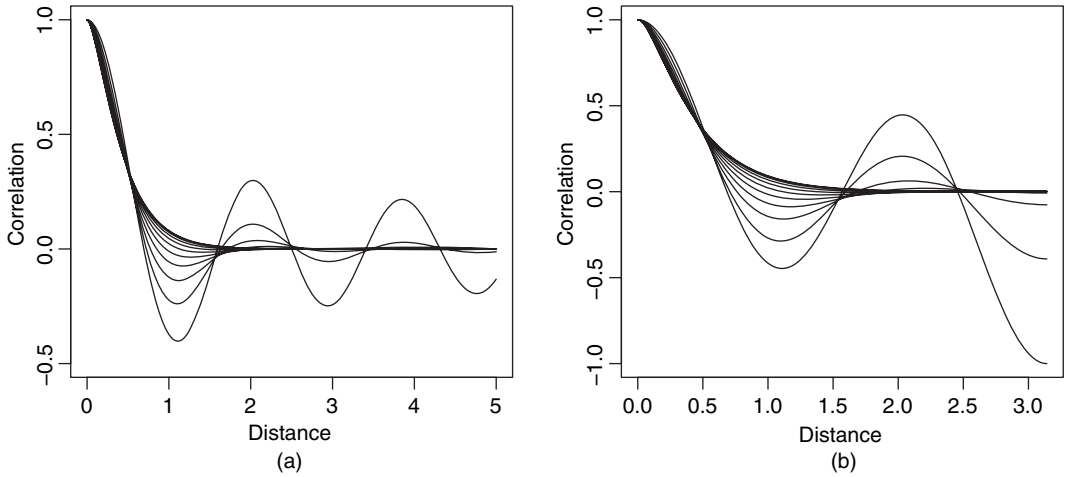

Fig. 7. Correlation functions from oscillating SPDE models, for $\theta = 0, 0.1, \ldots, 1$, on (b) $\mathbb{R}^2$ and (b) $\mathbb{S}^2$, with $\kappa^2 = 12$, $\nu = 1$

of the construction, which are given in Appendix 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 that was defined in Section 3.2, has locally isotropic correlations, despite having globally non-stationary correlations. This can be relaxed by widening the class of SPDEs considered, 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 that was 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 $\tilde{\Omega} \subseteq \mathbb{R}^d$, with $\mathbf{u} = f(\tilde{\mathbf{u}})$, $\mathbf{u} \in \Omega$, $\tilde{\mathbf{u}} \in \tilde{\Omega}$. Restricting to the case $\alpha = 2$, consider the stationary SPDE on the deformed space $\tilde{\Omega}$,

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

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

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

where $F(\mathbf{u})$ is the Jacobian of the deformation function f. This non-stationary SPDE exactly reproduces the deformation method with Matérn covariances (Sampson and Guttorp, 1992). A sparse GMRF approximation can be constructed by using the same principles as for the simpler non-stationary model in Section 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 parameterizing the model without explicit construction of a deformation function is to control the major axis of the local deformation given by $F(\mathbf{u})$ through a vector field, given either from covariate information or as a weighted sum of vector basis functions. Addition or subtraction of a directional derivative term further generalizes 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 Section 3.2. The model class can be interpreted as changes of metric in Riemannian manifolds, which is a natural generalization 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 characterized 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 dependence structures. Whereas it is difficult to construct non-separable non-stationary covariance functions explicitly, 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

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

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

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

which is strictly non-separable even with $\mathbf{m} = \mathbf{0}$ and $\mathbf{H} = \mathbf{I}$. The driving noise is an important part of the specification and may require an additional layer in the model. 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 $(\kappa^2 - \nabla\cdot\nabla)^{\alpha/2}\mathcal{E}(\mathbf{u},t) = \mathcal{W}(\mathbf{u},t)$, for some $\alpha \geqslant 1$, where $\mathcal{W}(\mathbf{u},t)$ is space–time white noise. A GMRF representation can be obtained by first applying the ordinary spatial method, and then discretizing the resulting system of coupled temporal stochastic differential equation 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 that were 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 data set (`http://www.ncdc.noaa.gov/ghcn/ghcn.html`) (Peterson and Vose, 1997) is commonly used. On August 8th, 2010, the data contained meteorological observations from 7280 stations spread across continents, where each of the 597 373 rows of observations contains the monthly mean temperatures from a specific station and year. The data span the period

1702–2010, 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 before 1880 up to 3700 in the 1970s. For each station, covariate information such as location, elevation and land use is available.

The Global Historical Climatology Network data are 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 (which is a procedure knows as *homogenization*), and the information about the temperature anomaly (the difference in weather from 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 by 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 shall 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 spatiotemporal 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 shall here restrict ourselves to a rotationally invariant covariance model, which reduces the computational burden. However, we shall allow for regional deviations from rotational symmetry in the expectations. The model separates weather from climate by assuming that the climate can be parameterized by non-stationary expectation and covariance parameters $\mu(\mathbf{u})$, $\kappa(\mathbf{u})$ and $\tau(\mathbf{u})$, for $\mathbf{u} \in \mathbb{S}^2$, and assuming that the yearly weather follows the model defined by equaiton (12), given the climate. Using the triangulation from Fig. 4 with piecewise linear basis function, the GMRF representation that is given in Appendix A.3 will be used, with $\mathbf{x}_t$ denoting the discretized 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 homogenization procedure into the model would go far beyond the scope of this paper, and we therefore use the 'adjusted' Global Historical Climatology Network 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 $\boldsymbol{\theta} = \{\boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_\tau, \boldsymbol{\theta}_s, \theta_\varepsilon\}$, and we denote the yearly temperature fields $\mathbf{x} = \{\mathbf{x}_t\}$ and the yearly observations $\mathbf{y} = \{\mathbf{y}_t\}$, with $t = 1970, \ldots, 1989$. Using basis function matrices $\mathbf{B}_\mu$ (all 49 spherical harmonics up to and including order 6; see Wahba, (1981)), $\mathbf{B}_\kappa$ and $\mathbf{B}_\tau$ (B-splines of order 2 in sin(latitude), shown in Fig. 8), the expectation field is given by $\boldsymbol{\mu}_{\mathbf{x}|\boldsymbol{\theta}} = \mathbf{B}_\mu \boldsymbol{\theta}_\mu$, the local spatial dependence $\kappa(\mathbf{u})$ is defined through $\log \boldsymbol{\kappa}^2 = \mathbf{B}_\kappa \boldsymbol{\theta}_\kappa$ and the local variance scaling $\tau(\mathbf{u})$ is defined through $\log(\boldsymbol{\tau}) = \mathbf{B}_\tau \boldsymbol{\theta}_\tau$. The prior distribution for the climate field is chosen as approximate solutions to the SPDE $\Delta\mu(\mathbf{u}) = \sigma_\mu \mathcal{W}(\mathbf{u})$, where $\sigma_\mu \gg 0$, which provides natural relative prior weights for the spherical harmonic basis functions.

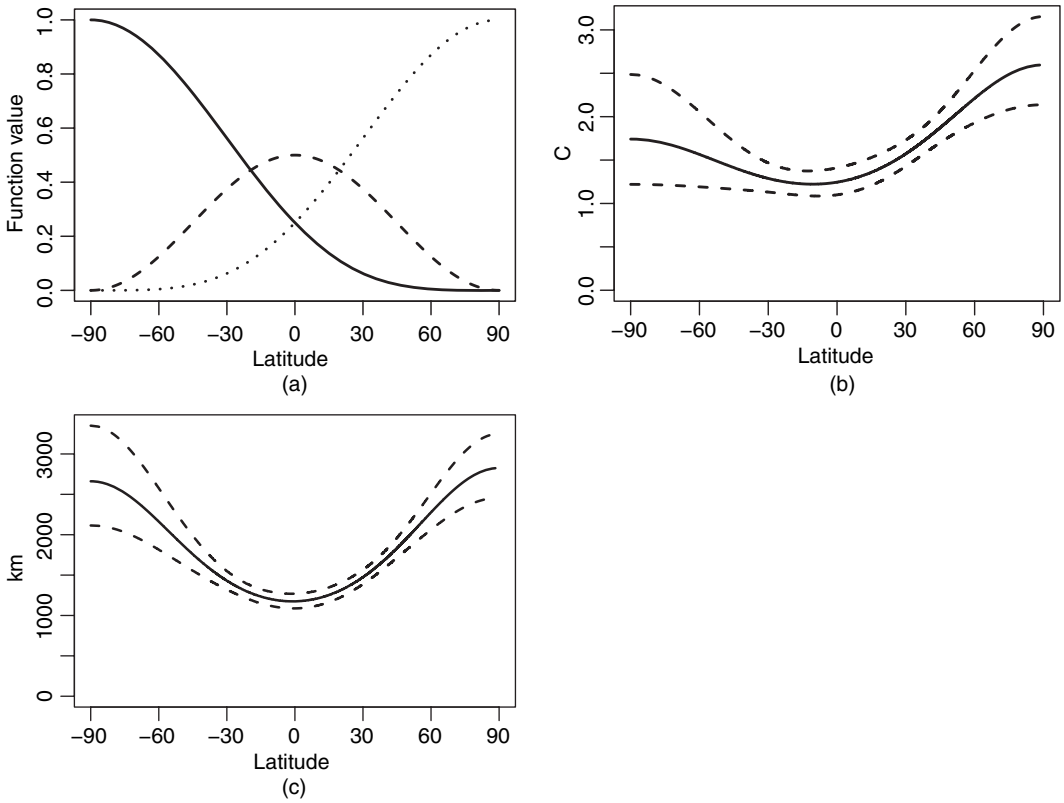

Fig. 8. (a) Three transformed *B*-spline basis functions of order 2, and approximate 95% credible intervals for (b) standard deviation and (c) approximate correlation range of the yearly weather, as functions of latitude

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

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

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

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

where $\mathbf{S}_t\boldsymbol{\theta}_s$ are station-specific effects and $\mathbf{Q}_{\mathbf{y}|\mathbf{x},\boldsymbol{\theta}} = \mathbf{I}\exp(\theta_\varepsilon)$ is the observation precision. Since we use the data only for illustrative purposes here, we shall ignore all station-specific effects except for elevation. We also ignore any remaining residual dependences 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 $(\boldsymbol{\theta}|\mathbf{y})$ and $(\mathbf{x}|\mathbf{y})$, so all uncertainty about the weather $\mathbf{x}_t$ is included in the distribution for the model parameters $\boldsymbol{\theta}$, and conversely for $\boldsymbol{\theta}$ and $\mathbf{x}_t$. One of the most important steps is how to determine the conditional distribution for the weather given observations and model parameters,

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

where $\mathbf{Q}_{\mathbf{x}|\mathbf{y},\theta} = \mathbf{Q}_{\mathbf{x}|\theta} + \mathbf{A}_t^{\mathrm{T}}\mathbf{Q}_{\mathbf{y}|\mathbf{x},\theta}\mathbf{A}_t$ is the conditional precision, and the expectation is the kriging estimator of $\mathbf{x}_t$. Owing to the compact support of the basis functions, which determined by the triangulation, each observation depends on at most three neighbouring nodes in $\mathbf{x}_t$, which makes the conditional precision have the same sparsity structure as the field precisions $\mathbf{Q}_{\mathbf{x}|\theta}$. 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 by using `R-inla`. Since $(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$ is Gaussian, the results are only approximate with regard to the numerical integration of the covariance parameters $(\boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_\tau, \boldsymbol{\theta}_\varepsilon)$. Owing to the large size of the data set, this initial analysis is based on data only from the period 1970–1989, 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 1 h to compute on a 12-core computer, with a peak memory use of about 50Gbytes 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 Section 3.4 took more than a week on a supercomputer.

The 95% credible interval for the measurement standard deviation, including local unmodelled effects, was calculated as $(0.628, 0.650)$ °C, with posterior expectation 0.634 °C. The spatial covariance parameters are more difficult to interpret individually, but we instead show the resulting spatially varying field 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 that were used for the climate model are not of sufficiently high 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 1970–1989, in effect using the period average as reference. Thus, instead of evaluating the distributions of $(\boldsymbol{\mu}|\mathbf{y})$ and $(\mathbf{x}_t - \boldsymbol{\mu}|\mathbf{y})$, we instead consider $(\bar{\mathbf{x}}|\mathbf{y})$ and $(\mathbf{x}_t - \bar{\mathbf{x}}|\mathbf{y})$, where $\bar{\mathbf{x}} = \Sigma_{t=1970}^{1989}\mathbf{x}_t/20$. In Figs 9(a) and 9(b), the posterior expectation of the empirical climate, $\mathrm{E}(\bar{\mathbf{x}}|\mathbf{y})$, is shown (including the estimated effect of elevation), together with the posterior expectation of the temperature anomaly for 1980, $\mathrm{E}(\mathbf{x}_{1980} - \bar{\mathbf{x}}|\mathbf{y})$. The corresponding standard deviations are shown in Figs 9(c) and 9(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

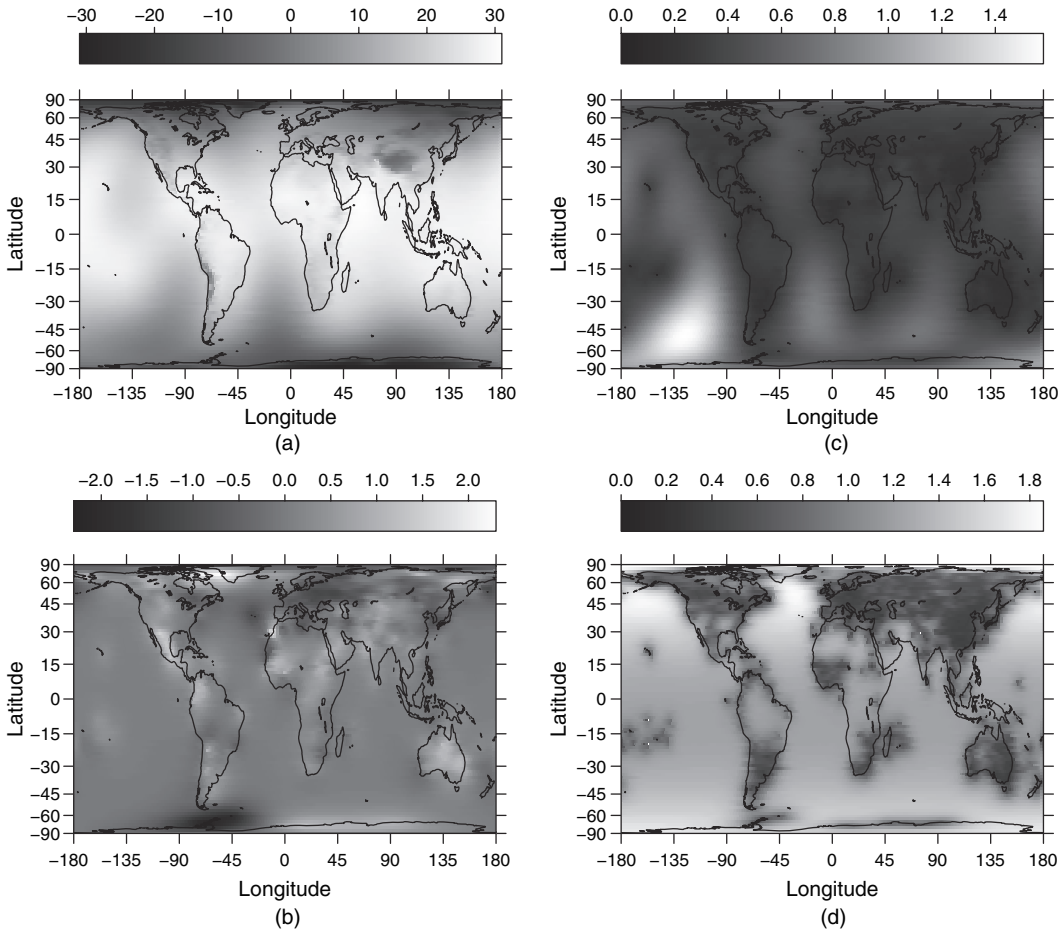

Fig. 9. Posterior means for (a) the empirical 1970–1989 climate and (b) the empirical mean anomaly 1980 with (c) and (d) the corresponding posterior standard deviations respectively: the climate includes the estimated effect of elevation; an area preserving cylindrical projection is used

estimated cooling effect of 5.2 °C per kilometre of increased elevation. It is clear from Figs 9(c) and 9(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, we might expect that the period of analysis and data coverage are too restricted to allow detection of global trends, especially since the simple model that 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 20-year period analysed. 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 that was 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 distinguish between these two cases by using only statistical methods on the single realization of the actual climate and weather system that is available, a full analysis should incorporate knowledge from climate system physics to balance properly 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) GFs and GMRFs by using an approximate weak solution of the corresponding SPDE. 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 GFs much more practically applicable, as we might model and interpret the model by using covariance functions while doing the computations by using the GMRF representation which allows for sparse matrix numerical linear algebra. In most cases, we can make use of the integrated nested Laplace approximation 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 in bridging the literature of (continuously indexed) GFs and geostatistics on one side, and GMRFs or 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 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, which has not yet been discussed, is that the SPDE approach might help to interpret external covariates (e.g. wind speed) as an appropriate drift term or similar in the related SPDE and then this covariate would enter the spatial dependence models correctly. 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.

Acknowledgements

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 that are commonly used in spatial statistics.

The authors thank the Research Section Committee and 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) and the key reference in Appendix C.5 and for invaluable discussions and comments, and to Georg Lindgren for numerous comments on the manuscript.

Appendix A: Explicit results

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

A.1. Regular lattices

Here we shall give some explicit precision expressions for grid-based models on $\mathbb{R}$ and $\mathbb{R}^2$. Consider the SPDE

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

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

For any given ordered discretization $u_1, \ldots, u_n$ 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 $\mathbf{H} = H \geqslant 0$, and the elements on row i, around the diagonal, of the precision are given by

$$\mathbf{Q}_1 : s_i \cdot [\, -a_i \quad c_i \quad -b_i \,],$$
$$\mathbf{Q}_2 : s_i \cdot [\, a_i a_{i-1} \quad -a_i(c_{i-1} + c_i) \quad a_i b_{i-1} + c_i^2 + b_i a_{i+1} \quad -b_i(c_i + c_{i+1}) \quad b_i b_{i+1} \,]$$

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 generalization of Lindgren and Rue (2008).

For $\mathbb{R}^2$, assume a given regular grid discretization, with horizontal (co-ordinate component 1) distances γ and vertical (co-ordinate 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

$$\mathbf{Q}_1 : \mathbf{s} \cdot \begin{vmatrix} -b & \\ c & -a \end{vmatrix}$$

$$\mathbf{Q}_2 : \mathbf{s} \cdot \begin{vmatrix} b^2 & & \\ -2bc & & 2ab \\ 2a^2 + 2b^2 + c^2 & -2ac & a^2 \end{vmatrix}$$

$$\mathbf{Q}_3 : \mathbf{s} \cdot \begin{vmatrix} -b^3 & & & \\ 3b^2 c & & -3ab^2 & \\ -3b(2a^2 + b^2 + c^2) & 6abc & & -3a^2 b \\ c(6a^2 + 6b^2 + c^2) & -3a(a^2 + 2b^2 + c^2) & 3a^2 c & -a^3 \end{vmatrix}$$

If the grid distances are proportional to the square root of the corresponding diagonal elements of $\mathbf{H}$ (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 Appendix C, we need to calculate

$$\left.\begin{aligned} \tilde{C}_{ij} &= \langle \psi_i, 1 \rangle_\Omega, \\ C_{ij} &= \langle \psi_i, \psi_j \rangle_\Omega, \\ G_{ij} &= \langle \nabla\psi_i, \nabla\psi_j \rangle_\Omega, \\ B_{ij} &= \langle \psi_i, \partial_{\mathbf{n}}\psi_j \rangle_{\partial\Omega}. \end{aligned}\right\} \tag{19}$$

For 2-manifolds such as regions in $\mathbb{R}^2$ or on $\mathbb{S}^2$, we require a triangulation with a set of vertices $\mathbf{v}_1, \ldots, \mathbf{v}_n$, embedded in $\mathbb{R}^3$. Each vertex $\mathbf{v}_k$ is assigned a continuous piecewise linear basis function ψ_k with support on the triangles attached to $\mathbf{v}_k$. To obtain explicit expressions for equation (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 $T = (\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2)$. The edge vectors opposite each corner are

$$\begin{aligned}\mathbf{e}_0 &= \mathbf{v}_2 - \mathbf{v}_1,\\ \mathbf{e}_1 &= \mathbf{v}_0 - \mathbf{v}_2,\\ \mathbf{e}_2 &= \mathbf{v}_1 - \mathbf{v}_0,\end{aligned}$$

and the corner angles are θ_0, θ_1 and θ_2.

The triangle area $|T|$ can be obtained from the formula $|T| = \|\mathbf{e}_0 \times \mathbf{e}_1\|/2$, i.e. half the length of the vector product in $\mathbb{R}^3$. The contributions from the triangle to the $\tilde{\mathbf{C}}$ and $\mathbf{C}$ matrices are given by

$$\begin{aligned}[\tilde{C}_{i,i}(T)]_{i=0,1,2} &= \frac{|T|}{3}(1 \quad 1 \quad 1),\\ [C_{i,j}(T)]_{i,j=0,1,2} &= \frac{|T|}{12}\begin{pmatrix}2 & 1 & 1\\ 1 & 2 & 1\\ 1 & 1 & 2\end{pmatrix}.\end{aligned}$$

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

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

and the entire contribution from the triangle is

$$[G_{i,j}(T)]_{i,j=0,1,2} = \frac{1}{4|T|}\begin{pmatrix}\|\mathbf{e}_0\|^2 & \mathbf{e}_0^{\mathrm{T}}\mathbf{e}_1 & \mathbf{e}_0^{\mathrm{T}}\mathbf{e}_2\\ \mathbf{e}_1^{\mathrm{T}}\mathbf{e}_0 & \|\mathbf{e}_1\|^2 & \mathbf{e}_1^{\mathrm{T}}\mathbf{e}_2\\ \mathbf{e}_2^{\mathrm{T}}\mathbf{e}_0 & \mathbf{e}_2^{\mathrm{T}}\mathbf{e}_1 & \|\mathbf{e}_2\|^2\end{pmatrix} = \frac{1}{4|T|}(\mathbf{e}_0 \quad \mathbf{e}_1 \quad \mathbf{e}_2)^{\mathrm{T}}(\mathbf{e}_0 \quad \mathbf{e}_1 \quad \mathbf{e}_2).$$

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

$$[B_{i,j}(T)]_{i,j=0,1,2} = \frac{-1}{4|T|}\begin{pmatrix}\mathbf{0} & \mathbf{e}_0 & \mathbf{e}_0\\ \mathbf{e}_1 & \mathbf{0} & \mathbf{e}_1\\ \mathbf{e}_2 & \mathbf{e}_2 & \mathbf{0}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}b_0\mathbf{I}\\ b_1\mathbf{I}\\ b_2\mathbf{I}\end{pmatrix}(\mathbf{e}_0 \quad \mathbf{e}_1 \quad \mathbf{e}_2),$$

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 $\tilde{\mathbf{C}}$, $\mathbf{C}$, $\mathbf{G}$, and $\mathbf{B}$-matrices.

For the anisotropic version, parameterized as in Appendix A.1 and Appendix C.4, the modified $\mathbf{G}$-matrix elements are given by

$$[G_{i,j}(T)]_{i,j=0,1,2} = \frac{1}{4|T|}(\mathbf{e}_0 \quad \mathbf{e}_1 \quad \mathbf{e}_2)^{\mathrm{T}}\mathrm{adj}(\mathbf{H})(\mathbf{e}_0 \quad \mathbf{e}_1 \quad \mathbf{e}_2), \tag{20}$$

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

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

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

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

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

where $\boldsymbol{\kappa}^2$ and $\boldsymbol{\tau}$ are diagonal matrices, with $\kappa_{ii}^2 = \kappa(\mathbf{u}_i)^2$ and $\tau_{ii} = \tau(\mathbf{u}_i)$. As shown in Appendix C.5, all the $\mathbf{C}$ should be replaced by $\tilde{\mathbf{C}}$ 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

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

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

$$\operatorname{cov}\{x(u),x(v)\} = \sum_{k=-\infty}^{\infty}\{r_M(u,v-2kL)+r_M(u,2kL-v)\}$$

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

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

$$\operatorname{cov}\{x(u),x(v)\} \approx r_{\mathrm{M}}(u,v)+r_{\mathrm{M}}(u,-v)+r_{\mathrm{M}}(u,2L-v) \tag{24}$$

$$= r_{\mathrm{M}}(0,v-u)+r_{\mathrm{M}}(0,v+u)+r_{\mathrm{M}}\{0,2L-(v+u)\}. \tag{25}$$

Moreover, the resulting covariance is nearly indistinguishable from the stationary Matérn covariance at distances greater than twice the range away from 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

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

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

$$r(\mathbf{u},\mathbf{v}) = \frac{1}{4\pi\sin(\pi\theta)\kappa^2 i}[K_0\{\kappa\|\mathbf{v}-\mathbf{u}\|\exp(-i\pi\theta/2)\}-K_0\{\kappa\|\mathbf{v}-\mathbf{u}\|\exp(i\pi\theta/2)\}] \tag{27}$$

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

Appendix B: Manifolds, random fields and operator identities

B.1. Manifold calculus

To state concisely the theory needed for constructing solutions to SPDEs 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 who are 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 co-ordinate-free manners. Here, we give a brief overview of the concepts that are used 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 consider only 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 that is embedded in $\mathbb{R}^3$. The d-dimensional hypercube $[0, 1]^d$ is a compact manifold with a closed boundary.

From the metric that is 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 $\mathbf{u}$ is a vector $\nabla\phi(\mathbf{u})$ defined indirectly via directional derivatives. In $\mathbb{R}^d$ with Euclidean metric, the gradient operator ∇ is formally given by the column vector $(\partial/\partial u_1, \ldots, \partial/\partial u_d)^{\mathrm{T}}$. The *Laplacian* Δ of ϕ at $\mathbf{u}$ (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 $\Delta\phi(\mathbf{u}) = \nabla \cdot \nabla\phi(\mathbf{u})$. In Euclidean metric on $\mathbb{R}^d$, we can write $\Delta = \partial^2/\partial u_1^2 + \ldots + \partial^2/\partial u_d^2$. At the boundary of Ω, the vector $\mathbf{n}_\partial(\mathbf{u})$ denotes the unit length outward normal vector at the point $\mathbf{u}$ on the boundary $\partial\Omega$. The *normal derivative* of a function ϕ is the directional derivative $\partial_{\mathbf{n}}\phi(\mathbf{u}) = \mathbf{n}_\partial(\mathbf{u}) \cdot \nabla\phi(\mathbf{u})$.

An alternative to defining integration on general manifolds through mapping subsets into $\mathbb{R}^d$ is to replace Lebesgue integration with integrals defined through normalized Hausdorff measures (Federer, 1951, 1978), here denoted $H^d_\Omega(\cdot)$. This leads to a natural generalization 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

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

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

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

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

Typical statements of the identity require $\phi \in C^1(\Omega)$ and $\psi \in C^2(\Omega)$, but we shall 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 (generalized fourier representation). The Fourier transform pair for functions $\{\phi \in L^2 : \mathbb{R}^d \mapsto \mathbb{R}\}$ is given by

$$\hat{\phi}(\mathbf{k}) = (\mathcal{F}\phi)(\mathbf{k}) = \frac{1}{(2\pi)^d} \langle \phi(\mathbf{u}), \exp(-i\mathbf{k}^{\mathrm{T}}\mathbf{u}) \rangle_{\mathbb{R}^d(\mathrm{d}u)},$$

$$\phi(\mathbf{u}) = (\mathcal{F}^{-1}\hat{\phi})(\mathbf{u}) = \langle \hat{\phi}(\mathbf{k}) \exp(i\mathbf{k}^{\mathrm{T}}\mathbf{u}) \rangle_{\mathbb{R}^d(\mathrm{d}\mathbf{k})}.$$

(Here, we briefly abuse 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 normalized eigenfunctions to the negated Laplacian, $-\Delta E_k = \lambda_k E_k$, can be chosen as basis, and the Fourier representation for a function $\phi \in L^2 : \Omega \mapsto \mathbb{R}$ is given by

$$\hat{\phi}(k) = (\mathcal{F}\phi)(k) = \langle \phi, E_k \rangle_\Omega,$$

$$\phi(\mathbf{u}) = (\mathcal{F}^{-1}\hat{\phi})(\mathbf{u}) = \sum_{k=0}^{\infty} \hat{\phi}(k)\, E_k(\mathbf{u}).$$

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

Definition 2. The Hilbert space $\mathcal{H}^1(\Omega, \kappa)$, for a given $\kappa \geqslant 0$, is the space of functions $\{\phi : \Omega \mapsto \mathbb{R}\}$ with $\nabla\phi \in L^2(\Omega)$, 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, which is 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 $\|\phi\|_{\mathcal{H}^1(\Omega,\kappa)}$ is a seminorm, 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. Generalized Gaussian random fields

We now turn to the problem of characterizing *random fields* on Ω. We restrict ourselves to GFs 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 GFs, the following definition can be used.

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

The complicating issue in dealing with the fractional SPDEs that are 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 generalized definition based on generalized functions.

Definition 4 (generalized function). For a given function space $\mathcal{F}$, an $\mathcal{F}$-generalized 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 that is 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 generalized functions, without any need to introduce heavy theoretical notational machinery, and provides a straightforward way of generalizing definition 3 to the kind of entities that we need for the subsequent analysis.

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

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

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

$$E(\langle \phi_i, \mathcal{W} \rangle_\Omega) = 0,$$
$$\operatorname{cov}(\langle \phi_i, \mathcal{W} \rangle_\Omega, \langle \phi_j, \mathcal{W} \rangle_\Omega) = \langle \phi_i, \phi_j \rangle_\Omega.$$

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 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 $E\{\mathcal{W}^*(A)\} = 0$ and $\operatorname{cov}\{\mathcal{W}^*(A), \mathcal{W}^*(B)\} = |A \cap B|_\Omega$. This definition is equivalent to that above (Adler and Taylor, 2007), and the Brownian sheet is a special case that considers only 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 generalized functions and random fields. Here, we present the two fundamental identities that are needed for the subsequent SPDE analysis; Green's first identity and a scalar product characterization of the half-Laplacian.

B.3.1. Stochastic Green's first identity

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

Lemma 1. If $\nabla f \in L^2(\Omega)$ and Δx is $L^2(\Omega)$ bounded, then (with probability 1)

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

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

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

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 $\tilde{f}$ and $\tilde{x}$, where q in each case is sufficiently large for the regular Green's identity to hold for $\tilde{f}$ and $\tilde{x}$. 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. Half-Laplacian

In defining and solving the SPDEs considered, the half-Laplacian operator needs to be characterized in a way that permits practical calculations on general manifolds. The fractional modified Laplacian operators $(\kappa^2 - \Delta)^{\alpha/2}$, $\kappa, \alpha \geqslant 0$, are commonly (Samko *et al.* (1992), page 483) defined through the Fourier transform, as defined above:

$$\{\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2}\phi\}(\mathbf{k}) = (\kappa^2 + \|\mathbf{k}\|^2)^{\alpha/2}(\mathcal{F}\phi)(\mathbf{k}),$$

on $\mathbb{R}^d$;

$$\{\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2}\phi\}(k) = (\kappa^2 + \lambda_k)^{\alpha/2}(\mathcal{F}\phi)(k),$$

on compact Ω, 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 generalized 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

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

whenever either

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

For a proof, see Appendix D.2. Lemma 2 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 generalizes in a natural way to random fields x with $L^2(\Omega)$-bounded ∇x, as well as to suitably well-behaved unbounded manifolds.

It would be tempting to eliminate the qualifiers in part (c) of lemma 2 by subtracting the average of the two boundary integrals to the relationship, and to extend lemma 2 to a complete equivalence relationship.

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 generalization necessarily yields the same half-Laplacian as the Fourier definition for functions that are not of the class $\Delta\phi \in L^2(\Omega)$. See Ilić *et al.* (2008) for a partial result.

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

C.1. Weak convergence and stochastic finite element methods

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

Definition 7 (dense subspace sequences). A finite subspace $\mathcal{H}_n^l(\Omega,\kappa) \subset \mathcal{H}^1(\Omega,\kappa)$ is spanned by a finite set of basis functions $\Psi_n = \{\psi_1,\ldots,\psi_n\}$. We say that a sequence of subspaces $\{\mathcal{H}_n^1\}$ is dense in $\mathcal{H}^1$ if for every $f \in \mathcal{H}^1$ there is a sequence $\{f_n\}$, $f_n \in \mathcal{H}_n^1$, 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 $\mathcal{H}_n^1$, we can choose the projection of $f \in \mathcal{H}^1$ onto $\mathcal{H}_n^1$, i.e. the f_n that minimizes $\|f - f_n\|_{\mathcal{H}^1}$. The error $f - f_n$ is orthogonal to $\mathcal{H}_n^1$, and the basis co-ordinates 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 $L^2(\Omega)$-bounded generalized GFs $\{x_n\}$ is said to converge *weakly* to an $L^2(\Omega)$-bounded generalized GF x if, for all $f, g \in L^2(\Omega)$,

$$E(\langle f, x_n\rangle_\Omega) \to E(\langle f, x\rangle_\Omega),$$

$$\operatorname{cov}(\langle f, x_n\rangle_\Omega, \langle g, x_n\rangle_\Omega) \to \operatorname{cov}(\langle f, x\rangle_\Omega, \langle g, x\rangle_\Omega),$$

as $n \to \infty$. We denote such convergence by

$$x_n \overset{D\{L^2(\Omega)\}}{\rightharpoonup} x.$$

Definition 9 (finite element approximations). Let $\mathcal{L}$ be a second-order elliptic differential operator, and let $\mathcal{E}$ be a generalized GF on Ω. Let $x_n = \Sigma_j \psi_j w_j \in \mathcal{H}_n^1(\Omega,\kappa)$ denote approximate weak solutions to the SPDE $\mathcal{L}x = \mathcal{E}$ on Ω.

(a) The weak *Galerkin solutions* are given by Gaussian $\mathbf{w} = \{w_1,\ldots,w_n\}$ such that

$$E(\langle f_n, \mathcal{L}x_n\rangle_\Omega) = E(\langle f_n, \mathcal{E}\rangle_\Omega),$$

$$\operatorname{cov}(\langle f_n, \mathcal{L}x_n\rangle_\Omega, \langle g_n, \mathcal{L}x_n\rangle_\Omega) = \operatorname{cov}(\langle f_n, \mathcal{E}\rangle_\Omega, \langle g_n, \mathcal{E}\rangle_\Omega)$$

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

(b) The *weak least squares solutions* are given by Gaussian $\mathbf{w} = \{w_1,\ldots,w_n\}$ such that

$$E(\langle \mathcal{L}f_n, \mathcal{L}x_n\rangle_\Omega) = E(\langle \mathcal{L}f_n, \mathcal{E}\rangle_\Omega),$$

$$\operatorname{cov}(\langle \mathcal{L}f_n, \mathcal{L}x_n\rangle_\Omega, \langle \mathcal{L}g_n, \mathcal{L}x_n\rangle_\Omega) = \operatorname{cov}(\langle \mathcal{L}f_n, \mathcal{E}\rangle_\Omega, \langle \mathcal{L}g_n, \mathcal{E}\rangle_\Omega)$$

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

C.2. 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

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

For integers $\alpha = 1, 2, 3, \ldots$, y_α is a solution to the original SPDE. To avoid solutions in the null space of $(\kappa^2 - \Delta)$, we shall 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, which show what the precision matrices should look like for given basis functions (theorem 2), that the finite Hilbert representations converge to the true distributions for $\alpha = 1$ and $\alpha = 2$ and dense Hilbert space sequences (theorem 3) and finally that the iterative constructions for $\alpha \geqslant 3$ also converge (theorem 4). A sequence $\mathcal{H}_n^1(\Omega, \kappa)$ 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 $\mathbf{C}$, $\mathbf{G}$ and $\mathbf{K}$ through

$$\begin{aligned} C_{i,j} &= \langle \psi_i, \psi_j \rangle_\Omega, \\ G_{i,j} &= \langle \nabla \psi_i, \nabla \psi_j \rangle_\Omega, \\ \mathbf{K} &= \kappa^2 \mathbf{C} + \mathbf{G} \end{aligned}$$

and denote the distribution for $\mathbf{w}$ with $N(\mathbf{0}, \mathbf{Q}^{-1})$, where the precision matrix $\mathbf{Q}$ is the inverse of the covariance matrix, and let $x_n = \Sigma_k \psi_k w_k$ be a weak $\mathcal{H}_n^1(\Omega, \kappa)$ approximation to $\mathcal{L}^{\alpha/2} x = \mathcal{E}$, $\mathcal{L} = (\kappa^2 - \Delta)$, with Neumann boundaries, and $\partial_{\mathbf{n}} \psi_k = 0$ on $\partial\Omega$.

(a) When $\alpha = 2$ and $\mathcal{E} = \mathcal{W}$, the weak Galerkin solution is obtained for $\mathbf{Q} = \mathbf{K}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{K}$.
(b) When $\alpha = 1$ and $\mathcal{E} = \mathcal{W}$, the weak least squares solution is obtained for $\mathbf{Q} = \mathbf{K}$.
(c) When $\alpha = 2$ and $\mathcal{E}$ is an $L^2(\Omega)$-bounded GF in $\mathcal{H}_n^1(\Omega, \kappa)$ with mean 0 and precision $\mathbf{Q}_{\mathcal{E},n}$, the weak Galerkin solution is obtained for $\mathbf{Q} = \mathbf{K}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{Q}_{\mathcal{E},n} \mathbf{C}^{-1} \mathbf{K}$.

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 $\mathcal{H}_n^1(\Omega, \kappa)$ approximation, when $\mathcal{W}$ is Gaussian white noise. Then,

$$x_n \overset{D\{L^2(\Omega)\}}{\rightharpoonup} x, \tag{28}$$

$$\mathcal{L}^{\alpha/2} x_n \overset{D\{L^2(\Omega)\}}{\rightharpoonup} \mathcal{L}^{\alpha/2} x, \tag{29}$$

if the sequence $\{\mathcal{H}_n^1(\Omega, \kappa), n \to \infty\}$ 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 $\mathcal{H}_n^1(\Omega, \kappa)$ approximation to y such that

$$y_n \overset{D\{L^2(\Omega)\}}{\rightharpoonup} y, \tag{30}$$

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

$$x_n \overset{D\{L^2(\Omega)\}}{\rightharpoonup} x, \tag{31}$$

$$\mathcal{L}x_n \overset{D\{L^2(\Omega)\}}{\rightharpoonup} \mathcal{L}x. \tag{32}$$

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

C.3. Intrinsic cases

When $\kappa = 0$, the Hilbert space from definition 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 $\alpha \geqslant 2$ and free boundaries, the matrix $\mathbf{K}$ in theorem 2 should be replaced by $\mathbf{G} - \mathbf{B}$ (owing Green's identity), where the elements of the (possibly asymmetric) boundary integral matrix $\mathbf{B}$ are given by $B_{i,j} = \langle \psi_i, \partial_{\mathbf{n}} \psi_j \rangle_{\partial\Omega}$. The formulations and proofs of theorem 3 and theorem 4 remain unchanged, but with the convergence defined only 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 first-order intrinsic conditional auto-regressive model on infinite lattices in $\mathbb{R}^2$ converges to the *de Wij* process, which is an intrinsic generalized Gaussian random field. As can be seen in Appendix A.1, for $\alpha = 1$ and $\kappa = 0$, the $\mathbf{Q}$-matrix (equal to $\mathbf{G}$) for a triangulated regular grid matches the ordinary intrinsic first-order conditional auto-regressive model. The null spaces of the half-Laplacian are constant functions. Choose non-trivial test functions f and g that integrate to 0 and apply theorem 3 and definition 8. This shows that the regular conditional auto-regressive 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. Oscillating and non-isotropic cases

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

$$\begin{pmatrix} h_1 - \nabla \cdot \mathbf{H}_1 \nabla & -h_2 + \nabla \cdot \mathbf{H}_2 \nabla \\ h_2 - \nabla \cdot \mathbf{H}_2 \nabla & h_1 - \nabla \cdot \mathbf{H}_1 \nabla \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \end{pmatrix} \tag{33}$$

which is equivalent to the complex SPDE

$$\{h_1 + \mathrm{i} h_2 - \nabla \cdot (\mathbf{H}_1 + \mathrm{i}\mathbf{H}_2)\nabla\}\{x_1(\mathbf{u}) + \mathrm{i}\, x_2(\mathbf{u})\} = \mathcal{E}_1(\mathbf{u}) + \mathrm{i}\, \mathcal{E}_2(\mathbf{u}). \tag{34}$$

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

To solve the coupled SPDE system (33) 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 $(x_1 \;\; x_2)^{\mathrm{T}}$ as

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

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

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

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

$$\begin{pmatrix} \mathbf{K}_1 & -\mathbf{K}_2 \\ \mathbf{K}_2 & \mathbf{K}_1 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Q}_\mathcal{E} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_\mathcal{E} \end{pmatrix} \begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{K}_1 & -\mathbf{K}_2 \\ \mathbf{K}_2 & \mathbf{K}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix},$$

where $\mathbf{Q} = \mathbf{Q}(h_1, \mathbf{H}_1) + \mathbf{Q}(h_2, \mathbf{H}_2)$, and $\mathbf{Q}(\cdot, \cdot)$ is the precision that is 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 construction of the precision is straightforward, and the local support make the basic matrices sparse. Since they are not orthogonal, the $\mathbf{C}$-matrix will be non-diagonal, and therefore the FEM construction does not directly yield Markov fields for $\alpha \geqslant 2$, since $\mathbf{C}^{-1}$ is not sparse. However, following standard practice in FEMs, $\mathbf{C}$ can be approximated with a diagonal matrix as follows. Let $\tilde{\mathbf{C}}$ be a diagonal matrix, with $\tilde{C}_{ii} = \Sigma_j C_{ij} = \langle \psi_i, 1 \rangle_\Omega$, and note that this preserves the interpretation of the matrix as an integration matrix. Substituting $\mathbf{C}^{-1}$ with $\tilde{\mathbf{C}}^{-1}$ 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 proofs of convergence. 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 $\mathbf{w}_f$ and $\mathbf{w}_g$. Taking the difference between the covariances for the Markov ($\tilde{x}_n$) and the full FEM solution (x_n) for $\alpha = 2$ yields the error

$$\operatorname{cov}(\langle f, \mathcal{L}\tilde{x}_n \rangle_\Omega, \langle g, \mathcal{L}\tilde{x}_n \rangle_\Omega) - \operatorname{cov}(\langle f, \mathcal{L}x_n \rangle_\Omega, \langle g, \mathcal{L}x_n \rangle_\Omega) = \mathbf{w}_f (\tilde{\mathbf{C}} - \mathbf{C}) \mathbf{w}_g.$$

Requiring $\|f\|_{\mathcal{H}^1(\Omega,\kappa)}, \|g\|_{\mathcal{H}^1(\Omega,\kappa)} \leqslant 1$, 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 expression (11) will not be affected by the Markov approximation. In practice, the $\mathbf{C}$ matrix in $\mathbf{K}$ should also be replaced by $\tilde{\mathbf{C}}$. 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 various methods, showing negligible differences between the exact FEM representation and the Markov approximation.

Appendix D: Proofs

D.1. Folded covariance: proof of theorem 1

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

$$\operatorname{cov}\{x(u), x(v)\} = \lambda_0 + \sum_{k=1}^{\infty} \cos(u\pi k/L) \cos(v\pi k/L) \lambda_k, \tag{35}$$

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 statement of theorem 1, 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

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

Thus, with $\tilde{r}(u, v)$ denoting the folded covariance in the statement of theorem 1

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

Rewriting the cosines via Euler's formulae, we obtain

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

where we used the Dirac measure representation

$$\sum_{k=-\infty}^{\infty}\exp(\mathrm{i}ks) = 2\pi\sum_{k=-\infty}^{\infty}\delta(s-2\pi k).$$

Finally, combining the results yields

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

which is precisely the expression sought in equation(35).

D.2. Modified half-Laplacian equivalence: proof of lemma 2

For brevity, we present only the proof for compact manifolds, as 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 $\lambda_k \geqslant 0$, $k = 0, 1, 2, \ldots$, be the eigenvalue corresponding to eigenfunction E_k of $-\Delta$ (definition 1). Then, with $\hat{\phi}(k) = (\mathcal{F}\phi)(k)$, the modified half-Laplacian from Appendix B.3.2. is defined through $\mathcal{F}\{(\kappa^2 - \Delta)^{1/2}\phi\}(k) = (\kappa^2+\lambda_k)^{1/2}\hat{\phi}(k)$, and we obtain

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

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

$$\langle(\kappa^2-\Delta)^{1/2}\phi(\kappa^2-\Delta)^{1/2}\psi\rangle_\Omega = \sum_{k=0}^{\infty}(\kappa^2+\lambda_k)\,\hat{\phi}(k)\,\hat{\psi}(k),$$

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

Now, starting from the Hilbert space inner product,

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

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

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

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

$$\langle \phi, \psi \rangle_{\mathcal{H}^1(\Omega,\kappa)} = \kappa^2 \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} \hat{\phi}(k)\, \hat{\psi}(k') \langle E_k E_{k'} \rangle_{\Omega} + \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} \hat{\phi}(k)\, \hat{\psi}(k') \langle \Omega \nabla E_k \nabla E_{k'} \rangle_{\Omega}.$$

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

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

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

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

By the boundary requirements in lemma 2, whenever Green's identity holds, the boundary integral vanishes, either because the boundary is empty (if the manifold is closed), or the integrand is 0, so collecting all the terms we obtain

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

and the proof is complete.

D.3. Hilbert space convergence

D.3.1. *Proof of theorem 2 (finite element precisions)*

The proofs for theorem 2 are straightforward applications of the definitions. Let $\mathbf{w}_f$ and $\mathbf{w}_g$ be the Hilbert space co-ordinates 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

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

owing to Green's identity, and

$$\operatorname{cov}(\langle f_n, \mathcal{L} x_n \rangle_{\Omega}, \langle g_n, \mathcal{L} x_n \rangle_{\Omega}) = \mathbf{w}_f^{\mathrm{T}} \mathbf{K} \operatorname{cov}(\mathbf{w}, \mathbf{w}) \mathbf{K}^{\mathrm{T}} \mathbf{w}_g.$$

This covariance is equal to

$$\begin{aligned} \operatorname{cov}(\langle f_n, \mathcal{W} \rangle_{\Omega}, \langle g_n, \mathcal{W} \rangle_{\Omega}) &= \langle f_n, g_n \rangle_{\Omega} \\ &= \sum_{i,j} w_{f,i} \langle \psi_i, \psi_j \rangle_{\Omega} w_{g,j} \\ &= \sum_{i,j} w_{f,i} C_{i,j} w_{g,j} \\ &= \mathbf{w}_f^{\mathrm{T}} \mathbf{C} \mathbf{w}_g \end{aligned}$$

for every pair of test functions f_n, g_n when $\mathbf{Q} = \operatorname{cov}(\mathbf{w}, \mathbf{w})^{-1} = \mathbf{K}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{K}$.

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, $\langle \mathcal{L}^{1/2} f_n, \mathcal{L}^{1/2} x_n \rangle_{\Omega} = \langle f_n x_n \rangle_{\mathcal{H}^1(\Omega,\kappa)} = \mathbf{w}_f^{\mathrm{T}} \mathbf{K} \mathbf{w}$ and

$$\operatorname{cov}(\langle \mathcal{L}^{1/2} f_n, \mathcal{W} \rangle_{\Omega}, \langle \mathcal{L}^{1/2} g_n, \mathcal{W} \rangle_{\Omega}) = \langle \mathcal{L}^{1/2} f_n, \mathcal{L}^{1/2} g_n \rangle_{\Omega} = \langle f_n g_n \rangle_{\mathcal{H}^1(\Omega,\kappa)} = \mathbf{w}_f^{\mathrm{T}} \mathbf{K} \mathbf{w}_g$$

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

Finally, for case (c), $\alpha = 2$ and $\mathcal{E} = \mathcal{E}_n$ is a GF on $\mathcal{H}_n^1(\Omega, \kappa)$ with precision $\mathbf{Q}_{\mathcal{E},n}$. Using the same technique as for (a),

$$\mathrm{cov}(\langle f_n, \mathcal{L}x_n\rangle_\Omega, \langle g_n, \mathcal{L}x_n\rangle_\Omega) = \mathbf{w}_f^{\mathrm{T}}\mathbf{K}\,\mathrm{cov}(\mathbf{w},\mathbf{w})\mathbf{K}^{\mathrm{T}}\mathbf{w}_g.$$

and the finite basis representation of the noise $\mathcal{E}_n$ gives

$$\mathrm{cov}(\langle f_n, \mathcal{E}_n\rangle_\Omega, \langle g_n, \mathcal{E}_n\rangle_\Omega) = \mathbf{w}_f^{\mathrm{T}}\mathbf{C}\mathbf{Q}_{\mathcal{E},n}^{-1}\mathbf{C}\mathbf{w}_g.$$

Requiring equality for all pairs of test functions yields $\mathbf{Q} = \mathbf{K}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{Q}_{\mathcal{E},n}\mathbf{C}^{-1}\mathbf{K}$. Here, keeping the transposes allows the proof to apply also to the intrinsic free-boundary cases.

D.3.2. Proof of theorem 3 (convergence)

First, we show that expression (28) follows from expression (29). Let $\mathcal{L} = (\kappa^2 - \Delta)$, let f and g be functions in $\mathcal{H}^1(\Omega, \kappa)$, and let $\tilde{f}$ the solution to the PDE

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

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

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

and

$$\langle f, x\rangle_\Omega = \langle \mathcal{L}\tilde{f}, x\rangle_\Omega = \langle \tilde{f}, x\rangle_{\mathcal{H}^1(\Omega,\kappa)} = \langle \tilde{f}, \mathcal{L}x\rangle_\Omega,$$

where the last equality holds when $\alpha = 2$, since $\mathcal{W}$ is $L^2(\Omega)$ bounded. The convergence of x_n to x follows from expression (29). In the Galerkin case (a), we have

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

and similarly for the least squares case (b).

For expression (29), let $f_n = \Sigma_k \psi_k w_{f,k}$ and $g_n = \Sigma_k \psi_k w_{g,k}$ be the orthogonal projections of f and g onto $\mathcal{H}_n^1\Omega, \kappa$. 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

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

as $n \to \infty$. Similarly in case (b), for any $f \in \mathcal{H}^1(\Omega, \kappa)$ fulfilling the requirements of lemma 2,

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

and

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

as $n \to \infty$.

D.3.3. Proof of theorem 4 (iterative convergence)
First, we show that expression (31) follows from expression (32). Let $\tilde{f}$ and $\tilde{g}$ be defined as in the proof of theorem 3. Then, since $\mathcal{L} = \kappa^2 - \Delta$,

$$\langle f, x_n \rangle_\Omega = \langle \tilde{f}, \mathcal{L} x_n \rangle_\Omega$$

and

$$\langle f, x \rangle_\Omega = \langle \tilde{f}, \mathcal{L} x \rangle_\Omega,$$

and the convergence of x_n to x follows from expression (32). For expression (32) as in the proof of theorem 3, $\langle f, \mathcal{L} x_n \rangle_\Omega = \langle f_n, x_n \rangle_{\mathcal{H}^1(\Omega,\kappa)}$, and

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

as $n \to \infty$, due to requirement (30).

References

Adler, R. J. (2009) *The Geometry of Random Fields*. Philadelphia: Society for Industrial and Applied Mathematics.
Adler, R. J. and Taylor, J. (2007) *Random Fields and Geometry*. New York: Springer.
Allcroft, D. J. and Glasbey, C. A. (2003) A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *Appl. Statist.*, **52**, 487–498.
Arjas, E. and Gasbarra, D. (1996) Bayesian inference of survival probabilities, under stochastic ordering constraints. *J. Am. Statist. Ass.*, **91**, 1101–1109.
Auslander, L. and MacKenzie, R. E. (1977) *Introduction to Differentiable Manifolds*. New York: Dover Publications.
Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman and Hall.
Banerjee, S., Gelfand, A. E. Finley, A. O. and Sang, H. (2008) Gaussian predictive process models for large spatial data sets. *J. R. Statist. Soc.* B, **70**, 825–848.
Bansal, R., Staib, L. H., Xu, D., Zhu, H. and Peterson, B. S. (2007) Statistical analyses of brain surfaces using Gaussian random fields on 2-D manifolds. *IEEE Trans. Med. Imgng*, **26**, 46–57.
Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B, **36**, 192–236.
Besag, J. (1975) Statistical analysis of non-lattice data. *Statistician*, **24**, 179–195.
Besag, J. (1981) On a system of two-dimensional recurrence equations. *J. R. Statist. Soc.* B, **43**, 302–309.
Besag, J. and Kooperberg, C. (1995) On conditional and intrinsic autoregressions. *Biometrika*, **82**, 733–746.
Besag, J. and Mondal, D. (2005) First-order intrinsic autoregressions and the de Wijs process. *Biometrika*, **92**, 909–920.
Besag, J., York, J. and Mollié, A. (1991) Bayesian image restoration with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.*, **43**, 1–59.
Bolin, D. and Lindgren, F. (2009) Wavelet Markov models as efficient alternatives to tapering and convolution fields. *Preprint* 2009:13. Lund University, Lund.
Bolin, D. and Lindgren, F. (2011) Spatial models generated by nested stochastic partial differential equations. *Ann. Appl. Statist.*, to be published.
Brenner, S. C. and Scott, R. (2007) *The Mathematical Theory of Finite Element Methods*, 3rd edn. New York: Springer.
Brohan, P., Kennedy, J., Harris, I., Tett, S. and Jones, P. (2006) Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.*, 111.
Chen, C. M. and Thomée, V. (1985) The lumped mass finite element method for a parabolic problem. *J. Aust. Math. Soc.* B, **26**, 329–354.
Chilés, J. P. and Delfiner, P. (1999) *Geostatistics: Modeling Spatial Uncertainty*. Chichester: Wiley.
Ciarlet, P. G. (1978) *The Finite Element Method for Elliptic Problems*. Amsterdam: North-Holland.
Cressie, N. A. C. (1993) *Statistics for Spatial Data*. New York: Wiley.
Cressie, N. and Huang, H. C. (1999) Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Am. Statist. Ass.*, **94**, 1330–1340.
Cressie, N. and Johannesson, G. (2008) Fixed rank kriging for very large spatial data sets. *J. R. Statist. Soc.* B, **70**, 209–226.

Cressie, N. and Verzelen, N. (2008) Conditional-mean least-squares fitting of Gaussian Markov random fields to Gaussian fields. *Computatnl Statist. Data Anal.*, **52**, 2794–2807.
Dahlhaus, R. and Künsch, H. R. (1987) Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, **74**, 877–882.
Das, B. (2000) Global covariance modeling: a deformation approach to anisotropy. *PhD Thesis*. Department of Statistics, University of Washington, Seattle.
Davis, T. A. (2006) *Direct Methods for Sparse Linear Systems*. Philadelphia: Society for Industrial and Applied Mathematics.
Diggle, P. J. and Ribeiro, P. J. (2006) *Model-based Geostatistics*. New York: Springer.
Duff, I. S., Erisman, A. M. and Reid, J. K. (1989) *Direct Methods for Sparse Matrices*, 2nd edn. New York: Clarendon.
Edelsbrunner, H. (2001) *Geometry and Topology for Mesh Generation*. Cambridge: Cambridge University Press.
Eidsvik, J., Finley, A. O., Banerjee, S. and Rue, H. (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, Trondheim.
Federer, H. (1951) Hausdorff measure and Lebesgue area. *Proc. Natn Acad Sci. USA*, **37**, 90–94.
Federer, H. (1978) Colloquium lectures on geometric measure theory. *Bull. Am. Math. Soc.*, **84**, 291–338.
Fuentes, M. (2001) High frequency kriging for nonstationary environmental processes. *Environmetrics*, **12**, 469–483.
Fuentes, M. (2008) Approximate likelihood for large irregular spaced spatial data. *J. Am. Statist. Ass.*, **102**, 321–331.
Furrer, R., Genton, M. G. and Nychka, D. (2006) Covariance tapering for interpolation of large spatial datasets. *J. Computnl Graph. Statist.*, **15**, 502–523.
George, A. and Liu, J. W. H. (1981) *Computer Solution of Large Sparse Positive Definite Systems*. Englewood Cliffs: Prentice Hall.
Gneiting, T. (1998) Simple tests for the validity of correlation function models on the circle. *Statist. Probab. Lett.*, **39**, 119–122.
Gneiting, T. (2002) Nonseparable, stationary covariance functions for space-time data. *J. Am. Statist. Ass.*, **97**, 590–600.
Gneiting, T., Kleiber, W. and Schlather, M. (2010) Matérn cross-covariance functions for multivariate random fields. *J. Am. Statist. Ass.*, **105**, 1167–1177.
Gschlößl, S. and Czado, C. (2007) Modelling count data with overdispersion and spatial effects. *Statistical Papers*. (Available from `http://dx.doi.org/10.1007/s00362-006-0031-6`.)
Guttorp, P. and Gneiting, T. (2006) Studies in the history of probability and statistics XLIX: on the Matérn correlation family. *Biometrika*, **93**, 989–995.
Guyon, X. (1982) Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, **69**, 95–105.
Hansen, J., Ruedy, R., Glascoe, J. and Sato, M. (1999) GISS analysis of surface temperature change. *J. Geophys. Res.*, **104**, 30997–31022.
Hansen, J., Ruedy, R., Sato, M., Imhoff, M., Lawrence, W., Easterling, D., Peterson, T. and Karl, T. (2001) A closer look at United States and global surface temperature change. *J. Geophys. Res.*, **106**, 23947–23963.
Hartman, L. and Hössjer, O. (2008) Fast kriging of large data sets with Gaussian Markov random fields. *Computnl Statist. Data Anal.*, **52**, 2331–2349.
Heine, V. (1955) Models for two-dimensional stationary stochastic processes. *Biometrika*, **42**, 170–178.
Henderson, R., Shimakura, S. and Gorst, D. (2002) Modelling spatial variation in leukemia survival data. *J. Am. Statist. Ass.*, **97** 965–972.
Higdon, D. (1998) A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environ. Ecol. Statist.*, **5**, 173–190.
Higdon, D., Swall, J. and Kern, J. (1999) Non-stationary spatial modelling. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 761–768. New York: Oxford University Press.
Hjelle, Ø. and Dæhlen, M. (2006) *Triangulations and Applications*. Berlin: Springer.
Hrafnkelsson, B. and Cressie, N. A. C. (2003) Hierarchical modeling of count data with application to nuclear fall-out. *Environ. Ecol. Statist.* **10**, 179–200.
Hughes-Oliver, J. M., Gonzalez-Farias, G. Lu, J. C. and Chen, D. (1998) Parametric nonstationary correlation models. *Statist. Probab. Lett.*, **40**, 267–278.
Ilić, M., Turner, I. W. and Anh, V. (2008) A numerical solution using an adaptively preconditioned Lanczos method for a class of linear systems related with the fractional Poisson equation. *J. Appl. Math. Stoch. Anal.*, 104525.
Jones, R. H. (1963) Stochastic processes on a sphere. *Ann. Math. Statist.*, **34**, 213–218.
Jun, M. and Stein, M. L. (2008) Nonstationary covariancs models for global data. *Ann. Appl. Statist.*, **2**, 1271–1289.

Karypis, G. and Kumar, V. (1998) *METIS: a Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-reducing Orderings of Sparse Matrices, Version 4.0*. Minneapolis: University of Minnesota. (Available from `http://www-users.cs.umn.edu/~karypis/metis/index.html`.)

Kneib, T. and Fahrmeir, L. (2007) A mixed model approach for geoadditive hazard regression. *Scand. J. Statist.*, **34**, 207–228.

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

Lindgren, F. and Rue, H. (2008) A note on the second order random walk model for irregular locations. *Scand. J. Statist.*, **35**, 691–700.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

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

Peterson, T. and Vose, R. (1997) An overview of the Global Historical Climatology Network temperature database. *Bull. Am. Meteorol. Soc.*, **78**, 2837–2849.

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

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

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

Rue, H. (2001) Fast sampling of Gaussian Markov random fields. *J. R. Statist. Soc.* B **63**, 325–338.

Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman and Hall.

Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *J. R. Statist. Soc.* B, **71**, 319–392.

Rue, H. and Tjelmeland, H. (2002) Fitting Gaussian Markov random fields to Gaussian fields. *Scand. J. Statist.*, **29**, 31–50.

Samko, S. G., Kilbas, A. A. and Maricev, O. I. (1992) *Fractional Integrals and Derivatives: Theory and Applications*. Yverdon: Gordon and Breach.

Sampson, P. D. and Guttorp, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Statist. Ass.*, **87**, 108–119.

Smith, T. (1934) Change of variables in Laplace's and other second-order differential equations. *Proc. Phys. Soc.*, **46**, 344–349.

Song, H., Fuentes, M. and Gosh, S. (2008) A compariative study of Gaussian geostatistical models and Gaussian Markov random field models. *J. Multiv. Anal.*, **99**, 1681–1697.

Stein, M. (2005) Space-time covariance functions. *J. Am. Statist. Ass.*, **100**, 310–321.

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

Stein, M. L., Chi, Z. and Welty, L. J. (2004) Approximating likelihoods for large spatial data sets. *J. R. Statist. Soc.* B, **66**, 275–296.

Vecchia, A. V. (1988) Estimation and model identification for continuous spatial processes. *J. R. Statist. Soc.* B, **50**, 297–312.

Wahba, G. (1981) Spline interpolation and smoothing on the sphere. *SIAM J. Scient. Statist. Comput.*, **2**, 5–16.

Wall, M. M. (2004) A close look at the spatial structure implied by the CAR and SAR models. *J. Statist. Planng and Inf.*, **121**, 311–324.

Weir, I. S. and Pettitt, A. N. (2000) Binary probability maps using a hidden conditional autoregressive Gaussian process with an application to Finnish common toad data. *Appl. Statist.*, **49**, 473–484.

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

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

Yue, Y. and Speckman, P. (2010) Nonstationary spatial Gaussian Markov random fields. *J. Computnl Graph. Statist.*, **19**, 96–116.