

# WILEY

---

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

Author(s): Finn Lindgren, Håvard Rue, Johan Lindström, John T. Kent, Peter J. Diggle, J. B. Illian, D. P. Simpson, Tilmann Gneiting, Michael Scheuerer, R. Furrer, E. Furrer, D. Nychka, Paul Fearnhead, Peter Challenor, Yiannis Andrianakis, Gemma Stephenson, Jesper Møller, Xiangping Hu, Daniel Simpson, David Bolin, Patrick E. Brown, Michela Cameletti, Sara Martino, Daniel Cooley, Jennifer A. Hoeting, Rosa M. Crujeiras, Andrés Prieto, Marco A. R. Ferreira...

Source: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 73, No. 4 (SEPTEMBER 2011), pp. 423-498

Published by: Wiley for the Royal Statistical Society

Stable URL: <http://www.jstor.org/stable/41262260>

Accessed: 09-01-2018 08:45 UTC

## REFERENCES

Linked references are available on JSTOR for this article:

[http://www.jstor.org/stable/41262260?seq=1&cid=pdf-reference#references\\_tab\\_contents](http://www.jstor.org/stable/41262260?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

*Royal Statistical Society, Wiley* are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*

# 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. Titterton 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

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  $\pi(x_i | \mathbf{x}_{-i})$ ,  $i = 1, \dots, n$ , depend only on a set of neighbours  $\partial i$  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  $\{s_i\}$ , to construct a discretized GF with covariance matrix  $\Sigma$ .
- (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  $\Sigma$  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 follow 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

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

Here,  $K_\nu$  is the modified Bessel function of the second kind and order  $\nu > 0$ ,  $\kappa > 0$  is a scaling parameter and  $\sigma^2$  is the marginal variance. The integer value of  $\nu$  determines the mean-square differentiability of the underlying process, which matters for predictions that are made by using such a model. However,  $\nu$  is usually fixed since it is poorly identified in typical applications. A more natural interpretation of the scaling parameter  $\kappa$  is as a range parameter  $\rho$ ; the Euclidean distance where  $x(\mathbf{u})$  and  $x(\mathbf{v})$  are 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  $\rho$ , for all  $\nu$ .

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

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

where  $(\kappa^2 - \Delta)^{\alpha/2}$  is a pseudodifferential 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,  $\Delta$  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  $\kappa \rightarrow 0$  or  $\nu \rightarrow 0$  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  $\alpha \geq 2$  the null space of the differential operator is non-trivial, containing, for example, the functions  $\exp(\kappa \mathbf{e}^T \mathbf{u})$ , 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  $\phi$  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 informal 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

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

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

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

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

$$\text{cov}(x_{ij}, x_{i'j'}) \approx \frac{1}{2\pi} K_0\{l\sqrt{(a-4)}\}, \quad l \neq 0,$$

where  $l$  is the Euclidean distance between  $ij$  and  $i'j'$ . Evaluated for continuous distances, this is a generalized covariance function, which is obtained from equation (1) in the limit  $\nu \rightarrow 0$ , with  $\kappa^2 = a - 4$  and  $\sigma^2 = 1/4\pi$ , 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 representation (6):  $\nu = 1$ ,

$$\begin{vmatrix} 1 & & & & \\ -2a & 2 & & & \\ 4+a^2 & -2a & 1 & & \\ & & & & \\ & & & & \end{vmatrix}$$

$\nu = 2$ ,

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

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

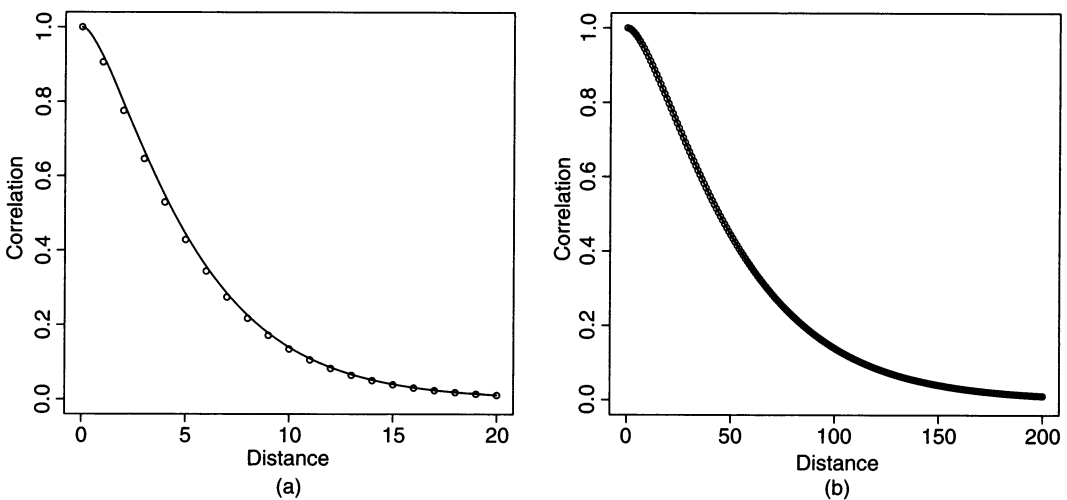


Fig. 1. Matérn correlations (—) for range (a) 10 and (b) 100, and the correlations for the GMRF representation (o)

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  $\nu = 1, 2, \dots$ , is found by convolving model (6) by itself  $\nu$  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](http://www.r-inla.org)) is based on Hjelle and Dæhlen (2006).

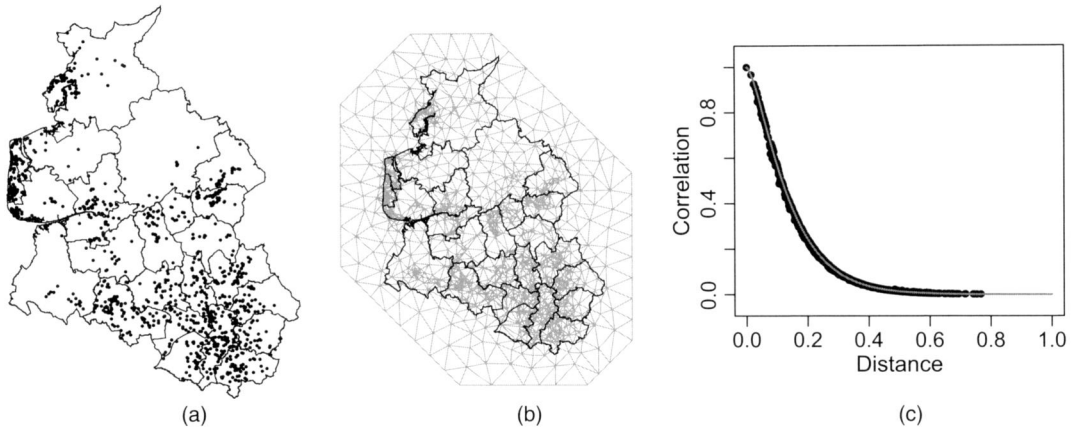
To illustrate the process of triangulation of  $\mathbb{R}^2$ , we shall use an example from Henderson *et al.* (2002) which models spatial variation in leukaemia survival data in north-west England. Fig. 2(a) displays the locations of 1043 cases of acute myeloid leukaemia in adults who were diagnosed 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

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

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

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



**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*  $\{\phi_j(\mathbf{u}), j = 1, \dots, m\}$ , where ‘=’<sup>d</sup> 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  $\psi_k$  that are piecewise linear in each triangle, defined such that  $\psi_k$  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  $\alpha \geq 3$ , 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 **C**, **G** and **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  $\alpha = 1, 2, \dots$ , as a function of  $\kappa^2$ . Then the finite dimensional representations of the solutions to equation (2) have precisions

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

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  $\kappa^2$ . 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  $3 \times 3$  for  $\alpha = 1$ , is  $5 \times 5$  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  $\nu$  in the Matérn covariance function, these results correspond to  $\nu = 1/2, 3/2, 5/2, \dots$ , in  $\mathbb{R}^1$  and  $\nu = 0, 1, 2, \dots$ , in  $\mathbb{R}^2$ .
- (f) We are currently unable to provide results for other values of  $\alpha$ ; 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  $\alpha = 1, 2, 3, \dots$ ; see equation (3). This result indicates that a different approach may be needed to provide representation results when  $\alpha$  is not an integer, such as approximating the spectrum itself. Given approximations for general  $0 \leq \alpha \leq 2$ , the recursive approach could then be used for general  $\alpha > 2$ .

Although the approach does give a GMRF representation of the Matérn field on the triangulated region, it is truly an approximation to the stochastic weak solution as we 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  $\alpha = 2$ ,

$$\sup_{f \in \mathcal{H}^1: \|f\|_{\mathcal{H}^1} \leq 1} \{E(\langle f, x_n - x \rangle_{\mathcal{H}^1}^2)\} \leq ch^2. \quad (11)$$

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  $d \geq 1$ , with  $h$  proportional to the edge lengths between the vertices, when the minimal mesh angles are bounded away from zero.

To see how well we 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).

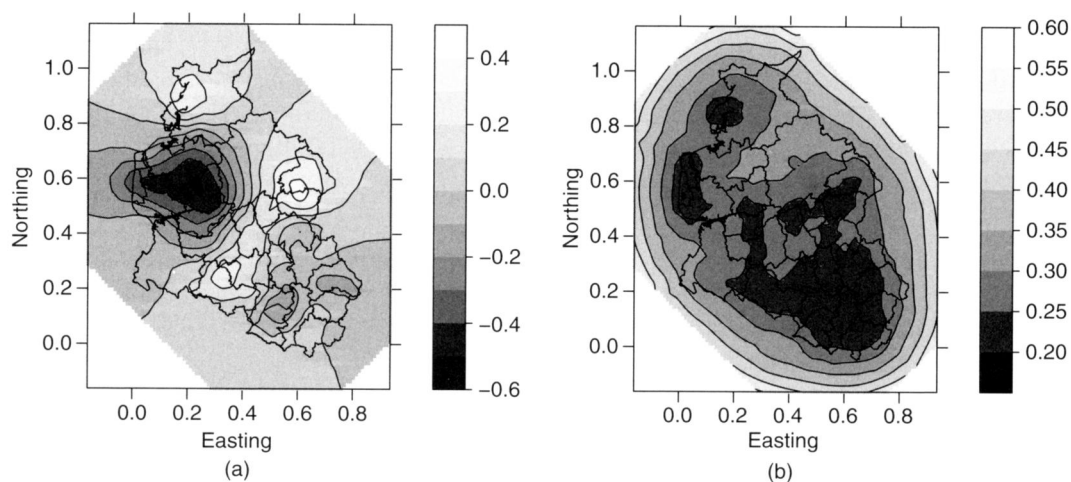
#### 2.4. Leukaemia example

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

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

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  $1749 \times 1749$  precision matrix for the spatial component. We ran the model in R-*inla* ([www.r-inla.org](http://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  $2797 \times 2797$  (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 expressions (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.

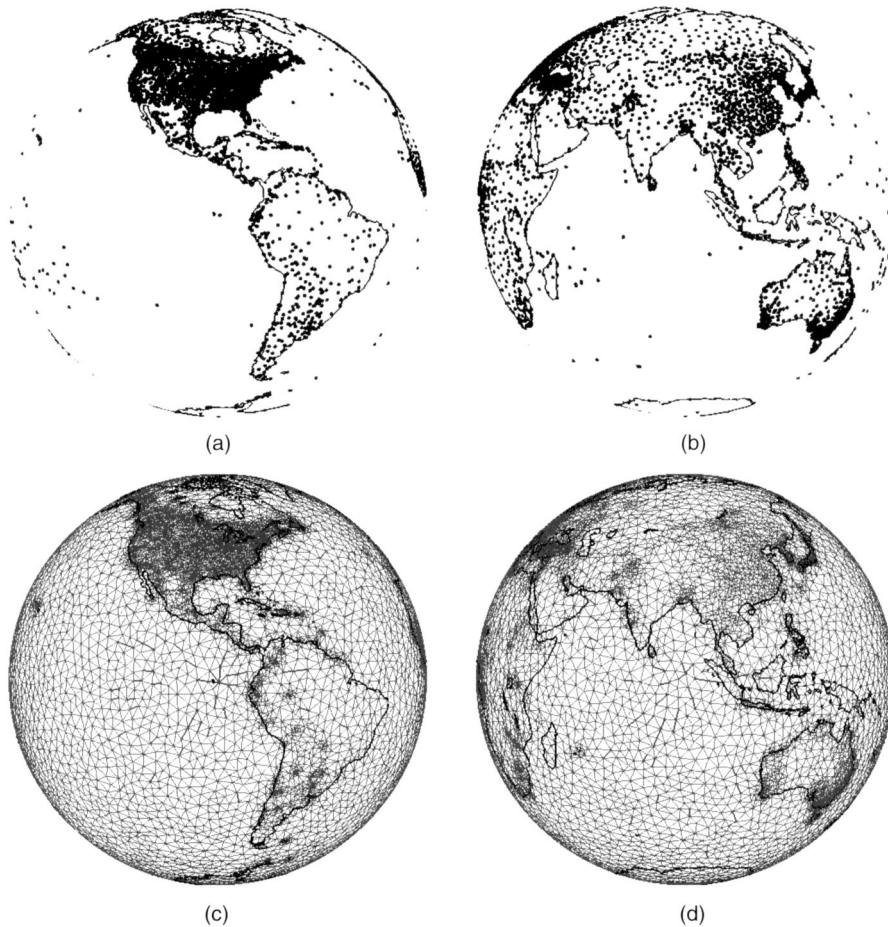
#### 3.1. Matérn fields on manifolds

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, 2010). 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.

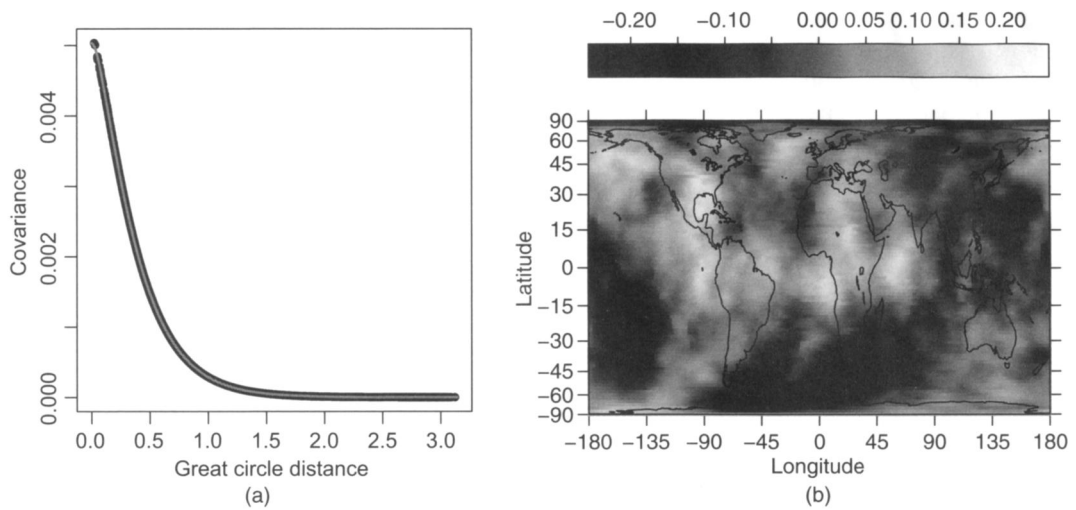
To illustrate the continuous index definition and the Markov representation of Matérn fields on a sphere, Fig. 4 shows the locations of 7280 meteorological measurement stations on the globe, together with an irregular triangulation. The triangulation was constrained to have



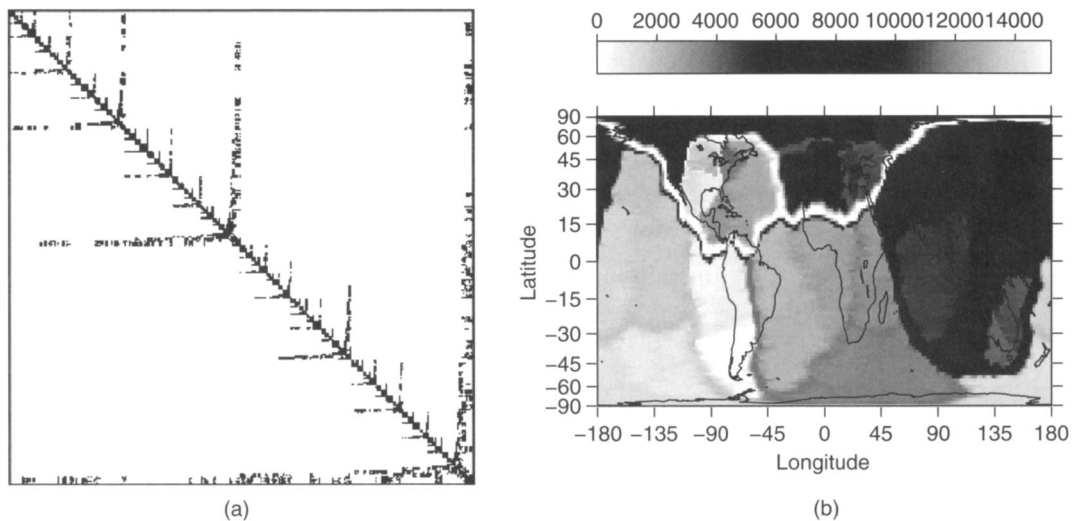


**Fig. 4.** (a), (b) Data locations and (c), (d) triangulation for the global temperature data set analysed in Section 4, with a coastline map superimposed

minimal angles  $21^\circ$  and maximum edge lengths corresponding to 500 km based on an average Earth radius of 6370 km. The triangulation includes all the stations more than 10 km apart, requiring a total of 15182 vertices and 30360 triangles. The resulting 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, 1999), 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  $\kappa^2$  and the innovation variance are constant in space. In general, we can allow both parameters to depend on the co-ordinate  $\mathbf{u}$ , 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  $\kappa^2$  and  $\tau$  can be considered as constant within the support of the basis functions  $\{\psi_k\}$ , and hence

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

for a naturally defined  $\mathbf{u}_j^*$  in the support of  $\psi_i$  and  $\psi_j$ . 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  $\theta$ , the complex version becomes

$$\{\kappa^2 \exp(i\pi\theta) - \Delta\} \{x_1(\mathbf{u}) + ix_2(\mathbf{u})\} = \mathcal{W}_1(\mathbf{u}) + i\mathcal{W}_2(\mathbf{u}), \quad 0 \leq \theta < 1. \tag{14}$$

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 is given in Appendix A.

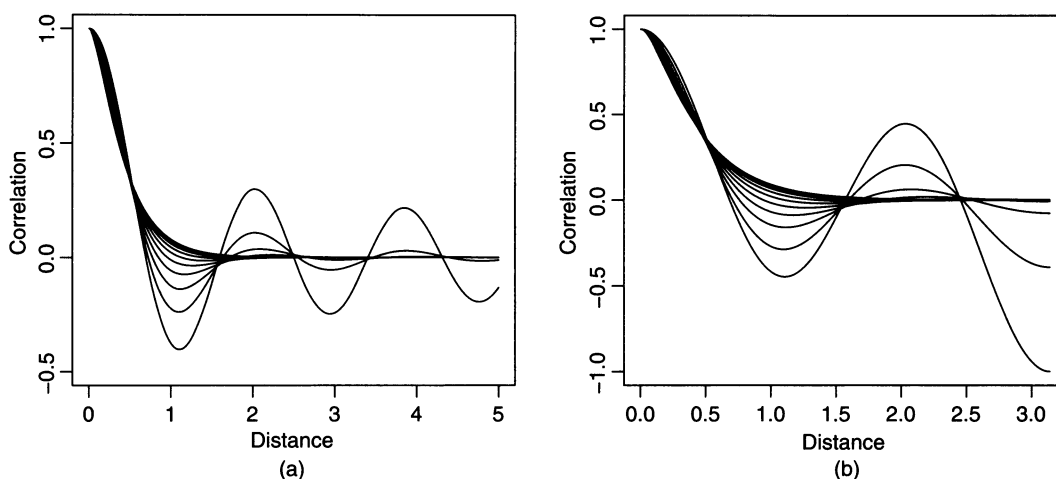


Fig. 7. Correlation functions from oscillating SPDE models, for  $\theta = 0, 0.1, \dots, 1$ , on (a)  $\mathbb{R}^2$  and (b)  $\mathbb{S}^2$ , with  $\kappa^2 = 12$  and  $\nu = 1$

The details 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  $\theta$ . 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  $\kappa$ .

### 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 be 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  $\Omega$  yields (Smith, 1934)

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

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 noise, 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

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

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

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

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  $\alpha \geq 1$ , 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 equations 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 597373 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.

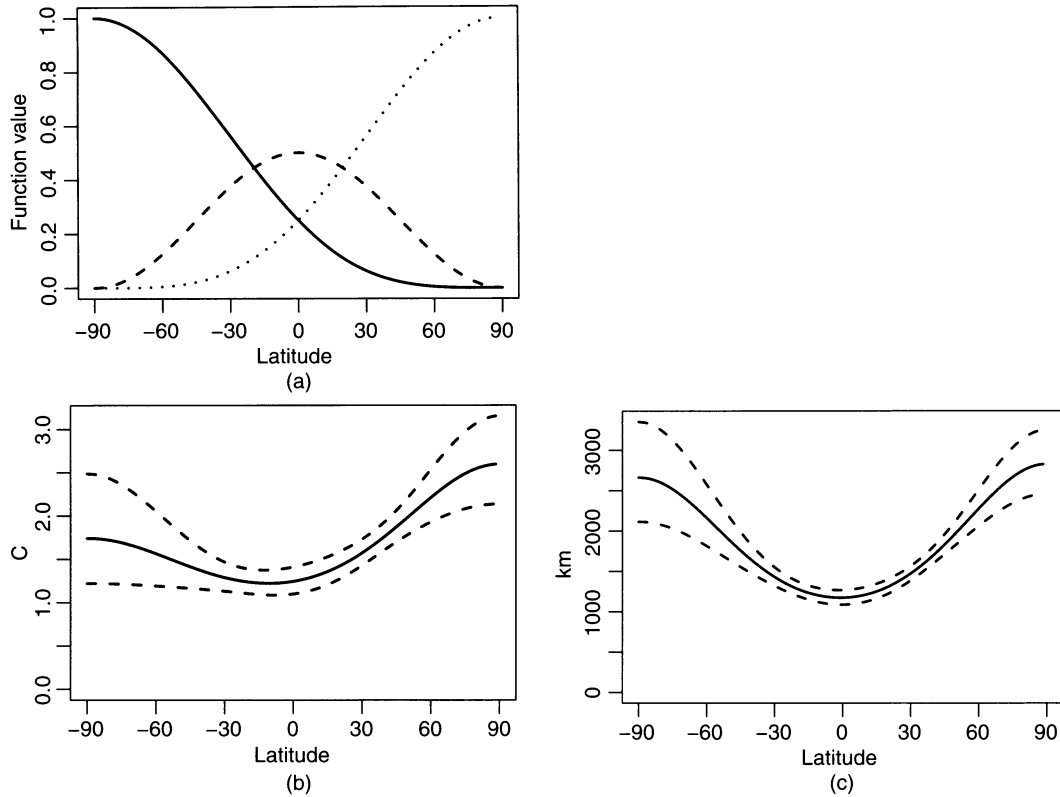
The Global Historical Climatology Network data are used to analyse regional and global temperatures in the GISTEMP temperature series (see <http://data.giss.nasa.gov/gistemp>) (Hansen *et al.*, 1999, 2001) and HadCRUT3 (Brohan *et al.*, 2006) global temperature series, together with additional data such 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 known 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 gridding 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 equation (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  $\theta = \{\theta_\mu, \theta_\kappa, \theta_\tau, \theta_s, \theta_\varepsilon\}$ , and we denote the yearly temperature fields  $\mathbf{x} = \{\mathbf{x}_t\}$  and the yearly observations  $\mathbf{y} = \{\mathbf{y}_t\}$ , with  $t = 1970, \dots, 1989$ . 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  $\sin(\text{latitude})$ , shown in Fig. 8), the expectation field is given by  $\mu_{\mathbf{x}|\theta} = \mathbf{B}_\mu \theta_\mu$ , the local spatial dependence  $\kappa(\mathbf{u})$  is defined through  $\log(\kappa^2) = \mathbf{B}_\kappa \theta_\kappa$  and the local variance scaling  $\tau(\mathbf{u})$  is defined through  $\log(\tau) = \mathbf{B}_\tau \theta_\tau$ . 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 yearly weather, as functions of latitude

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

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

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

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

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,

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

where  $\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}} = \mathbf{Q}_{\mathbf{x}|\boldsymbol{\theta}} + \mathbf{A}_t^T \mathbf{Q}_{\mathbf{y}|\mathbf{x},\boldsymbol{\theta}} \mathbf{A}_t$  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 is 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  $\mathbf{Q}_{\mathbf{x}|\boldsymbol{\theta}}$ . The computational cost of the kriging estimates is  $\mathcal{O}(n)$  in the number of observations, and approximately  $\mathcal{O}(n^{3/2})$  in the number of basis functions. If basis functions with non-compact support had been used, such as a Fourier basis, the posterior precisions would have been fully dense matrices, with computational cost  $\mathcal{O}(n^3)$  in the number of basis functions, regardless of the sparsity of the prior precisions. This shows that when constructing computationally efficient models it is not enough to consider the theoretical properties of the prior model, but instead the whole sequence of computations needs to be taken into account.

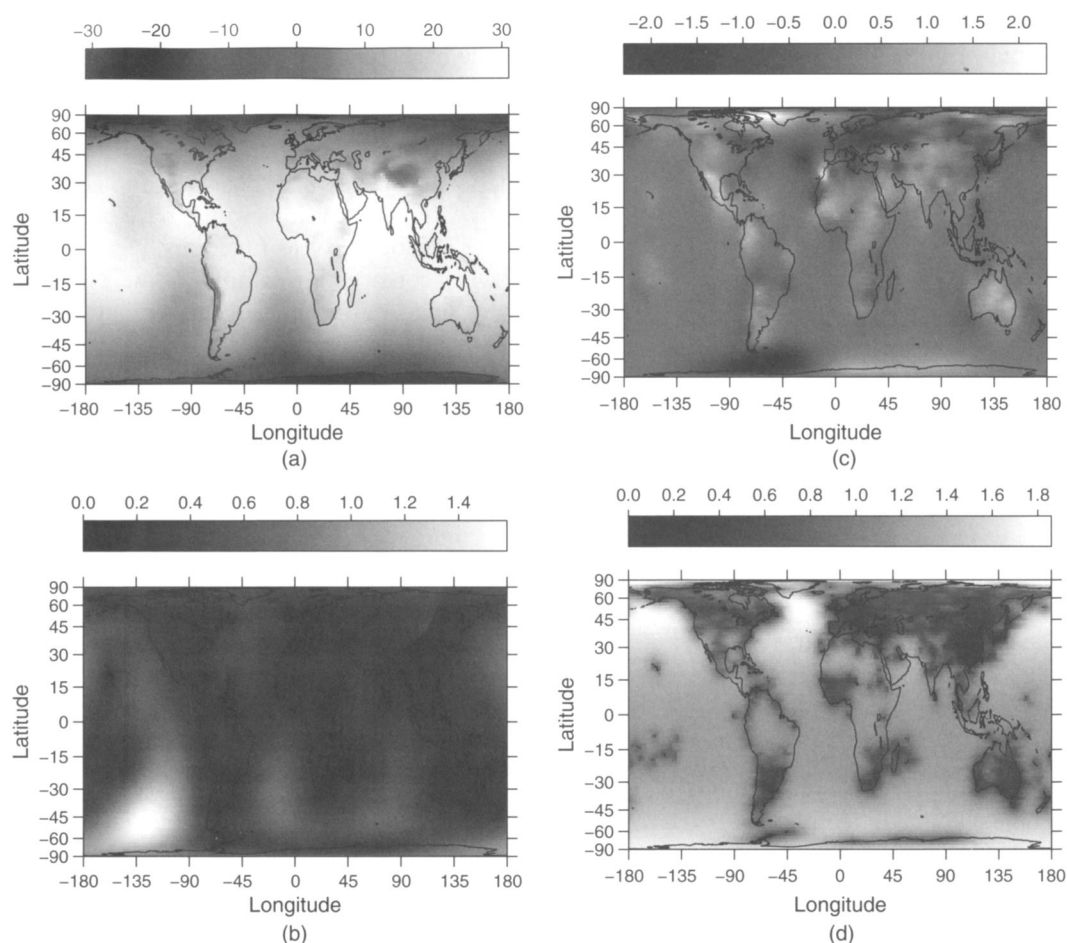
### 4.3. Results

We implemented the model 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 50 Gbytes 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 consider  $(\bar{\mathbf{x}}|\mathbf{y})$  and  $(\mathbf{x}_t - \bar{\mathbf{x}}|\mathbf{y})$ , where  $\bar{\mathbf{x}} = \sum_{t=1970}^{1989} \mathbf{x}_t / 20$ . In Figs 9(a) and 9(b), the posterior expectation of the empirical climate,  $E(\bar{\mathbf{x}}|\mathbf{y})$ , is shown (including the estimated effect of elevation), together with the posterior expectation of the temperature anomaly for 1980,  $E(\mathbf{x}_{1980} - \bar{\mathbf{x}}|\mathbf{y})$ . 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 estimated cooling effect





**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

of  $5.2\text{ }^{\circ}\text{C}$  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)  $^{\circ}\text{C}$  per century (expectation 1.52  $^{\circ}\text{C}$ ) 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  $^{\circ}\text{C}$ . Comparing the values with the corresponding estimates in the GISS series, which has an observed trend of 1.48  $^{\circ}\text{C}$  per century for this period, yields a standard deviation for the differences between the series of only 0.04  $^{\circ}\text{C}$ . 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

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

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

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

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

where  $a_i = H/\gamma_i s_i$ ,  $b_i = H/\delta_i s_i$  and  $c_i = \kappa^2 + a_i + b_i$ . If the spacing is regular,  $s = \delta = \gamma$ , and  $a = a_i = b_i \equiv H/\delta^2$  and  $c = c_i \equiv \kappa^2 + 2a$ . The special case  $\alpha = 2$  with  $\kappa = 0$  and irregular spacing is a generalization of Lindgren and Rue (2008).

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

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

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

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

For 2-manifolds such as regions in  $\mathbb{R}^2$  or on  $\mathbb{S}^2$ , we require a triangulation with a set of vertices  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , embedded in  $\mathbb{R}^3$ . Each vertex  $\mathbf{v}_k$  is assigned a continuous piecewise linear basis function  $\psi_k$  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  $\theta_0, \theta_1$  and  $\theta_2$ .

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

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

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

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

and the entire contribution from the triangle is

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

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

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

where  $b_k = \mathbb{1}(\text{edge } k \text{ in } T \text{ lies on } \partial\Omega)$ . 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

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

where  $\text{adj}(\mathbf{H})$  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  $\Omega$ . 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

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

where  $\kappa^2$  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

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

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  $\alpha$ , then

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

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:

$$\text{cov}\{x(u), x(v)\} \approx r_M(u, v) + r_M(u, -v) + r_M(u, 2L - v)
 \tag{24}$$

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

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

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

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

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

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

**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  $\Omega$  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  $\phi$  denote a function  $\phi: \Omega \mapsto \mathbb{R}$ . The *gradient* of  $\phi$  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  $\nabla$  is formally given by the column vector  $(\partial/\partial u_1, \dots, \partial/\partial u_d)^T$ . The *Laplacian*  $\Delta$  of  $\phi$  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  $\Delta = \partial^2/\partial u_1^2 + \dots + \partial^2/\partial u_d^2$ . At the boundary of  $\Omega$ , 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  $\phi$  is the directional derivative  $\partial_n\phi(\mathbf{u}) = \mathbf{n}_\partial(\mathbf{u}) \cdot \nabla\phi(\mathbf{u})$ .

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  $H_\Omega^d(\cdot)$ . 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  $|A|_\Omega = H_\Omega^d(\mathbf{1}_A)$ , and the Hausdorff integral of a (measurable) function  $\phi$  as  $H_\Omega^d(\phi)$ . An inner product between scalar- or vector-valued functions  $\phi$  and  $\psi$  is defined through

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

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

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

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

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

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

If  $\Omega$  is a compact manifold, a countable subset  $\{E_k, k = 0, 1, 2, \dots\}$  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

$$\begin{aligned} \hat{\phi}(k) &= (\mathcal{F}\phi)(k) = \langle \phi, E_k \rangle_\Omega, \\ \phi(\mathbf{u}) &= (\mathcal{F}^{-1}\hat{\phi})(\mathbf{u}) = \sum_{k=0}^{\infty} \hat{\phi}(k) E_k(\mathbf{u}). \end{aligned}$$

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  $\kappa \geq 0$ , 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  $\|\phi\|_{\mathcal{H}^1(\Omega, \kappa)} = \langle \phi, \phi \rangle_{\mathcal{H}^1(\Omega, \kappa)}^{1/2}$ . 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  $\Omega$ . 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  $\Omega$  is a GF if  $\{x(\mathbf{u}_k), k = 1, \dots, n\}$  are jointly Gaussian random vectors for every finite set of points  $\{\mathbf{u}_k \in \Omega, k = 1, \dots, n\}$ . If there is a constant  $b \geq 0$  such that  $E\{x(\mathbf{u})^2\} \leq b$  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  $\Omega$ ,  $x$  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  $\Omega$  is a random  $L^2(\Omega)$  generalized function such that, for every finite set of test functions  $\{\phi_i \in L^2(\Omega), i = 1, \dots, n\}$ , the inner products  $\langle \phi_i, x \rangle_{\Omega}, i = 1, \dots, n$ , are jointly Gaussian. If there is a constant  $b \geq 0$  such that  $E(\langle \phi, x \rangle_{\Omega}^2) \leq b \|\phi\|_{\Omega}^2$  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  $\Omega$  is an  $L^2(\Omega)$ -bounded generalized GF such that, for any set of test functions  $\{\phi_i \in L^2(\Omega), i = 1, \dots, n\}$ , the integrals  $\langle \phi_i, \mathcal{W} \rangle_{\Omega}, i = 1, \dots, n$ , are jointly Gaussian, with expectation and covariance measures given by

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

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  $\text{cov}\{\mathcal{W}^*(A), \mathcal{W}^*(B)\} = |A \cap B|_{\Omega}$ . 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  $C^\infty$ , 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  $\Delta x$  is  $L^2(\Omega)$  bounded, then (with probability 1)

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

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

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

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  $(\kappa^2 - \Delta)^{\alpha/2}$ ,  $\kappa, \alpha \geq 0$ , 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  $\Omega$ , where  $\lambda_k, k = 0, 1, 2, \dots$ , 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  $\phi$  and  $\psi$  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)  $\Omega$  is closed or
- (c)  $\Omega$  is compact and  $\langle \phi, \partial_n \psi \rangle_{\partial\Omega} = \langle \partial_n \phi, \psi \rangle_{\partial\Omega} = 0$ .

For a proof, see Appendix D.2. Lemma 2 shows that, for functions  $\psi$  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  $\nabla x$ , 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  $\mathcal{H}_n^l(\Omega, \kappa) \subset \mathcal{H}^l(\Omega, \kappa)$  is spanned by a finite set of basis functions  $\Psi_n = \{\psi_1, \dots, \psi_n\}$ . We say that a sequence of subspaces  $\{\mathcal{H}_n^l\}$  is dense in  $\mathcal{H}^l$  if for every  $f \in \mathcal{H}^l$  there is a sequence  $\{f_n\}$ ,  $f_n \in \mathcal{H}_n^l$ , such that  $\lim_{n \rightarrow \infty} (\|f - f_n\|_{\mathcal{H}^l(\Omega, \kappa)}) = 0$ .

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

*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  $f, g \in L^2(\Omega)$ ,

$$\begin{aligned} E(\langle f, x_n \rangle_\Omega) &\rightarrow E(\langle f, x \rangle_\Omega), \\ \text{cov}(\langle f, x_n \rangle_\Omega, \langle g, x_n \rangle_\Omega) &\rightarrow \text{cov}(\langle f, x \rangle_\Omega, \langle g, x \rangle_\Omega), \end{aligned}$$

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

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

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

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

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

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

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

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

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

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  $\alpha$ -values, be unravelled into an iterative formulation

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

For integers  $\alpha = 1, 2, 3, \dots$ ,  $y_\alpha$  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  $\Omega$ . 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  $\alpha \geq 3$  also converge (theorem 4). A sequence  $\mathcal{H}_n^1(\Omega, \kappa)$  of piecewise linear Hilbert spaces defined on non-degenerate triangulations of  $\Omega$  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  $x_n = \sum_k \psi_k w_k$  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  $\partial_n \psi_k = 0$  on  $\partial\Omega$ .

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

*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  $\Omega$ , and let  $x_n$  be a weak  $\mathcal{H}_n^1(\Omega, \kappa)$  approximation, when  $\mathcal{W}$  is Gaussian white noise. Then,

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

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

if the sequence  $\{\mathcal{H}_n^1(\Omega, \kappa), n \rightarrow \infty\}$  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  $\Omega$ , for some  $L^2(\Omega)$ -bounded random field  $\mathcal{E}$ , and let  $x$  be a weak solution to the SPDE  $\mathcal{L}_x x = \mathcal{E}$ , where  $\mathcal{L} = \kappa^2 - \Delta$ . Further, let  $y_n$  be a weak  $\mathcal{H}_n^1(\Omega, \kappa)$  approximation to  $y$  such that

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

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  $\Omega$ . Then,

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

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

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  $\alpha$  or more general manifolds, the intrinsicness will only be approximately represented. How to construct models with more fine-tuned control of the null space is a subject for further research.

To approximate intrinsic fields with  $\alpha \geq 2$  and free boundaries, the matrix  $\mathbf{K}$  in theorem 2 should be replaced by  $\mathbf{G} - \mathbf{B}$  (owing to Green’s identity), where the elements of the (possibly asymmetric) boundary integral matrix  $\mathbf{B}$  are given by  $B_{i,j} = \langle \psi_i, \partial_n \psi_j \rangle_{\partial\Omega}$ . 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

$$\{h_1 + ih_2 - \nabla \cdot (\mathbf{H}_1 + i\mathbf{H}_2) \nabla\} \{x_1(\mathbf{u}) + ix_2(\mathbf{u})\} = \mathcal{E}_1(\mathbf{u}) + i\mathcal{E}_2(\mathbf{u}). \tag{34}$$

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  $\{\psi_k, k = 1, \dots, n\}$  of basis functions for  $\mathcal{H}_n^1(\Omega, \kappa)$  and construct a basis for the solution space for  $(x_1 \ x_2)^T$  as

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

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

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

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

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

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 makes 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  $\alpha \geq 2$ , 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  $\tilde{C}_{ii} = \sum_j C_{ij} = \langle \psi_i, 1 \rangle_\Omega$ , and note that this preserves the interpretation of the matrix as an integration matrix. Substituting  $\mathbf{C}^{-1}$  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  $\mathcal{H}_n^1(\Omega, \kappa)$ , 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

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

Requiring  $\|f\|_{\mathcal{H}^1(\Omega, \kappa)}, \|g\|_{\mathcal{H}^1(\Omega, \kappa)} \leq 1$ , it follows from lemma 1 in Chen and Thomée (1985) that the covariance error is bounded by  $ch^2$ , where  $c$  is some constant and  $h$  is the diameter of the largest circle that can be inscribed in a triangle of the triangulation. This shows that the convergence rate from 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  $\kappa$  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,

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

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  $\cos(u\pi k/L)$ ,  $k = 0, 1, 2, \dots$

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

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

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

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

Rewriting the cosines via Euler’s formulae, we obtain

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

where we used the Dirac measure representation

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

Finally, combining the results yields

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

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  $\lambda_k \geq 0, k = 0, 1, 2, \dots$ , 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,

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

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

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

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

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

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

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

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

owing to Green's identity, and

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

This covariance is equal to

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

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

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

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

so  $\mathbf{Q} = \mathbf{K}^T \mathbf{K}^{-1} \mathbf{K} = \mathbf{K}$ , 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),

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

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

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

Requiring equality for all pairs of test functions yields  $\mathbf{Q} = \mathbf{K}^\top \mathbf{C}^{-1} \mathbf{Q}_{\mathcal{E},n} \mathbf{C}^{-1} \mathbf{K}$ . 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  $\tilde{f}$  be the solution to the PDE

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

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

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

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

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

and similarly for the least squares case (b).

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

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

as  $n \rightarrow \infty$ . 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

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

as  $n \rightarrow \infty$ .

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

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

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

## References

- Adler, R. J. (2010) *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. Imngng*, **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. *Mathematical Sciences Preprint 2009:13*. Lund University, Lund.
- Bolin, D. and Lindgren, F. (2011) Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Ann. Appl. Statist.*, **5**, 523–550.
- 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. *Computnl 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. *Statist. Pap.*, **49**, 531–552.
- 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 modeling. 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.
- Hrafinkelsson, 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 covariance models for global data. *Ann. Appl. Statist.*, **2**, 1271–1289.

- Karypis, G. and Kumar, V. (1999) A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J. Scient. Comput.*, **20**, 359–392.
- 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 data-base. *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 Marichev, 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 comparative 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. Planning 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.

## Discussion on the paper by Lindgren, Rue and Lindström

John T. Kent (*University of Leeds*)

This paper uses finite element methods to give a principled construction of Matérn-type Gaussian models in a variety of spatial settings. A key property of such models is that they have a sparse inverse covariance or precision matrix. The paper gives a comprehensive treatment of these models and it seems destined to become a landmark paper in spatial analysis. In many ways the paper is a natural sequel to Besag's (1974) paper that was read to the Society, and it forms a fitting tribute to his memory.

In earlier work often a great distinction was made between conditional auto-regression (CAR) and simultaneous auto-regression (SAR) models. In terms of a zero-mean Gaussian process  $\{x_{ij}\}$  on the integer lattice in  $\mathbb{R}^2$  and the notation  $A(x)_{ij} = (x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1})/a$  for the shrunken first-order neighbourhood average ( $a > 4$ ), the one-parameter versions are

$$E(x_{ij}|x_{-i,j}) = A(x)_{ij},$$

$$\text{var}(x_{ij}) = \sigma_{\text{CAR}}^2,$$

(CAR) and

$$x_{ij} - A(x)_{ij} = \varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_{\text{SAR}}^2)$$

(SAR). In the first, the conditional distribution of  $x_{ij}$  given the values of the process at all the remaining sites depends only on the nearest neighbours; in the second a filtered version of the  $x$ -process equals (discrete) white noise. The second process is a discrete approximation to the stochastic partial differential equation (2) based on the differential operator  $D_{\kappa,\alpha} = (\kappa^2 - \Delta)^{\alpha/2}$  with  $\alpha = 2$ .

The difference between the two processes can be seen most clearly in the spectral domain. Setting  $\tilde{A}(\omega) = \{2 \cos(\omega_1) + 2 \cos(\omega_2)\}/a$  and  $g(\omega) = 1 - \tilde{A}(\omega)$  for  $\omega \in (-\pi, \pi)^2$ , the two spectral densities are

$$\begin{aligned} f_{\text{CAR}}(\omega) &\propto g(\omega)^{-1}, \\ f_{\text{SAR}}(\omega) &\propto g(\omega)^{-2}. \end{aligned}$$

By taking suitable convolutions, both models can be extended to higher order neighbourhoods as described in the paper and, letting the even integer  $\alpha \geq 2$  denote the order of neighbourhood, we have

$$\text{SAR}(\alpha/2) = \text{CAR}(\alpha) \approx M(\nu), \quad \nu = \alpha - d/2,$$

where the dimension is  $d = 2$  here and  $M(\nu)$  stands for the Matérn model of index  $\nu$ . In higher dimensions, I wonder to what extent we need to restrict  $\alpha$  to be sufficiently large that  $\nu > 0$ , or at least  $\nu \geq 0$ . One of the clever observations in the paper is to note that this SAR and CAR identification can be extended to odd integers  $\alpha \geq 1$  where, for the approximate weak solution of the stochastic partial differential equation, the finite element ideas from Section 2.3 are used.

The role of the null space of a differential operator is not clear to me. Consider the stationary process on all of  $\mathbb{R}^2$  generated by  $D_{\kappa,\alpha}$  with  $\kappa > 0$  and  $\alpha = 2$ . As noted in the paper, this differential operator has a null space which includes certain exponential functions. However, the random field is well defined even when specified to have mean 0. Further, if  $D_{\kappa,\alpha}$  is used to motivate a process on a finite domain using the finite element construction, it is not clear to me what happens to the null space.

The situation is more delicate for intrinsic processes (Kent and Mardia, 1994). Consider the self-similar process on  $\mathbb{R}^d$  generated by  $D_{0,\alpha}$ , which is intrinsic provided that the real parameter  $\alpha$  is sufficiently large that  $\nu = \alpha - d/2 > 0$ . The intrinsic order is  $p = [\nu]$ , where  $[\cdot]$  denotes the integer part. An intrinsic process is defined only up to the null space of polynomials of degree  $p$ . When  $\alpha$  is an even integer, the differential operator  $D_{0,\alpha}$  has a null space given by the polynomials of degree  $r = \alpha - 1$ , and this result extends to all integers  $\alpha \geq 1$ . For integer  $\alpha$  in dimensions  $d = 1, 2$ , it follows that  $p = r$ . However, if  $d \geq 3$ , then  $p < r$ . Thus, in the intrinsic setting, the question about the role of the null space of the differential operator also arises.

Recall that a differential operator  $D$  can be used to define a smoothing spline with penalty term  $\int (D\varphi)^2$  for sufficiently smooth functions  $\varphi$  on  $\mathbb{R}^d$ . For example,  $D_{0,2}$  leads to cubic and thin plate splines in dimensions  $d = 1, 2$ . As the paper continually emphasizes, this construction is computationally intensive for large  $n$  when  $d \geq 2$ . Hence it is natural to ask how successfully the finite element ideas of this paper can be used to yield a computationally efficient approximation.

The paper mentions briefly how deformations can be used to introduce non-stationarity in a real-valued Gaussian process. Gaussian models can also be used to construct deformations of Euclidean space. Bookstein (1989) suggested the use of thin plate splines, and Michael Miller and his colleagues (e.g. Christensen *et al.* (1996)) developed more sophisticated models using non-linear partial differential equations motivated by continuum mechanics. My former doctoral student, Godwin (2000), constructed deformations which were constrained by elastic penalties and discretized by using finite element methods. In this case a pair of interacting partial differential equations for the horizontal and vertical displacement is obtained, depending on two parameters called the Lamé coefficients; changing the ratio between them can have a dramatic effect on the fitted deformation.

In summary, I found this to be an extremely stimulating paper and it gives me great pleasure to propose the vote of thanks.

**Peter J. Diggle** (*Lancaster University*)

This paper is an important contribution to an important topic. Latent Gaussian fields are widely used as components of geostatistical models (Diggle *et al.*, 1998) and of point process models (Møller *et al.*, 1998). I believe that they should also be more widely used for the analysis of spatially discrete data. Consider,

for example, data consisting of counts  $Y_i$  associated with each of  $n$  subregions  $A_i$  that partition a region of interest,  $A$ . A standard class of models for data of this kind is that

$$Y_i|X \sim \text{Pois}(\mu_i), \quad \mu_i = \exp(\alpha + X_i), \tag{36}$$

where the  $X_i$  form a Markov random field in which

$$X_i|\{X_j: j \neq i\} \sim N(\bar{X}_i, \tau^2/n_i),$$

where  $\bar{X}_i$  is the average of the values of  $X_j$  associated with the  $n_i$  subregions  $A_j$  that are considered to be neighbours of  $A_i$ . The usual approach to defining the neighbours is through contiguities:  $A_i$  and  $A_j$  are neighbours if they share a common boundary. This is appealing in a regular geography; less so when the  $A_i$  vary substantially in size and shape. An alternative is to assume that

$$Y_i|X(\cdot) \sim \text{Pois}(\mu_i), \quad \mu_i = \int_{A_i} \exp\{\alpha + X(u)\} du, \tag{37}$$

where now  $X(\cdot)$  is a spatially continuous Gaussian field.

A pragmatic reason for preferring class (36) over class (37) has long been that the associated computations are much less burdensome. The remarkable computational efficiency that the methods in this paper achieve should instead allow a choice between classes (36) and (37) to be made on the basis of their merits as models.

Another important aspect of the paper is its delivery of appealing non-stationary constructions through the paper's equation (12). In this respect, it is a pity that the restriction to integer  $\alpha$  excludes the exponential correlation function ( $\kappa=0.5$ ) in the two-dimensional case. The popularity of the Matérn family of correlation functions stems from the fact that the integer part of  $\kappa$  corresponds to the mean-square differentiability of  $X(\cdot)$ . But  $\kappa$  is difficult to estimate, and a widely used strategy is to choose its value among a small number of qualitatively different candidates, say  $\kappa = 0.5, 1.5, 2.5$ , corresponding to continuous, differentiable and twice-differentiable fields  $X(\cdot)$  respectively.

The paper also gives a new slant on low rank approximations by representing the field  $X(\cdot)$  as

$$X(u) \approx \sum_{k=1}^n \psi_k(u) W_k, \quad W \sim \text{MVN}(0, \Sigma).$$

Here again, computational efficiency is crucial as the number of terms in the summation needs to be large when the range of the spatial correlation is small relative to the dimensions of the region of interest.

The paper uses integrated nested Laplace approximations as the basis for inference, building on Rue *et al.* (2009). For the applications that are considered in tonight's paper, the focus on marginal posteriors is a little too restrictive. If, as is often the case in applications, the focus of scientific interest is on predictive inference for one or more non-linear functionals of the unobserved realization of  $X(\cdot)$ , we need a reliable approximation to the joint predictive distribution of the whole field  $X(\cdot)$ , which is in practice approximated by its joint distribution on a fine grid spanning the region of interest.

The ability to fit models of this kind without resorting to Markov chain Monte Carlo methods is very welcome. My strong impression is that, for problems of this degree of complexity, the tuning and empirical assessment of convergence of Markov chain Monte Carlo algorithms remains something of a black art. However, is tuning still necessary for the methods that are proposed in the paper to deliver accurate inferences, and if so how delicate is this tuning?

The acute myeloid leukaemia (AML) data that are analysed in the paper consist of the residential locations and (censored) survival times of all recorded cases in an area of north-west England. The authors' analysis follows earlier published analyses in treating these as geostatistical data, implicitly assuming that the locations have been sampled independently of the survival process. But this may not be so—the data are a realization of a marked point process with carrier space  $\mathcal{N} \times \mathcal{R}^+$ , where  $\mathcal{N}$  is the set of all locations of people resident in the study region who are at risk of contracting AML. This does not invalidate the authors' analysis, which addresses the spatial variation in survival prognosis conditional on contraction of AML. But if the wider objective is to identify spatially varying factors that are involved in the aetiology of AML it potentially tells only half of the story, as there may be unrecognized spatially varying factors that affect both disease risk and survival prognosis; for a discussion of some of the methodological issues that are involved, see Diggle *et al.* (2010). For the AML data, although there is clear evidence of a marginal association between a simple circle counting estimate of the local density of cases and the hazard for survival ( $p < 0.001$  in a Cox proportional hazards analysis), this is accounted for by the authors' adjustments for sex, white blood cell count and deprivation ( $p = 0.349$ ). However, a thorough joint analysis of risk and

survival prognosis should also take into account the spatial variation in the local density of the population at risk.

It is with great pleasure that I second the vote of thanks for what will be, I am sure, a very influential paper.

The vote of thanks was passed by acclamation.

**J. B. Illian** (*St Andrews University*) and **D. P. Simpson** (*Norwegian University of Science and Technology, Trondheim*)

*Explicitly linking Gaussian fields and Gaussian Markov random fields—relevance for point process modelling*  
We congratulate the authors for this inspiring paper which we are sure will have a strong influence on spatial and spatiotemporal modelling for years to come. The stochastic partial differential equation approach provides an alternative representation for a large class of non-stationary Gaussian random-field models without needing explicitly to derive a covariance function. A piecewise linear Gaussian Markov random-field approximation is constructed that globally approximates the true random field up to a given resolution. This is a particularly interesting feature in the context of spatial point process modelling.



**Fig. 10.** Map indicating the structure of the Zackenberg valley and its boundaries

A log-Gaussian Cox process is a spatial point process that models the log-intensity field as a Gaussian random field defined *continuously* over the whole observation window. A common model fitting approach is to place a fine lattice over the observation window (Møller and Waagepetersen, 2007) and to count the number of points that are present in each grid box (Rue *et al.*, 2009; Illian and Rue, 2010; Illian *et al.*, 2011). This count value has a Poisson distribution and the latent spatial structure is modelled through a Gaussian Markov random field. Using this approach, the computational lattice has two roles: to estimate the latent Gaussian field, and to approximate the position of the points through a binning procedure. The first of these aims is entirely natural, whereas the second is an artefact of the approximation and leads to finer lattices than are really necessary. *This is wasteful.* As the stochastic partial differential equation approach constructs a *continuously indexed* random field, it is no longer necessary to bin the data over the lattice. It is then possible to construct a numerical approximation to the point process likelihood and to perform inference as normal. In particular, it is possible to perform extremely fast approximate inference for realistically complicated marked point process models by using the integrated nested Laplace approximations framework of Rue *et al.* (2009).

This approach begets a whole host of modelling questions. Consider the point pattern that was discussed in Illian and Hendrichsen (2010) and in Illian *et al.* (2011) consisting of the locations of muskoxen herds in Zackenberg valley in eastern Greenland; see Fig 10 for the location of the study area within Greenland and a map of the area. The observation window has a complicated boundary structure; it contains hard boundaries (the sea to the south), permeable boundaries (a river that the muskoxen may, reluctantly, cross) and artificial boundaries (the end of the observation window in the north). The stochastic partial differential equation approach allows us to incorporate this important information into our models. This is impossible with standard covariance-based models.

We thank Mads Forchhammer from Greenland Ecosystem Monitoring, and Ditte Hendrichsen for introducing us to the Zackenberg muskoxen data as an example of different boundary characteristics within a study site.

#### Tilmann Gneiting and Michael Scheuerer (*Universität Heidelberg*)

We congratulate Lindgren, Rue and Lindström on an exceptionally rich and original paper that builds bridges between statistics, probability, approximation theory, numerical analysis and applied fields, and opens up a wealth of new perspectives.

We share the authors' excitement about the ease with which the stochastic partial differential equation approach allows for the modelling of non-stationarity, both on Euclidean spaces and on manifolds. As the authors note, an appealing interpretation of the non-stationary model (12) is that of a locally stationary Matérn field, whereas the actual form of its correlation function is unknown. Would members of the locally stationary Matérn correlation function that was developed by Stein (2005) and Anderes and Stein (2011) be candidates?

We second the authors' call for 'more physics-based spatial modelling' (Section 5) with enthusiasm. In this context, Balgovind *et al.* (1983) derived a spatial statistical model for the errors in numerical weather prediction schemes from the physics of large-scale atmospheric flow. Their arguments lead to the stochastic dynamic differential equation (2.7) for the geopotential error field, which is of form (12) in this paper with  $\alpha = 2$ , where the  $\kappa(\mathbf{u})$  term varies smoothly with geographic latitude. This defines a spatial model on the sphere and so provides an applied instance where the manifold setting is essential, similarly to the global temperature example in the paper. On more general types of manifolds, approaches developed by approximation theorists may prove related, and useful; see, for example, Narcowich (1995).

Perhaps the strongest limitation of the authors' ingenious approach is the restriction to a small set of feasible values for the Matérn smoothness parameter (Section 2.3, remark (e)). In view of well-known sample path properties of Matérn fields (see Guttorp and Gneiting (2006), and references therein), this creates what could be called a 'roughness gap', particularly in  $\mathbb{R}^2$ , where smoothness parameters  $\nu = 1, 2, \dots$  allow for Gaussian fields with differentiable sample paths only. What value of  $\nu$  was used in the global temperature example, where the spatial domain is the sphere, and what are the implications?

In a way, the restriction just mentioned is natural, as any probabilistically principled approximation of Gaussian fields by discretely indexed Gaussian Markov random fields can be expected to yield Markov models in the continuum limit, which is indeed what happens, leading to processes with reciprocally polynomial spectral densities (Section 2.3, remark (f)) and the (intrinsic, generalized) de Wijs field (Appendix C.3).

**R. Furrer and E. Furrer** (*University of Zurich*) and **D. Nychka** (*National Center for Atmospheric Research, Boulder*)

The authors are to be congratulated for this timely paper describing a very useful and practical methodological approach and computational procedure. There are many statisticians who are still working with Gaussian random fields simply because it is ‘virtually’ trivial to embed non-stationarity in the covariance function. With the stochastic partial differential equation approach, non-stationary modelling using Gaussian Markov random fields has become straightforward as well and this will take much wind out of the so-called ‘big  $n$ ’ problem research.

We would like to point out a link from the presented work with recent developments concerning the asymptotics of the kriging predictor. It is well known that spline smoothing and kriging prediction are related in the sense that kriging prediction can be interpreted similarly to smoothing splines as a penalized minimization problem where the roughness penalty is determined by the covariance function of the underlying spatial process. For both settings the basic idea lies in representing the estimator or predictor as a weighted average of the observations by using a weight function, which is—in contrast with kernel methods—not known in closed form. On the basis of Nychka (1995), we have exploited an approach of how to approximate this weighting function by using its reproducing kernel properties. For the kriging predictor in a stationary setting this so-called equivalent kernel is shown to have a simple form in terms of the Fourier transform of the covariance function of the process. The equivalent kernel can also be characterized as the Green function of a pseudodifferential operator, which is closely related to the one of the paper.

In our research we aim to use the equivalent kernel to analyse the asymptotic behaviour of the kriging predictor in terms of bias and variance (Furrer and Nychka, 2007; Furrer, 2008). For this we need to show that it satisfies the so-called exponential envelope condition, which is accomplished for fractional smoothing parameters  $\nu$  by using complex analysis (Furrer *et al.*, 2011). It is our hope that the use of reproducing kernel methodology and of techniques from complex analysis, especially in the case of fractional values of  $\alpha$ , could also be beneficial to the presented work.

On a second note, we were surprised to see that the range parameter of the fitted climate field was smallest around the equator and increased towards both poles. Although the observational data cannot be directly compared with general circulation model output, the latter usually exhibits the inverse behaviour. This is partially due to larger ocean proportions in the mid-latitudes, compared with the high latitudes in the northern hemisphere and due to sea ice components around the pole.

**Paul Fearnhead** (*Lancaster University*)

I congratulate the authors on their principled approach and elegant (yet easily implementable) solution to the important computational bottleneck of inverting covariance matrices that arises in spatial statistics. I have two comments, one relating to experience of constructing their Gaussian Markov random-field (GMRF) approximations in one dimension, and the other to the restrictions on the covariance models for which GMRF approximations can be calculated.

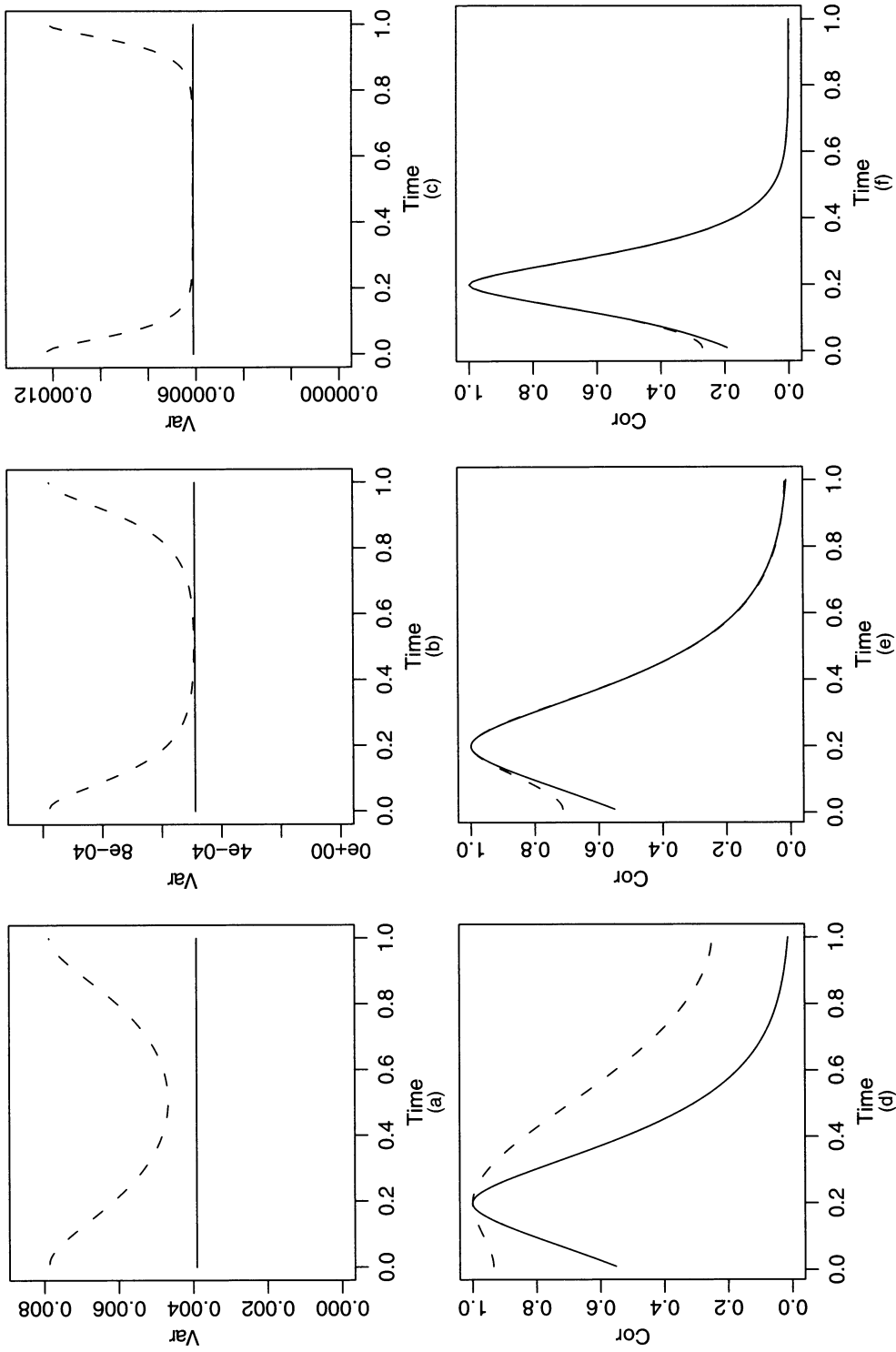
To see how easy it is to construct the GMRF approximation, and how accurate it is, I ran some experiments for approximating one-dimensional Gaussian processes. Firstly, constructing the approximation was relatively simple to implement, and computationally cheap. (I imagine that the main difficulty in higher dimensions will come from needing to construct the triangulation.) In terms of accuracy, results for  $\nu = \frac{3}{2}$  and three different values of  $\rho = \sqrt{(8\nu)/\kappa}$  are shown in Fig. 11.

The covariance structure of the GMRF approximation varied little with the number of basis vectors used, and Fig. 11 shows the result by using 100 basis vectors. The key observation is that the GMRF approximation is excellent except for within a distance  $\rho$  of the boundary of the interval considered; this is particularly noticeable for the marginal variance. A question to the authors is how should the GMRF approximation be constructed in practice to avoid these issues of deteriorating accuracy at the boundary? One simple approach would be to construct the GMRF approximation over an interval that extends a distance  $\rho$  beyond the data.

Secondly, a comment on the generality of the method: Stein (1999) argued for the use of the Matérn covariance model on the basis of its flexibility. The argument is based on the ability to vary  $\nu$  to allow for different degrees of smoothness in the underlying spatial model. So, is the restriction of the authors’ approach to integer values of  $\nu + d/2$  a practically important restriction on the flexibility of models that can be fitted to the data?

One way to extend the class of covariance models that can be used is to consider a spatial process  $X(\mathbf{u})$ , defined as

$$X(\mathbf{u}) = Y(\mathbf{u}) + Z(\mathbf{u}),$$



**Fig. 11.** Accuracy of GMRF approximation for  $\nu = \frac{2}{3}$  (all results are for using 100 basis functions in the GMRF approximation): (a)–(c) variance of the true Gaussian process (—), and the GMRF approximation (---) as a function of time; (d)–(f) correlation of the Gaussian process (—) and the GMRF approximation (---) at different time points with its value at time 0.2; (a), (d)  $\rho = 1$ ; (b), (e)  $\rho = \frac{1}{2}$ ; (c), (f)  $\rho = \frac{1}{4}$



where  $Y(\mathbf{u})$  and  $Z(\mathbf{u})$  are independent Gaussian fields with different Matérn covariance functions (but each restricted to integer values of  $\nu + d/2$ ). It is possible to analyse such models by using Markov chain Monte Carlo methods (or even integrated nested Laplace approximations), in such a way that you need only to invert the covariance matrices of  $Y(\mathbf{u})$  or  $Z(\mathbf{u})$ , but not  $X(\mathbf{u})$ . Hence the GMRF approximations could be used to do the needed calculations efficiently.

**Peter Challenor, Yiannis Andrianakis and Gemma Stephenson** (*National Oceanography Centre, Southampton*)

We congratulate the authors on what we believe will be a significant paper. Our own application of Gaussian fields is in the statistical analysis of computer experiments (see for example Kennedy and O'Hagan (2001)). Large computer simulators are used in many areas of science. In essence we follow the following procedure.

- (a) We carry out a designed experiment, running the computer simulator as few times as possible to span its input space.
- (b) What is known as an emulator is built by fitting a Gaussian field to the results of this experiment, linking the simulator inputs to its outputs.
- (c) A second experiment is often used for diagnostic purposes to check that we have a good emulator.
- (d) The emulator is used to make inferences about the simulator.

For further details see [www.mucm.ac.uk](http://www.mucm.ac.uk). Although the methods used are very similar to the analysis of spatial data we work in a much larger number of dimensions. We have one dimension for each uncertain parameter in the computer code being analysed. Challenor *et al.* (2010) examined a climate simulator with 16 uncertain input parameters. However, because runs of the computer simulator are often expensive we have a limited number of data points from a designed experiment: usually about 10 per dimension (Loeppky *et al.*, 2009). This means that we have a different big  $n$  problem. Unlike the geostatistics problem we have a relatively small number of data points but a large dimension—a big  $d$  problem rather than a big  $n$ . The framework presented here is very attractive because of the ease of including, for example, non-stationarity through the stochastic partial differential equation formulation, it can handle data on manifolds and the theory is appropriate for  $R^d$ . We are not sure that the implementation in many dimensions will be as simple as for  $R^1$  or  $R^2$ . The triangulation will be complex and since we often work in a sequential way modifying the emulator as additional experiments are carried out. Since our experiments are designed we might be able to anticipate the positioning of our future data points when we build the triangulation and make it part of the design process. One possibility we are considering is whether a more complex choice of basis function than piecewise linear would allow a simpler triangulation and, at the expense of computational complexity, gain in the set-up costs.

**Jesper Møller** (*Aalborg University*)

This important and impressive paper by Lindgren, Rue and Lindström provides a computationally feasible approach for large spatial data sets analysed by a hierarchical Bayesian model. It involves a latent Gaussian field, a parameter  $\theta$  of low dimension, Gaussian Markov random-field approximations to stochastic partial differential equations for covariance functions, and the integrated nested Laplace approximations package for the computations instead of time-consuming Markov chain Monte Carlo methods. Other recent papers by the authors (Simpson *et al.*, 2010; Bolin and Lindgren, 2011a) compare the approach in this paper with kernel convolution methods (process convolution approaches) and covariance tapering methods, and conclude that the Gaussian Markov random-field approximation to stochastic partial differential equations is superior.

The Matérn covariance function plays a key role, where for example in the planar case the authors assume that the shape parameter  $\nu$  is a non-negative integer when considering the stochastic partial differential equation. Is there some link to the fact that this stationary planar covariance function is proportional to the mixture density of a zero-mean radially symmetric bivariate normal distribution  $N_2(0, W I_2)$  with the variance  $W$  following a  $(\nu + 1)$ -times convolution of an exponential distribution?

Despite its popularity and flexibility for modelling different degrees of smoothness, is this three-parameter Matérn covariance function really sufficiently flexible for modelling large spatial data sets? Would a flexible non-parametric Bayesian approach be more appropriate for 'huge' spatial data sets, although this of course may be computationally slow? The dimension of  $\theta$  may then be expected to be so high that integrated nested Laplace approximations (Rue *et al.*, 2009) would not work; as the dimension of  $\theta$  may even be varying, a reversible jump Markov chain Monte Carlo method (Green, 1995) may be needed when updating  $\theta$  from its full conditional. When updating the Gaussian field from its full conditional (corresponding to a finite

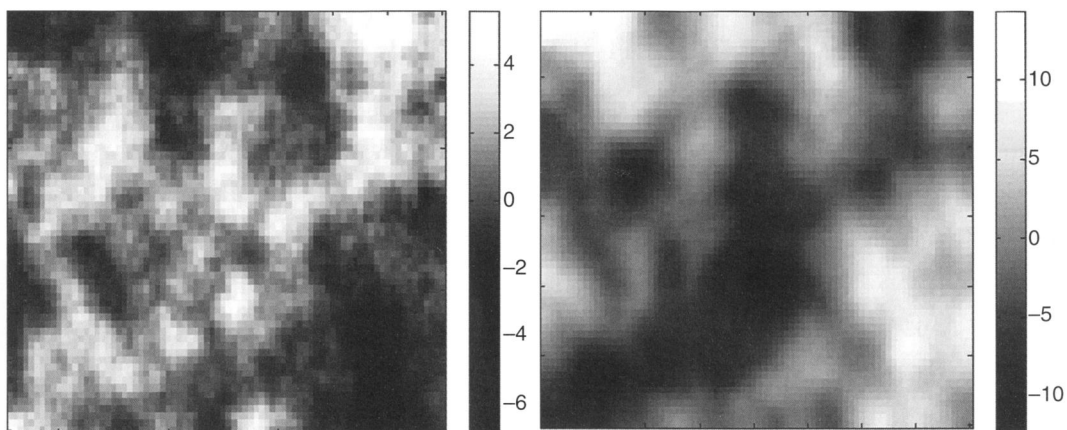


Fig. 12. Bivariate Matérn random field with negative correlation

set of locations), a Metropolis–Hastings algorithm may apply (Roberts and Tweedie, 1996; Møller and Waagepetersen, 2004).

The authors do not discuss model checking. The integrated nested Laplace approximation provides quick estimates of the marginal posterior distributions of the Gaussian field and of  $\theta$ . For model checking based on the joint posterior distribution, e.g. when comparing the data with simulations from the posterior predictive distribution, I presume that Markov chain Monte Carlo algorithms still are needed.

Finally, using a triangulation for a finite element representation of a Gaussian field is an appealing idea. For a spatial point pattern modelled by a log-Gaussian Cox process (Møller *et al.*, 1998), I expect that a regular triangulation would be used, since both the point pattern and the ‘empty space’ provide important information.

**Xiangping Hu and Daniel Simpson** (*Norwegian University of Science and Technology, Trondheim*)

We congratulate the authors on their excellent contribution to practical spatial statistics. We are particularly excited about the *local* specification of these random fields, which is markedly different from the constructs that are used for traditional spatial models. This stochastic partial differential equation specification is particularly useful in that it avoids any considerations of positive definiteness—if a solution exists it is automatically a Gaussian random field with a valid covariance function. Following from the comment in Section 3.3 of the paper, we have been investigating the extension of these methods to *multivariate* Gaussian random fields.

The construction of valid cross-covariance functions for multivariate Gaussian random fields is a very difficult problem and thus far only very specialized methods exist (see Gelfand and Banerjee (2010), and sources cited therein). The stochastic partial differential equation specification, however, *automatically* constructs valid cross-covariance functions! Inspired by the multivariate Matérn fields constructed by Gneiting *et al.* (2010), we define our multivariate random field  $\mathbf{x}(s) = (x_1(s), x_2(s))^T$  as the solution to

$$b_{11}(\kappa_{11}^2 - \Delta)^{\alpha_{11}/2} x_1(s) + b_{12}(\kappa_{12}^2 - \Delta)^{\alpha_{12}/2} x_2(s) = f_1(s), \tag{38}$$

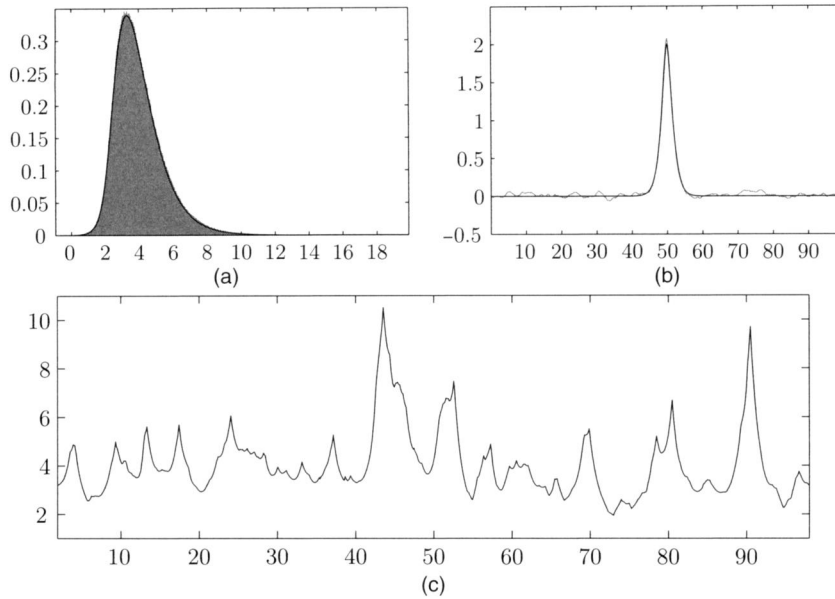
$$b_{22}(\kappa_{22}^2 - \Delta)^{\alpha_{22}/2} x_2(s) + b_{21}(\kappa_{21}^2 - \Delta)^{\alpha_{21}/2} x_1(s) = f_2(s), \tag{39}$$

where  $b_{ij} \in \mathbb{R}$ ,  $\kappa_{ij}^2 > 0$  and  $f_i(s)$  are independent (but not necessarily identical) noise processes. If we choose our exponents as  $\alpha_{ij} = 0, 2, 4, \dots$  and take the noise processes  $f_i(s)$  to be Markov, then following the procedure outlined in Section 2.3 we arrive at a bivariate *Markov* random field. A sample from a negatively correlated bivariate random field is given in Fig. 12.

We conclude this comment by noting that all the extensions mentioned in Section 3 of this excellent paper can be applied to this situation. In particular, we can construct non-stationary, spatiotemporal Gaussian random fields over fairly arbitrary manifolds. We plan to expand on this in future work.

**David Bolin** (*Lund University*)

The methods presented in the paper are indeed useful in a wide range of applications; however, if the Gaus-



**Fig. 13.** Simulation results using the SPDE formulation of a Laplace moving average model: (a) histogram of the samples from 1000 simulations together with the true density; (b) empirical covariance function for  $X(50)$  together with the true Matérn covariance; (c) simulation of the process on  $\mathbb{R}$

sianity assumption cannot be justified one cannot apply the methodology directly. Although the authors look only at Gaussian applications, the extension to certain non-Gaussian models is fairly straightforward.

A non-Gaussian model can be obtained by changing the Gaussian white noise to some other non-Gaussian process, i.e. define  $X$  as the solution to

$$(\kappa^2 - \Delta)^{\alpha/2} X(\mathbf{s}) = \dot{Z}(\mathbf{s}), \tag{40}$$

for some non-Gaussian process  $Z$ . What then differs in the method is the calculation of the elements on the right-hand side of the weak formulation of the stochastic partial differential equation, i.e. the integrals of the basis functions  $\varphi$  with respect to the noise,

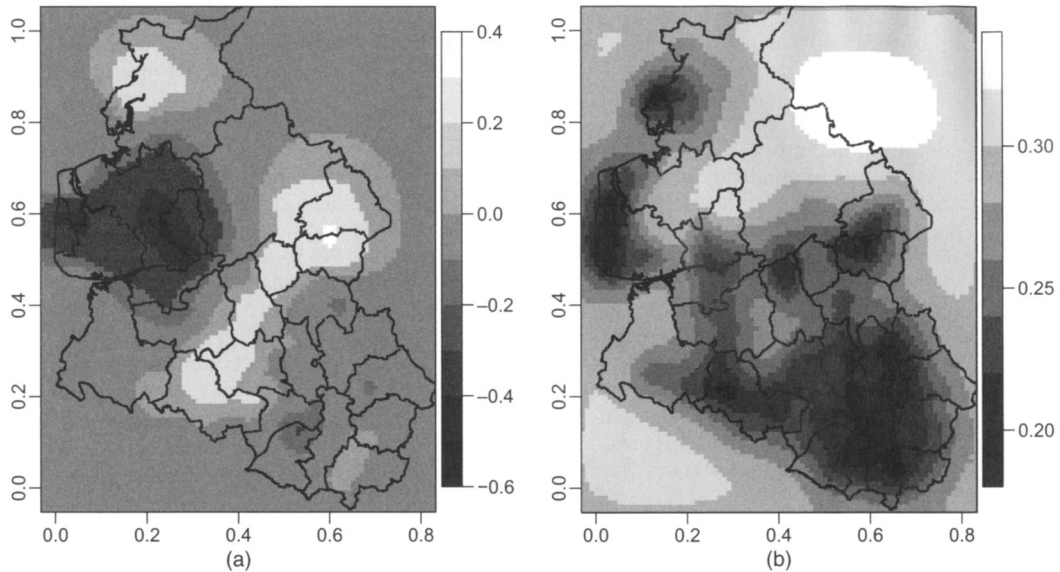
$$\int_{\Omega} \varphi(\mathbf{s}) Z(\mathbf{ds}). \tag{41}$$

An interesting class of non-Gaussian fields are the generalized asymmetric Laplace fields (Åberg *et al.*, 2009). If  $Z$  is of this type, integral (41) can be expressed as a Gaussian variable with mean  $\gamma \int_{\Omega} \varphi(\mathbf{s}) \mathbf{ds} + \mu \int_{\Omega} \varphi(\mathbf{s}) \Gamma(\mathbf{ds})$  and variance  $\sigma^2 \int_{\Omega} \varphi^2(\mathbf{s}) \Gamma(\mathbf{ds})$ , where  $\Gamma$  is a gamma process. Thus, the weak solution to equation (40) can be expressed as a GMRF conditionally on the integrals with respect to  $\Gamma$ . The solution to equation (40) can be viewed as a Laplace moving average model (Åberg *et al.*, 2009; Åberg and Podgórski, 2010) with a symmetric Matérn kernel  $f$ . The covariance function for  $X$  is Matérn and the marginal distribution for  $X(\mathbf{s})$  is given by the characteristic function

$$\varphi_X(u) = \exp \left[ \int_{\Omega} i\gamma f(t)u - \log \left\{ 1 - iu f(t) + \frac{\sigma^2 u^2}{2} f^2(t) \right\} dt \right]. \tag{42}$$

In Fig. 13, a simulation of the process on  $\mathbb{R}$  is shown when  $\mu = \gamma = \sigma = \kappa = 1$ . In Fig. 13 we can also observe close agreement between the empirical covariance function based on samples from 1000 simulations and the true Matérn covariance, as well as close similarity between the empirical density and the true density calculated by using numerical Fourier inversion of equation (42). This indicates that the method works for this non-Gaussian case.

Besides providing a computationally efficient method for simulation, an advantage with the stochastic partial differential equation formulation is that it simplifies parameter estimation, which usually must be done using the method of moments for these models (Wegener, 2010). However, the stochastic partial differential equation formulation facilitates estimation in a likelihood framework using the EM algorithm.



**Fig. 14.** Results of a GMRF approximation analysis of the leukaemia data, using a regular square grid assuming a piecewise constant random effect within grid cells: (a) posterior mean; (b) posterior standard deviation

Laplace processes are a special case of the Lévy processes of type G (see for example Wiktorsson (2002)), and the method extends to this larger class and possibly other non-Gaussian models as well. This is currently being investigated together with the parameter estimation problem for these models.

The following contributions were received in writing after the meeting.

**Patrick E. Brown** (*Cancer Care Ontario, Toronto, and University of Toronto*)

This is a very interesting paper which is certain to enable a wide range of spatial analyses which were previously intractable. Particular credit is due to the authors for making their excellent software available.

The interpretation of the Markov random-field approximation on the regular square grid could be slightly modified to allow for its use with irregularly spaced data. Rather than evaluating the surface  $X(s)$  only at a set of grid points  $s_{ij}$ , could we not use the  $x_{ij}$  to approximate the continuous  $X(s)$  by a surface which is piecewise constant within grid cells centred at  $s_{ij}$ ? This would be a special case of the method used for irregular data, with a regular square grid in place of irregular triangles and constant basis functions within cells in place of the linear basis functions. An important difference between the piecewise constant grid approximation and that used in the paper is that observations need not lie on the vertices of the grid.

Fig. 14 shows the regular grid, piecewise constant approximation fit to the leukaemia data by using the `R-inla` software. Grid cells are 1/100th the size of the  $y$ -dimension of the study region, and the grid is extended by 20 cells beyond the study region in each direction to avoid edge effects. The posterior means and standard deviations are nearly identical to those shown in Fig. 3 of the paper. The disadvantage of this analysis is that it appears to be more computationally demanding (taking roughly 10 min on an eight-core computer), though it is likely to scale quite well as increasing the number of observations would not increase the number of vertices on the lattice.

Can the authors offer advice on the choice of grid? How would we compare the quality of the approximation for a fine regular grid with piecewise constant bases to a much coarser triangular grid with linear basis functions? There is certainly a limit to roughness of a surface approximated by a coarse grid, and presumably a limit to how smooth a Markov surface on an extremely fine grid can be.

**Michela Cameletti** (*University of Bergamo*) and **Sara Martino** (*Norwegian University of Science and Technology, Trondheim*)

We congratulate the authors for this excellent paper that defines a link between Gaussian fields with Matérn covariance function and Gaussian Markov random fields. From our point of view, the stochastic partial

differential equation (SPDE) approach, combined with the integrated nested Laplace approximations algorithm proposed by Rue *et al.* (2009), introduces a new modelling strategy that is particularly useful for spatiotemporal geostatistical processes. The key point is that the spatiotemporal covariance function and the dense covariance matrix of a Gaussian field are substituted respectively by a neighbourhood structure and a sparse precision matrix, that together define a Gaussian Markov random field. In particular, the good computational properties of Gaussian Markov random fields and the computationally effective approximations of the integrated nested Laplace approximations algorithm make it possible to overcome the so-called ‘big  $n$  problem’. This issue refers to the infeasibility of linear algebra operations involving dense matrices and arises in many environmental fields where large spatiotemporal data sets are available.

The authors mention in Section 3.5 the possibility of extending the SPDE approach to non-separable spatiotemporal models. In this regard, we wonder whether it is possible to use the SPDE approach for approximating a spatiotemporal Gaussian field with a non-separable covariance function belonging to the general class defined by Gneiting (2002). As described in Cameletti *et al.* (2011) for air quality data, models characterized by these non-separable covariance functions are extremely computationally expensive because they involve matrices whose dimension is given by the number of data in space and time. Thus, parameter estimation and spatial prediction become infeasible by using Markov chain Monte Carlo methods. Moreover, such non-separable covariance functions are defined by a large number of parameters and the convergence when using Markov chain Monte Carlo methods can be an issue. Thus, if the SPDE approach could deal with the general class of non-separable covariance functions given in Gneiting (2002), it would be an important result for spatiotemporal geostatistical modelling.

**Daniel Cooley and Jennifer A. Hoeting** (*Colorado State University, Fort Collins*)

We congratulate the authors for establishing this important link between Gaussian fields (GFs) and Gaussian Markov random fields (GMRFs). To a large extent, GFs provide the foundation for geostatistics. Even when one restricts assumptions to only a mean and covariance function, a GF is not far removed from any geostatistical analysis since there is always a GF with these first- and second-order properties. For kriging the best linear unbiased predictor and the conditional expectation of a GF correspond and, if performing maximum likelihood estimation, the GF assumption is explicit.

Rather than dealing with point-located geostatistical data, GMRFs have their origin in modelling areal or lattice data. In comparison with geostatistical methods, statistical practice for areal data has seemed *ad hoc*. Whether assessing auto-correlation (e.g. Geary’s  $C$  or Moran’s  $I$ ) or constructing models (GMRFs or other auto-regressive models), areal data methods have relied on an adjacency matrix constructed from researchers’ *a priori* assumptions of spatial dependence. Modelling a GMRF often includes justifying a dependence structure described by a small number of parameters and constructing an adjacency matrix. The process of constructing a dependence structure from an adjacency matrix is typically heuristic at best and particularly difficult for irregular lattices. However, the computational advantages of GMRFs and other auto-regressive models usually outweighed the disadvantages. The authors’ link between GFs and GMRFs allows for the construction of a meaningful dependence structure in a GMRF setting for both regular and irregular lattices, allows for the use of GMRFs on point-located data and enables fast computation that has always been the advantage of GMRFs over GFs.

That the GF–GMRF link is made through a Matérn covariance function is important as the Matérn function is often defended as the most theoretically justifiable. In geostatistical practice, estimating the Matérn function’s smoothness parameter is challenging. Often the smoothness parameter  $\nu$  is fixed according to an *a priori* belief of the smoothness of process realizations. One can view the choice of template for a regular grid given in Section 2.2 as equivalent to setting the smoothness parameter at the outset, as is common practice.

A challenge in spatial statistics is that complex models require large sample sizes to estimate model parameters reliably (Irvine *et al.*, 2007), but many modelling procedures are too computationally complex for large sample sizes. The link between GFs and GMRFs will allow more researchers to investigate important statistical issues like model selection, precision of parameter estimates (especially spatial covariance parameters), sampling designs and more.

**Rosa M. Crujeiras and Andrés Prieto** (*University of Santiago de Compostela*)

We thank the authors for such an interesting paper, which forms a bridge between Gaussian fields and Gaussian Markov random fields by means of stochastic partial differential equations (SPDEs). The recent developments in numerical methods for PDEs may play an important role, by suitable modifications in the stochastic context. Our comments are focused on the procedure for solving the SPDE (2) by means

of finite element methods (FEMs), concerning the mesh construction, the FEM approximation and the boundary conditions imposed.

The construction of a triangulated lattice to approximate the Matérn field is required for constructing the FEM approximation. The authors point out that the vertices are placed on sample locations and a heuristic refinement is used to minimize the total number of triangles satisfying the mesh constraints. Such a generation is explicitly determined by the geometric requirements of sample locations and the geometry of the observation region. However, the mesh design could be driven additionally by the Matérn field itself, by means of an iterative procedure, which adapts the mesh to the behaviour of the SPDE solution minimizing an *a posteriori* error estimate. This is a well-known procedure in the engineering literature called *h*-adaptivity (see Ainsworth and Oden (2000) for further details).

Also in the FEM approximation, low order (piecewise linear) elements are used to approximate the Matérn field in the computational domain. These elements provide satisfactory approximations in the stationary case, but other choices are possible. For instance, discontinuous Galerkin methods (Hesthaven and Warburton, 2008), Petrov–Galerkin methods or finite volume methods (see LeVeque (2002)) are possible alternatives. In fact, finite volume methods may be more suitable in the non-separable space–time models, with a dominant transport term.

As the authors remarked, there is a boundary effect in the covariance approximation due to boundary conditions. In this approach, Neumann conditions are imposed on the boundary of the computational domain when solving the SPDE. Rue and Held (2005) showed that the effect of boundary conditions is negligible if the length of the computational domain is sufficiently large compared with the effective range of the covariance, which enables us to capture the variability of the process. Nevertheless, this embedding drawback due to the non-exact homogeneous Neumann boundary conditions (which are not satisfied by the Matérn class) could be overcome by settling other boundary values in the second-order elliptic problem: non-local DtN operators (Givoli, 1999), absorbing boundary conditions (Keller and Givoli, 1989) or perfectly matched layers (see Bermúdez *et al.* (2010)).

**Marco A. R. Ferreira** (*University of Missouri, Columbia*)

I congratulate Dr Lindgren and his colleagues for their valuable contribution to the area of spatial statistics modelling and computation. In addition, I commend the authors for making their methodology publicly available in the R-*inla* package.

Lindgren and colleagues build on a stochastic partial differential equation based explicit link between Gaussian fields and Gaussian Markov random fields to develop fast computations for Gaussian fields with Matérn covariance functions. In addition, the stochastic partial differential equation may be thought of as a generator process that produces Matérn Gaussian fields. Within this stochastic partial differential equation framework, Matérn fields on manifolds and non-stationary fields arise naturally.

However, in this work the authors consider the smoothness parameter  $\nu$  fixed and taking values only on the set of positive integers. This contrasts with one of the advantages of the use of the Matérn class of covariance functions: the smoothness parameter  $\nu$  may be estimated from the data. Thus, the use of the Matérn class allows the degree of mean-square differentiability of the process to be data adaptive. In the current work, this particular advantage of the Matérn class is lost in favour of fast computation.

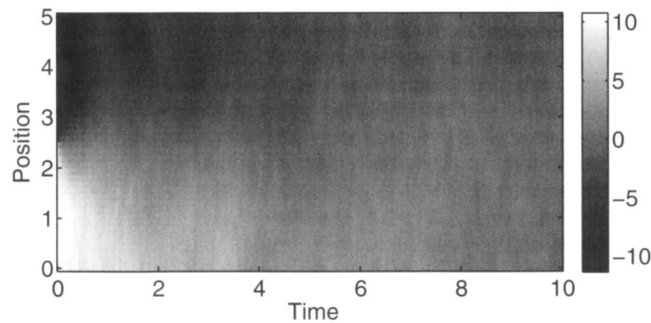
Finally, I am curious about what were the priors used by the authors for the hyperparameters. Even though the data sets considered are fairly large, some strange things may happen when one analyses spatial data. For example, improper uniform priors may lead to improper posterior distributions (Berger *et al.*, 2001).

**Geir-Arne Fuglstad** (*Norwegian University of Science and Technology, Trondheim*)

I congratulate the authors for their work on bringing together modelling with stochastic partial differential equations (SPDEs) and computations with efficient (sparse) Gaussian Markov random fields. This is an elegant way of creating (space–time) consistent models, which take advantage of the computational benefits of sparse precision matrices. I wish to comment on the brief statement at the end of Section 3.5 about how this method can be extended to a non-separable space–time SPDE. I used an approach that is similar to the one described there, but with a different spatial discretization. The steps that were used are described briefly in what follows.

Consider the non-separable space–time SPDE

$$\frac{\partial}{\partial t} u(x, t) - \frac{\partial^2}{\partial x^2} u(x, t) = \mathcal{W}(x, t), \quad (x, t) \in [0, A] \times [0, T], \quad (43)$$



**Fig. 15.** One realization of the SPDE (43), with  $A = 5$ ,  $T = 10$ , von Neumann boundary conditions  $\partial u / \partial x = 0$  on the spatial boundaries and the starting condition given in equation (44)

where  $\mathcal{W}$  is standardized Gaussian white noise and  $A, T > 0$  are constants. The spatial discretization is done by a finite volume method, and the Gaussian variable associated with each cell in the grid is allowed to be a Gaussian process in time. This reduces the space–time SPDE to a linear system of SDEs in time,

$$\frac{d}{dt}U(t) = C U(t) + D(t),$$

where  $U$  is the vector of Gaussian processes associated with the cells,  $C$  is a sparse matrix relating the derivative to the value at the cell itself and neighbouring cells and  $D$  is the vector of Gaussian white noise processes associated with each of the cells.

This system is then approximated with the backward Euler method. The use of the backward Euler method gives a Markov property in time, as each time step is only conditionally dependent on the closest time steps, thus still giving the desired speed-up compared with a dense precision matrix.

The results of this procedure proved quite good and the finite volume method extends to transport terms in a natural way. One simple example of using this procedure is the case with von Neumann boundary conditions at the spatial boundaries and a fixed starting state. Fig. 15 shows one simulation from such a situation with the starting condition

$$u(x, 0) = \begin{cases} 10, & 0 \leq x \leq 2.5, \\ -10, & 2.5 \leq x \leq 5. \end{cases} \quad (44)$$

Hence the approach briefly indicated in Section 3.5 does indeed seem to be a viable approach for space–time SPDEs.

**Andrew Gelman** (*Columbia University, New York*)

When using Bayesian inference (or, more generally, structured probability modelling) to obtain precise inferences from data, we have a corresponding responsibility to check the fit and the range of applicability of our models. Using expected values and simulated replicated data, we can graphically explore ways in which the model does not fit the data and broadly characterize particular inferences as interpolation or extrapolation. I am wondering whether the authors have considered using their powerful computational machinery to understand and explore the fit of their models. I think that graphical exploration and model checking could greatly increase the utility of these models in applications.

**Peter Guttorp** (*University of Washington, Seattle, and Norwegian Computing Center, Oslo*) and **Barnali Das** (*Westat, Rockville*)

This is a very important paper in spatial statistics. The late Julian Besag used to argue that the Markov random-field approach was better than the geostatistical approach, both from the computational and the conceptual point of view. The difficulty was always the requirement to have observations on a regular grid. Le and Zidek (2006) solved this by creating a grid containing the observations, and treating most of the grid points as missing, but then only a small fraction of the data are non-missing. The approach by Rue and his colleagues adapts the tessellation to the data and, in our opinion, proves that Julian once again was right.

The current approaches to estimating global mean temperature suffer from several difficulties. They are not truly global in character, they do not take into account the non-stationarity of the global temperature field and the covariance structure is local, not global. One of us (Das, 2000) developed an approach

to estimating non-stationary processes on the globe. The idea was based on the deformation approach (Sampson, 2010), by mapping the globe to a sphere, on which isotropic covariance functions are fully characterized (Obukhov, 1947). The implementation consisted of alternating transformations of latitude and longitude. Fig. 16 shows the resulting deformation, which expands the southern hemisphere (indicating smaller correlation length) and compresses the northern hemisphere.

Owing to the computational complexity of this fitting procedure, only 724 stations in the Global Historical Climate Network version 2 data set with a record of at least 40 years were used (115 randomly selected stations were reserved for model evaluation). Fig. 17 shows the resulting correlation estimates around three different points on the globe, showing a clear indication of inhomogeneity. The computational complexity of this fitting, not to mention spatial estimation of the global temperature field, has been reduced by several orders of magnitude in the work by Rue and his colleagues.

It appears that the temperature analysis by Rue and his colleagues has reasonable results; for example, the coefficient for regression on altitude is similar to what has been seen in the literature. We do wonder whether the estimates of the uncertainty of the various published global temperature series are similar to those computed by the new approach. Also, to what extent would inclusion of ocean temperature data change the estimates and their uncertainties?

**Ben Haaland** (*Duke–National University of Singapore Graduate Medical School*) and **David J. Nott** (*National University of Singapore*)

As the authors emphasize perhaps the greatest value of this paper lies in the non-stationary and other extensions their results provide. One potential application area is Gaussian field (GF) emulation of computer models. Modelling output of computer models as GFs is useful for constructing emulators, which are computationally cheap mimics of computer models providing appropriate descriptions of uncertainty (Santner *et al.*, 2003). The computer model involves inputs  $\lambda$ , and we have run the model at designed values  $\lambda_i$ ,  $i = 1, \dots, n$  say. There has been recent interest in dynamic emulation where the computer model output is a time series. Conti and O’Hagan (2010) discussed dynamic emulation with GFs. One approach treats time as an additional input, but for long time series this treats a large data set as an observation of a GF and we are in the realm of the ‘big  $n$  problem’. Correlations are often highly non-stationary in the input space and time. Although appropriate covariates in the mean sometimes help, currently used models do not seem sufficiently flexible. The methods provided have great potential, but we have some questions. First, the authors focus on the two-dimensional case and we wonder whether computational benefits decrease as the dimensionality increases. Secondly, we wonder in higher dimensions whether the boundary effects are more problematic.

Whereas the explicit representation discussed by the authors concerns positive integer  $\alpha$ , the representation has broad applicability. Draws from a GF with Matérn covariance are functions in a Sobolev space  $H^\alpha(\Omega)$  (Wendland, 2005). For integer  $\alpha$ , these functions have  $\alpha$  square integrable derivatives. The value of  $\alpha$  is typically selected so that the function estimate has sensible properties and fractional values are not ordinarily chosen. These function spaces are nested decreasing,  $H^\alpha(\Omega) \supset H^{\alpha+\varepsilon}(\Omega)$  for  $\varepsilon > 0$ . Hence, a

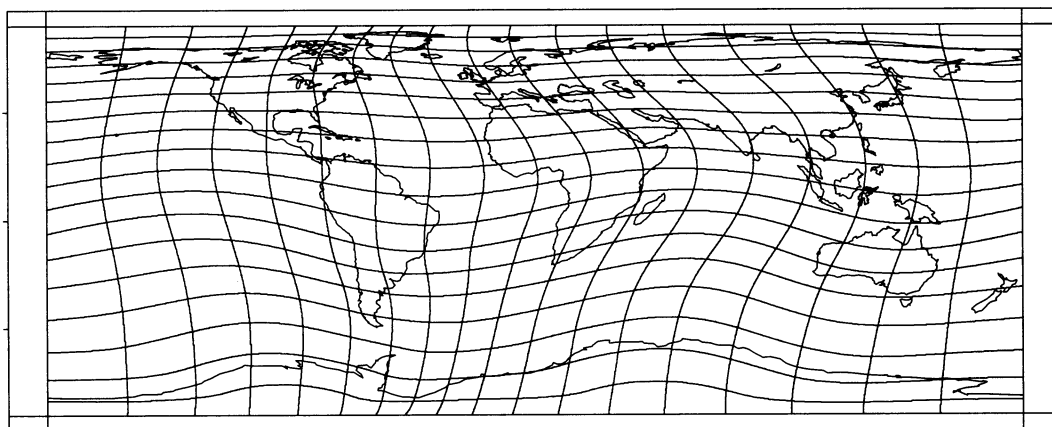
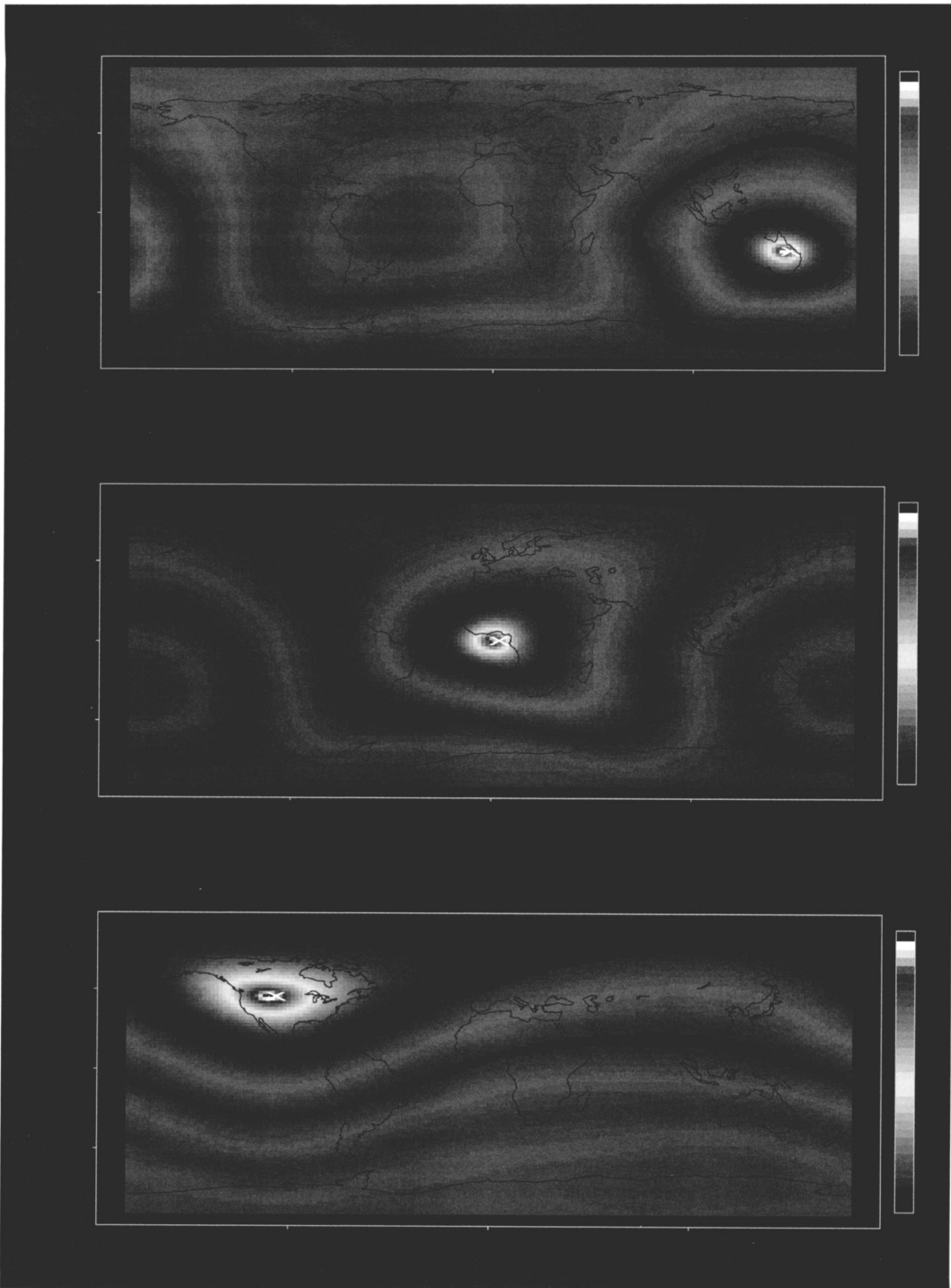


Fig. 16. Deformation of the globe (from Das (2000))





**Fig. 17.** Estimates of non-stationary spatial correlation around three points on the globe (Australia, West Africa and North America) (from Das (2000))

function with fractional smoothness  $\alpha > 1$  can be approximated by a function in  $H^{[\alpha]}(\Omega)$  and the convergence results hold. The infinitely smooth function space corresponding to the commonly used Gaussian covariance is contained in all Sobolev spaces. So, a Markov representation of a Matérn covariance with large smoothness could be used to approximate a Gaussian covariance. A balance must be found between better computational properties of less smoothness and better approximation properties of more smoothness. However, we wonder to what extent the convergence (theorems 3 and 4) and sparsity (Appendix C.5) results depend on using the typical finite element basis.

**John Haslett, Chaitanya Joshi and Vincent Garreta** (*Trinity College Dublin*)

We congratulate the authors on a most stimulating paper, and we see applications everywhere.

One application that concerns us at the moment is a multivariate space–time process in which the temporal changes can be much longer tailed than Gaussian. The simplest case is a univariate process at a single point in space. We have found that the normal inverse Gaussian distribution provides a flexible basis for temporal increments (see, for example, Barndorff-Nielsen and Halgreen (1977)). This model and its extensions are now widely used in finance (see, for example, Kumar *et al.* (2011)). Its multivariate extension (Wu *et al.*, 2009) provides a basis for the spatiotemporal version. The multivariate normal inverse Gaussian distribution can be motivated as a scale mixture of Gaussian distributions.

Impressed by the link between Gaussian fields and the Gaussian Markov random fields presented in this paper, we wonder whether it may be possible to extend the existing methodology to include non-Gaussian fields. If so, we suspect that one approach may be via *subordination*, which is one way of changing the distributional properties of a stochastic process.

Let  $\mathbf{X}(t)$  be a random process, and let  $\mathbf{T}(t)$  be a monotone, non-decreasing Lévy process; then the process  $\mathbf{X}\{\mathbf{T}(t)\}$  is said to be *subordinated* to the process  $\mathbf{X}(t)$  and  $\mathbf{T}(t)$  is called the *directing process*.

Subordination is an elegant way of producing both subdiffusive and superdiffusive motions from regular diffusive motions such as random walks and Brownian motion; see for example Sokolov (2000) and Eliazar and Klafter (2004). In particular, temporal subordination has been widely used to model systems whose subjective ‘operational time’ is different from the objective ‘physical time’, such as those often found in statistical physics (see for example Eliazar and Klafter (2005)) and econometrics (see for example Clark (1973) and Carr *et al.* (2003)). In particular, subordination has been shown to result in processes which are more *leptokurtic* (Clark, 1973).

We therefore wonder whether, by using a (spatially) subordinated Brownian motion  $W\{U(u)\}$ , it might be possible to modify the stochastic partial differential equation in equation (3)

$$(\kappa^2 - \Delta)^{\alpha/2} x(u) = W\{U(u)\}, \quad u \in \mathfrak{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0, \quad (45)$$

where the directing process  $U(u)$  is any suitable monotone, non-decreasing Lévy process, so that a (non-Gaussian) *weak* solution to this equation exists.

We would like to know the author’s views regarding this approach or any other approach leading to non-Gaussian fields.

**Michael Höhle** (*Robert Koch Institute, Berlin*)

The work presents innovative numerical solutions for geostatistical modelling. Despite the complexity of its mathematical detail, an open source implementation including an R interface is provided. This encourages potential users, including those without advanced mathematical knowledge, to apply the methods in practice. I congratulate the authors on their work regarding both the theoretical and the applied dimension.

As a comment, I miss explicit mention of how the methodology proposed can be used for inference in spatial point processes—specifically log-Gaussian Cox processes where Gaussian fields play an important role. In Rue *et al.* (2009), a proposal is given on how to use a Gaussian Markov random-field approximation in this case. It would therefore be of interest to see a comparison between this approximation and the current proposal. Furthermore, the leukaemia application in the paper provides the important link of using latent Gaussian fields in spatial regression modelling: it bridges between the joint posterior of the latent field and the hyperparameters as given by Rue *et al.* (2009), page 321. With specific application of Gaussian fields in spatiotemporal point process modelling in mind, how well does the integrated nested Laplace approximations approach allow for a likelihood where  $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$  is not simply a product of the individual observations anymore? Of particular interest would be whether the approach proposed is usable for adding Gaussian field random effects to regression models of the conditional intensity function, e.g. as in the stochastic epidemic model of Höhle (2009).

**Table 1.** Matérn inverse correlations—range = 30

Lag	Correlations for the following lags:				
	0	1	2	3	4
0	1.000	-0.510	0.187	-0.063	0.021
1	-0.510	0.167	-0.020	-0.005	0.005
2	0.187	-0.020	-0.011	0.006	-0.002
3	-0.063	-0.005	0.006	-0.001	0.000
4	0.021	0.005	-0.002	0.000	0.000

**Table 2.** Matérn inverse correlations—range = 100

Lag	Correlations for the following lags:				
	0	1	2	3	4
0	1.000	-0.512	-0.220	0.204	0.208
1	-0.512	0.179	0.062	0.072	0.045
2	-0.220	0.062	0.189	-0.047	-0.165
3	0.204	0.072	-0.047	-0.157	-0.142
4	0.208	0.045	-0.165	-0.142	0.054

**L. Ippoliti and R. J. Martin** (*University G. d'Annunzio Chieti–Pescara*) and **R. J. Bhansali** (*University of Liverpool*)

We congratulate the authors for developing a power Gaussian Markov random-field approximation to the correlation function of the Matérn class of Gaussian random fields, extending Besag's (1981) approximation using  $K_0$  to the completely symmetric conditional auto-regressive CAR(1) process, and Whittle's (1954) approximation using  $K_1$  to the completely symmetric simultaneous auto-regressive SAR(1) process.

We have two questions for the authors.

For large distances, both correlations (Matérn and approximating power CAR) are small, and the difference is small. From Abramowitz and Stegun (1965), equation (9.7.2),

$$z^\nu K_\nu(z) \sim cz^{\nu-0.5} \exp(-z).$$

Do the authors have any results on behaviour of the power CAR correlations for large distances?

The other question is on the interpolation. If the number of points to be interpolated is small the Gaussian Markov random-field specification has several advantages: compared with kriging, an inverse covariance-based predictor should be much quicker and more accurate, and avoid the common ill-conditioning problems of  $\Sigma$ . However, this assumes that the precision matrix of the Matérn function is well approximated by that of the power CAR with power  $\nu + 1$ . Consider the Matérn function with  $\nu = 2$ . Using a  $512 \times 512$  torus lattice Tables 1 and 2 show some low lag inverse correlations for ranges 30 and 100. The inverse correlations (the same values for both ranges to three decimal places) of the corresponding power CAR are given in Table 3.

Despite the good fit for the correlation function, the power CAR is not a good approximation to the inverse correlations within the CAR neighbourhood, and also the Matérn values can be appreciably different from 0 outside the neighbourhood. More importantly, the interpolation variance can be appreciably larger for the approximating power CAR—Table 4. Do the authors have any views on how much accuracy it is reasonable to sacrifice to obtain faster predictions?

Note that the Gaussian correlation structure, the limit as  $\nu \rightarrow \infty$  of the Matérn function, has an exact explicit form on the rectangular planar lattice for the inverse variance matrix, or for the finite CAR representation, which can be obtained from the results in Martin (2000).

**Table 3.** Inverse correlations of the completely symmetric power CAR

Lag	Inverse correlations for the following lags:				
	0	1	2	3	4
0	1.000	-0.508	0.107	-0.009	—
1	-0.508	0.214	-0.027	—	—
2	0.107	-0.027	0.000	—	—
3	-0.009	—	—	—	—
4	0.000	—	—	—	—

**Table 4.** Interpolation variance for Matérn and completely symmetric power CAR for two different values of  $\nu$  and ranges

$\nu$	Range	Interpolation variance, Matérn	Interpolation variance, CAR
1	30	0.0308	0.0498
1	100	0.0311	0.0500
2	30	0.0025	0.0089
2	70	0.0023	0.0089

Finally, some results comparing directly specified CAR models with integrated continuous processes over irregular areas are given in Martin and Dwyer (1994).

**Venkata K. Jandhyala and Stergios B. Fotopoulos** (*Washington State University, Pullman*)

We congratulate the authors for a timely and inspiring paper on a topic that is at the forefront of contemporary statistical modelling and computational statistics. The advantages of working with Gaussian Markov random fields have been well exploited for fitting hierarchical Bayesian models for spatial data. However, as pointed out by the authors, one often wishes to model data drawn from Gaussian fields (GFs). In the absence of the Markovian property, GFs are dense and are not easily amenable for computations. In this regard the authors have come up with a breakthrough result where they have demonstrated (with the necessary technical details) that any GF with a Matérn covariance function can be suitably replaced by an equivalent Gaussian Markov random field whose precision matrix is sparse. The depth of their result becomes evident by the general nature of spatial domains to which the result has been extended in the paper, e.g. to irregular lattices and manifolds. Thus, the authors have done a remarkable job.

We believe that the authors' results can be useful to address the change-point problem for spatiotemporal GFs, which is a problem that has been well formulated recently by Majumdar *et al.* (2005). Estimating the unknown change-point for retrospective data has been approached under both maximum likelihood estimation and Bayesian methods. Recent simulation studies performed by Fotopoulos *et al.* (2010) and Jandhyala *et al.* (2011) for estimating the change-point when changes occur in the mean vector alone or in the mean and/or covariance matrix of a multivariate Gaussian series have shown that the maximum likelihood estimate (MLE) performs marginally better than the Bayesian estimate, in most cases. Asymptotic distribution of the change-point MLE developed in both Fotopoulos *et al.* (2010) and Jandhyala *et al.* (2011) appear to be extendable to spatiotemporal Gaussian Markov random fields when changes occur in the elements of the precision matrix at some unknown time point. However, taking advantage of the main contribution of the present paper, it seems that asymptotic distribution of change-point MLEs can be extended to GFs with Matérn covariance functions. Thus, the results derived by the authors not

only are useful for Bayesian hierarchical model fitting but also seem to apply quite well to asymptotic distribution of the MLE in a change-point setting. One of the limitations of the change-point MLE in the treatment of Fotopoulos *et al.* (2010) and Jandhyala *et al.* (2011) is that its asymptotics require the assumption of independence over time. It seems that such an assumption about the MLE should extend to spatiotemporal GFs also, which may not be satisfactory, and this is where the Bayesian approach to change-point structure can be advantageous.

**Mikyong Jun** (*Texas A&M University, College Station*)

It is my honour to congratulate the authors on this excellent paper. Gaussian random-field models with a Matérn covariance function are popular, useful and commonly used in geostatistical applications, but the computation is problematic with large data sets. This paper provides a computationally efficient way for such a model through the Gaussian Markov random-field framework.

I have a few comments to add. My first comment is regarding the smoothness parameter  $\nu$ . As the authors acknowledge in Section 2.3, the current link between a Gaussian random field with Matérn covariance function and the Gaussian Markov random-field model is only possible for integer values of  $\alpha$  (and thus integer increments of  $\nu$ ), which, in my view, could be a little limited. Given that we cannot distinguish well  $\nu$ -values larger than 3 or so, this gives only two or three possible values of  $\nu$  for the application. I wonder how this might affect the model performance, especially parameter estimation and spatial prediction.

Regarding the above point, how is the estimation of the parameter  $\nu$  done? Is it fixed, or do you fit possible values separately and compare the fit? In Section 4.3, the authors mention that spatial covariance parameters are difficult to interpret individually (I am not sure whether I completely agree with this) and do not provide detail on how each parameter is estimated and what the estimates look like. I am also wondering, if you misspecify  $\nu$ -values, how this might affect the estimation of the rest of the parameters, and the spatial prediction.

I am particularly interested in the development for the processes on a sphere and non-stationary (and non-isotropic) models for the processes on a sphere. One of the limitations of the covariance models in Jun and Stein (2007, 2008) is that they are singular at the poles without some constraints. It appears that the approach suggested here may not suffer from this limitation. Is any constraint required for the model in this paper, to guarantee this? Do you need to have vertices at the poles or is any constraint required on the triangularization near the poles?

My last comment is that the third issue in Section 5 seems to me very interesting. Incorporating external covariates such as wind speed or wind direction in the covariance structure should be useful for environmental applications. However, as the authors point out, it may be difficult to do it in a way that the resulting covariance model is guaranteed to be valid. The method proposed may naturally provide ways to develop such physics-based covariance models.

**Håvard Wahl Kongsgård** (*Peace Research Institute, Oslo*)

I congratulate Lindgren, Rue and Lindström for an excellent paper and for the computational implementation. In conflict studies, microeconomics and related fields, spatial inquiries have become increasingly popular. For these fields, large data sets with multiple variables are often favoured.

However, most scholars nevertheless refrain from utilizing spatial regression as methods are often poorly computationally implemented and computer intensive. Consequently, spatial independence remains a major concern.

With great effect I recently applied the methods implemented in `R-inla` on a series of large point-based conflict data sets from Vietnam, Afghanistan and Iraq (Kongsgaard, 2011). When combined with spatial-temporal visualization, this type of analysis can be of value for tactical risk assessment, in a military or humanitarian capacity. Lindgren, Rue and Lindström's work is especially interesting as the new approaches make it possible to adjust for relative spatial distance.

Although the method introduced is an approximation with limitations, it is easy to use and can be applied in a streamlined fashion. Given rugged low level computational implementation and support for large sample sizes, the new functions will be a good supplement to the `R-inla` library. It is my opinion that these methods have great potential within fields of application.

**Giovanna Jona Lasinio** (*"Sapienza" University of Rome*) and **Alessio Pollice** (*"Aldo Moro" University of Bari*)

First, we congratulate the authors for their extremely interesting work that sheds new light on Gaussian

Markov random-field (GMRF) modelling. A connection is established with Matérn Gaussian fields through a weak solution of a linear fractional stochastic partial differential equation (SPDE) driven by a Gaussian white noise process. One of the main theoretical results is stated in theorems 3 and 4, where the authors prove the weak convergence of the finite Hilbert representation of the solution for positive integer values of the parameter  $\alpha$  to the continuous solution. Recently Bolin and Lindgren (2011a) have characterized a class of non-Gaussian random fields as the solution to a nested SPDE driven by Gaussian white noise. Could the authors comment on the possible further extension of their results to the case of SPDEs driven by more general non-Gaussian laws as stable or Lévy processes, in view of obtaining tools suitable for rates or concentrations, which do not typically follow a normal distribution? When  $d=2$ , the integer  $\alpha$  restriction of theorems 3 and 4 implies that also the parameter  $\nu$  is integer in the Matérn class specification. Within this class, the exponential covariance function, one of the most popular in many applied fields, corresponds to the value  $\nu=0.5$ . Is the exponential covariance model definitively excluded by the approach proposed? In the discussion the authors recognize that 'the approach comes with an implementation and preprocessing cost for setting up the models, as it involves the SPDE solution, triangulations and GMRF representation'. We would be interested in some more comments on the possible qualifications of this cost from an applied statistical point of view. A more detailed comparison of the GMRF solution to competing modelling approaches from a predictive perspective would also be of considerable interest. Here we mainly refer to low rank (Banerjee *et al.*, 2008; Cressie and Johannesson, 2008; Eidsvik *et al.*, 2010; Crainiceanu *et al.*, 2008) and process convolution methods (Higdon, 1998; Higdon *et al.*, 1999) when estimation is carried out with the same technique (either integrated nested Laplace approximation or Markov chain Monte Carlo methods).

Incidentally we think that the geographical interpretation of Fig. 6(b) would benefit from a more detailed description.

**Chihoon Lee** (*Colorado State University, Fort Collins*)

First of all, I congratulate the authors on their interesting and stimulating paper. Although this work is clearly an important contribution to the field of spatial statistics, its impact on the field of stochastic partial differential equations (SPDEs) cannot be overlooked as it ties in the probabilistic and mathematical treatment of SPDEs to spatial statistical modelling and geostatistics.

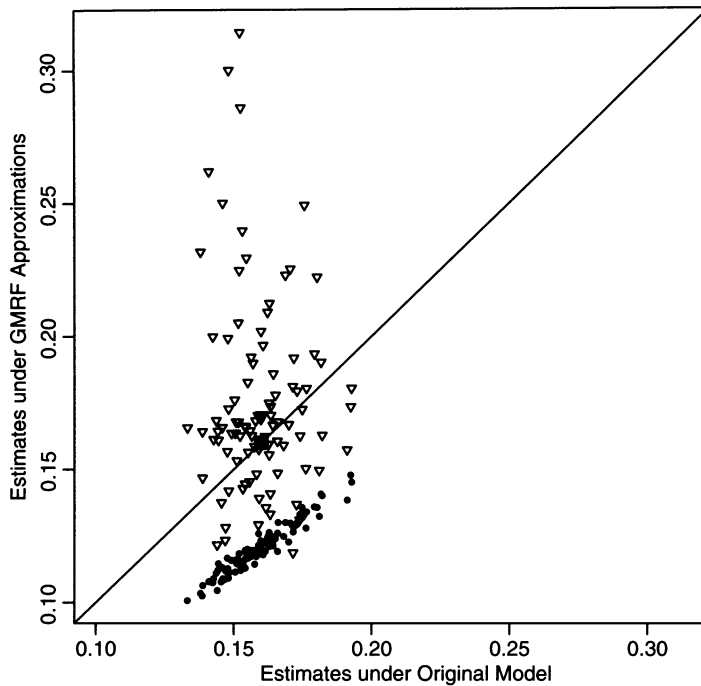
The authors' Gaussian Markov random-field construction via the SPDE link offers a natural definition of a Matérn field on manifolds with an intuitive interpretation. By employing a stochastic finite element approximation to the SPDE, practitioners can easily obtain finer resolution around any area of interest on the manifold. It seems promising that the authors' explicit mapping from the parameters of the Gaussian field model to the elements of a Gaussian Markov random-field precision matrix will yield new approaches to parameter estimation, for example, of a scaling parameter  $\kappa > 0$  of the Matérn field, which is of great importance in practice but generally difficult to analyse directly.

More precisely, one can construct an estimator (e.g. from maximum likelihood or the method of moments)  $\hat{\kappa}_n$  based on observations, using a finite element representation of the full solution to the SPDE as presented in expressions (9) and (10). The next step is to verify that  $\hat{\kappa}_n$  indeed converges (in an appropriate sense) to the true parameter  $\kappa$  as the finite Hilbert space, spanned by a finite set of basis functions  $\{\psi_1, \dots, \psi_n\}$ , approaches the full space. Such parameter estimation procedures circumvent the complexity that stems directly from dealing with the estimation problem of the fractional SPDE (2). Furthermore, this approach will shed light on estimating underlying physical parameters, corresponding to various covariates which could be incorporated into more physics-based SPDEs. For example, in a weather model, meteorological covariates such as wind speed or temperature may be incorporated (e.g. as appropriate drift terms) into underlying SPDEs. This would move beyond the purely statistical treatment of the SPDE (2) by combining theoretical (physical) models and statistical (data-driven) models.

**Wayne T. Lee and Cari G. Kaufman** (*University of California, Berkeley*)

We congratulate the authors on an important theoretical achievement with exciting computational implications. It is true that we tend to set  $n$  'a little higher than the value that gives a reasonable computation time'. We expect that 'reasonable' will now be redefined according to what is possible by using the authors' explicit link. For example, we simulated a Matérn-like field of size 750 000 (approximately the number of spatial grid boxes for a climate model with 25-km resolution) on a laptop in 8 s. Although a little slow for a long Markov chain Monte Carlo run, it is no longer completely unreasonable!

We would like to address the Gaussian Markov random field as an approximation in the context of likelihood-based estimation for geostatistical models. We hope that the authors can clarify the implications of



**Fig. 18.** Estimates for  $\sigma^2 \kappa^{2\nu}$  with effective range of 10 on a  $20 \times 20$  unit grid with  $\nu = 1$  and  $\sigma^2 = 2$  (the x-axis shows the maximum likelihood estimate, whereas the y-axis shows estimators maximizing two approximations to the likelihood):  $\nabla$ , naive implementation with no boundary correction;  $\bullet$ , implementation with embedding;  $\blacksquare$ , true parameter value

the treatment of the boundary points in this context. Although many will undoubtedly use these models in their non-stationary versions, non-stationarity that is an artefact of the boundary is undesirable and may introduce bias in estimating the geostatistical parameters.

We simulated 100 replications of a mean 0 Gaussian field on a  $20 \times 20$  unit grid, using a Matérn covariance function with  $\nu = 1$  and  $\sigma^2 = 2$ , and effective range 10 ( $\kappa \approx 0.283$ ). The large effective range highlights the boundary issue, but we think that high auto-correlation is also not uncommon in geostatistical data. We fixed  $\nu$  at its true value and considered three possible estimators for  $\sigma^2$  and  $\kappa$ . The first is the maximum likelihood estimator under the original model. The second maximizes an approximate likelihood that replaces the covariance matrix  $\Sigma(\sigma^2, \kappa)$  by  $\sigma^2 Q(\kappa)^{-1} / \sigma_M^2(\kappa)$ , where  $Q(k)$  is constructed by using the formulation from Section 2.2.1, with no special treatment of the boundary points, and  $\sigma_M^2(\kappa)$  is the marginal variance for the stochastic partial differential equation solution. The third treats the boundary points by using the embedding method suggested in Section 5.1.4 of Rue and Held (2005) under which we use the true model for a boundary set of thickness  $m = 2$  and the Gaussian Markov random-field approximation for the conditional distribution of the interior set. The drawback is the computational cost of calculating the boundary model.

Fig. 18 shows estimates of  $\sigma^2 \kappa^{2\nu}$ , the consistently estimable parameter under the fixed domain asymptotics (Zhang, 2004). We observe a clear bias in the embedding estimates, whereas the naive approximation estimates have less bias but have high variability. We have not implemented the authors' approach using the Neumann boundary conditions.

**Bo Li** (*Purdue University, West Lafayette*) and **Marc G. Genton** (*Texas A&M University, College Station*) In this very stimulating paper, the authors created a new path to deal with the computational challenge caused by large spatial data sets. Whereas most of the previous approaches mainly focused on screening out information that is relatively less important to gain computational efficiency (see Sun *et al.* (2011) for a recent review), this newly proposed method sought an explicit link between some Gaussian fields (GFs) and Gaussian Markov random fields (GMRFs) and thus enabled a direct application of the inherent

computational advantage in GMRFs to GFs. The GFs with Matérn covariance structure play a central role in spatial data modelling. Although the GMRF representation is developed only for the GFs with certain values of smoothness, we expect a wide application of this new approach since the smoothness parameter in the Matérn function is nevertheless difficult to estimate precisely. We genuinely appreciate the novelty and practical value of this paper. However, recently emerged data sets are often indexed by locations in both space and time, and many have more than one variable observed. Analyses with those data sets are more challenging owing to the cubic growth of computations in terms of the sample size. North *et al.* (2011) derived Matérn-like space–time correlations from evolving GFs governed by a white-noise-driven damped diffusion equation arising from simple energy balance climate models on the plane and on the sphere. It appears that those results could be used directly to extend the link between GFs and GMRFs to spatiotemporal data. Further extensions to a multivariate context remain open.

The authors gave an example of modelling non-stationary global temperature GFs and then making inference on the temperature process via GMRFs in conjunction with the integrated nested Laplace approximation. This can be very useful for the palaeoclimate community because one popular approach for large-scale palaeoclimate reconstructions is through Bayesian hierarchical models (Li *et al.*, 2010) where it is crucial to identify an appropriate model for the random process of climate variables. Such a model needs to be sufficiently flexible while still keeping the inference computationally feasible. The explicit link developed in this paper combined with the integrated nested Laplace approximation seems a promising direction. Since the proxy data used for the reconstruction carry various types of noise, a nugget effect may need to be considered. Would the approach be directly applicable if nuggets are included in the covariance model? An ambitious goal in palaeoclimate studies is simultaneously to reconstruct the entire space–time process of the temperature and other climate variables. Therefore, it again requires computational efficiency for spatiotemporal and multivariate data. We look forward to seeing further developments on this topic and in the mean time congratulate the authors for their outstanding work!

**Georg Lindgren** (*Lund University*)

I would like to add to the impressive list of applications of the Gaussian Markov random field–stochastic partial differential equation (SPDE) link, namely its use in ocean wave modelling. Traditionally, stochastic wave models have been based on linear Fourier analysis, possibly including low order interactions between the Fourier components. Such models are seen as approximations to the basic hydrodynamical (deterministic) partial differential equations for water waves. These equations have, in themselves, little room for stochastic forces.

One of the common spectra used in ocean modelling is the Pierson–Moskowitz spectrum

$$S(\omega, \theta) = A_{PM}\omega^{-5} \exp(-B_{PM}/\omega^{-4}) \cos^{2s}(\theta - \theta_0).$$

The SPDE approach, developed in the paper presented, offers a promising link between the hydrodynamic and Fourier view on random ocean waves. It has recently been shown by David Bolin and Finn Lindgren (see Bolin and Lindgren (2011b) for the general theory and Bolin (2009) and Lindgren *et al.* (2010) for the wave application) that the solutions to a *nested* SPDE,

$$(\kappa^2 - \Delta)^{(s+2)/2} X(\mathbf{t}) = (\mathbf{B}^T \nabla)^s W(\mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^2,$$

has a spectral density

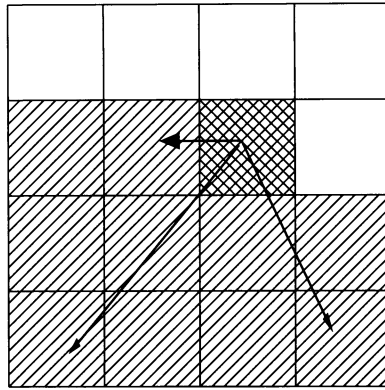
$$S(\omega, \theta) = C \frac{\omega^{4s+3}}{(g^2 \kappa^2 + \omega^4)^\alpha} \cos^{2s}(\theta - \theta_0),$$

with  $g$  equal to Earth’s acceleration. The vector  $\mathbf{B} = (b_1, b_2)^T$  determines the main direction  $\theta_0$  of the directional spectrum. For large  $s$ , this is close to the Pierson–Moskowitz wave spectrum and, thus, the SPDE approach could turn out to be a flexible alternative to the Fourier approach.

**K. V. Mardia** (*University of Leeds*)

I found the paper very timely and stimulating. The problem of dealing with large spatial data has a long history and the authors have given a comprehensive way forward. Mardia (2007) has given a historical background to the maximum likelihood methods for spatial data and pointed out that it seems there are still two main communities—one mining practitioners and the other mainstream statisticians.





**Fig. 19.** Strategy for composite likelihood for the criss-cross pixel by using a random selection of three points in a unilateral scheme,  $m = 3$

The Matérn covariance function has now been accepted by statisticians as a reasonable covariance function, and rightly, in general, the smoothing parameter  $\nu$  is taken as a tuning parameter. But there are some scientific problems where the estimation of  $\nu$  is the main interest, for example, when the geovariables are sampled intensively as for a computer-scanned image in petrology or in a high resolution image in remote sensing. Another example is of water head–piezometrics level  $\mathbf{u}$  where the hydraulic gradient map is the main interest (Pardo-Igúzquiza *et al.*, 2008); the deterministic equation which describes the movement of groundwater is a particular case of equation (2).

Pardo-Igúzquiza *et al.* (2008) have given the background to a computer program for the estimation of all the parameters of the Gaussian random field with Matérn covariance function ( $\kappa, \nu, \sigma$  and drift; geometrical anisotropy). The code is available at <http://www.iamg.org/CGEditor/index.htm> and as far as we are aware the geostatistics community has found this software useful. But the program works for  $n$  only up to 700. To extend this work for large data sets, in my joint work with Pardo-Igúzquiza, we have used various composite likelihood strategies extending Vecchia (1988) and Pardo-Igúzquiza and Dowd (1997). In particular, we have found unilateral random selection as shown in Fig. 19 useful for lattice data, and the method is extended for irregular data by using random permutation (even *a priori* for lattice data) where  $m$  is the number of the sites used in conditioning. We obtain the value of  $m$  through cross-validation but it is normally 5–15 for various large sets (images) we have experimented with. Indeed, the program takes a few minutes for  $n$  as large as 250 000, the total number of operations being proportional to  $n$ . The recent efficiency results given in Mardia *et al.* (2010) on composite likelihood are also encouraging and relevant. I welcome a comprehensive comparison of the composite likelihood approach and the finite element method that is used in this paper.

**Jorge Mateu** (*University Jaume I, Castellón*)

The authors are to be congratulated on a valuable contribution and a thought-provoking paper. Spatial data are frequently modelled as realizations of random fields. Gaussian fields (GFs) have a dominant role in spatial statistics and form an important building block in modern hierarchical spatial models. A common approach is to model spatial dependences through a covariance function. However, from the computational side, GFs are hampered by *the big n problem*, as the authors outline. In the way to overcome such a problem, rather than applying covariance tapering, we argue the use of compactly supported covariance functions which considerably reduce the computational burden of kriging, and allow the use of computationally efficient sparse matrix techniques, thus becoming a core aspect in spatial prediction when dealing with massive data sets. In addition, by considering a class of convolution-based flexible models, we can generate classes of non-stationary, compactly supported spatial covariance functions (Mateu *et al.*, 2011).

Rather than assuming that the GF is built through a Matérn covariance function, have the authors thought about considering other covariance radial basis functions with some positive definite kernel? Alternative stochastic partial differential equations to equation (2) could be brought into play. In their main result 2, the authors use a particular finite element method. An alternative could be using a Tikhonov regularization scheme with particular adapted functionals, like those shown in Montegranario and Espinosa

(2007). This regularization method has the mathematical advantage of using properties of compact operators in Hilbert spaces (Vapnik (1998), chapter 7).

I would like to draw the authors' attention to the use of basis functions to construct a finite element representation of the solution to the stochastic partial differential equation. First, they use  $n$  (the number of vertices) functions, and this is basically oversmoothing the solution. An objective criterion following a kind of cross-validation approach (as in functional geostatistics) should be considered. Second, the authors do not explicitly explain which appropriate basis functions could be used. They use harmonic and  $B$ -splines: why not wavelets or any other smoothing functions? Recall that some of these bases make orthogonality an issue in computation. Functional geostatistics is a field providing huge amounts of data, I would like to ask the authors whether they have ever thought about the possibility of having a GF of functions as in Giraldo *et al.* (2011). Do the authors have any tangible results on this emerging issue?

**Debashis Mondal** (*University of Chicago*)

The first main result of this paper was familiar to me (and to my late colleague Julian Besag) because of our work on the de Wijs process. In particular, note sections 4.2 and 5 of Besag and Mondal (2005) that allude to such results from a frequency domain perspective and from a computational point of view. Consider for example a sequence of Gaussian Markov random fields on sublattices  $\mathcal{Z}_m^2$  with spacing  $1/m, m = 1, 2, \dots$ , which have individual spectral densities of the form

$$f_m(\omega, \eta) = \frac{\sigma_m^2}{m^2(1 - 4\beta_m + 4\beta_m[\sin^2\{(1/2m)\omega\} + \sin^2\{(1/2m)\eta\}])^\nu}, \quad \omega, \eta \in (-\pi m, \pi m],$$

$$|\beta_m| < \frac{1}{4}, \quad \nu = 0, 1, 2, \dots$$

Here  $\nu = 1$  corresponds to the first-order conditional auto-regressions at increasingly dense sublattices,  $\nu = 2$  suggests Whittle's simultaneous auto-regressions at increasingly dense sublattices and so on. Now assume, as  $m \rightarrow \infty, \beta_m \rightarrow \frac{1}{4}, 4m^2(1 - 4\beta_m) \rightarrow \kappa^2$  and  $m^{\nu-1}\sigma_m \rightarrow \sigma/2^\nu$ . Then  $f_m(\omega, \eta)$  converges to

$$f(\omega, \eta) = \frac{\sigma^2}{(\kappa^2 + \omega^2 + \eta^2)^\nu},$$

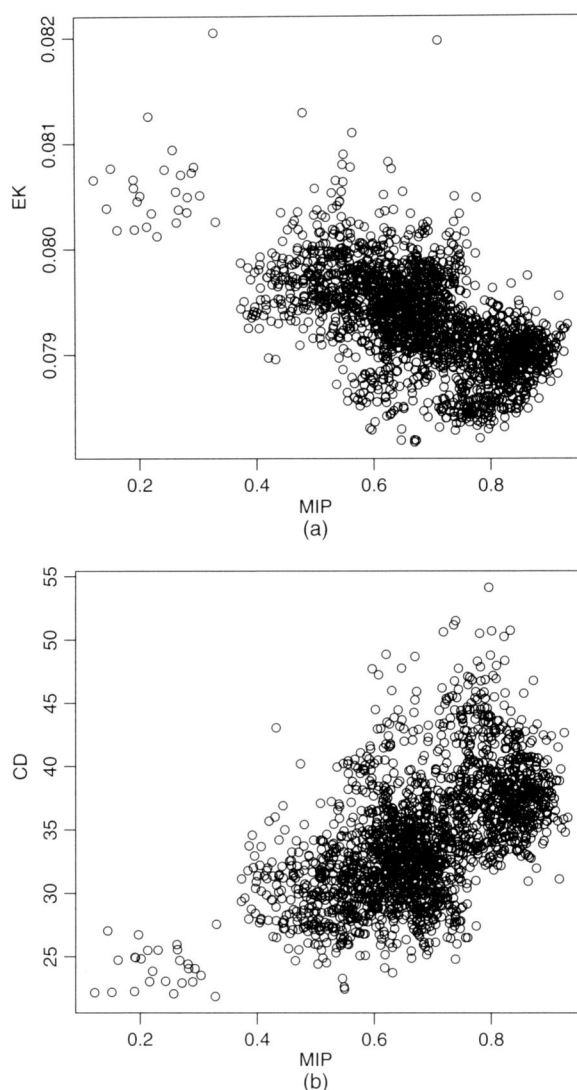
from which it follows that the corresponding continuum Gaussian random fields (generalized if  $\nu = 0, 1$  or if  $\nu > 1, \kappa = 0$ ) emerge as scaling limits of the Gaussian Markov random fields on regular lattices. One point I make here is that the above explicitly describes the rescaling of parameters needed, particularly when we want to choose a suitable sublattice (e.g. to embed irregular sampling locations into a grid or to approximate irregular regions by unions of grid cells when observations themselves are aggregates over such regions). This is important for both estimation and inference of continuum random fields from the approximate Gaussian Markov random field, but it did not receive much attention in the current paper. In addition, rescaling of parameters would also be required for the triangulation scheme discussed in the second main result of the paper, and it would be interesting to see the effect of numeric convergences as one considers denser triangulations.

Delaunay triangulation provides a way to place a dependence graph on irregularly sampled observations. Then the idea is to use a Gaussian Markov random field on this triangulation to approximate the continuum Gaussian random field. For this, the use of the graph Laplacian becomes important, and one can proceed with alternative calculations. Here I draw attention to the work of Coifman and Maggioni (2004) on diffusion wavelets that allow fast and approximate computations of functionals of inverse Laplacian and related diffusion-like operators on manifolds, graphs and other non-homogeneous media. A study of the strengths and weaknesses of this procedure will be a matter of future research.

**Werner G. Müller and Helmut Waldl** (*Johannes-Kepler-Universität Linz*)

The authors thankfully further strengthen the bridge connecting the somewhat disparate areas of geostatistics and spatial econometrics. We would like to draw attention towards a rather neglected aspect of establishing a link as above, namely the potential effect on the respective optimal sampling designs. We shall illustrate our points on the leukaemia survival example from Section 2.3, utilizing some of the authors' calculations.

In geostatistics the optimal sampling design is often based on the kriging variance over the region of interest, frequently by minimizing its maximum. Accounting for the additional uncertainty due to



**Fig. 20.** (a) MIP (horizontal) versus EK (vertical) criterion values and (b) MIP (horizontal) versus  $CD_{0.99}$  (vertical) criterion values

estimating covariance parameters Zhu and Stein (2006) and Zimmerman (2006) have employed a modification termed the empirical kriging (EK) criterion by the latter. In spatial econometrics it is common to test for spatial auto-correlation by specifying a spatial linkage matrix and to utilize an overall type of measure such as Moran's  $I$ . Therefore Gumprecht *et al.* (2009) have suggested to maximize the power of Moran's  $I$ -test under a hypothesized spatial lattice process, call it the Moran  $I$ -power (MIP) criterion.

From the example one sensible design question we could pose is which out of the 24 districts should we sample if we are limited to a number  $k < 24$  for financial reasons, say  $k = 3$ , which allows for  $\binom{24}{3} = 2024$  different configurations? Randomly sampling 20 locations from those as designs we can then draw a scatter plot of the values of the above criterion to judge for a potential linkage (Fig. 20(a)). EK reduces to scalar operations localized at  $\rho = 0.26$ , the only free covariance parameter. For the MIP we use the corresponding precision matrix provided by the authors, the linkage matrix assigned 1 to point pairs within the range  $\rho$ . Although the evident link between the criteria does not extend well into the lower right-hand corner

where the optima lie, it looks as if we could achieve reasonably high design efficiencies by substituting the criteria for each other.

Both criteria, however, are computationally quite intensive and it thus makes sense to look for cheaper alternatives. Motivated by traditional connections between estimation- and prediction-based criteria ('equivalence theory'), Müller and Stehlik (2010) have suggested replacing the EK criterion by a compound criterion for determinants of information matrices with a weighing factor  $\alpha$ , call it compound  $D$  optimality,  $CD_\alpha$ . A scatter plot of its values (assuming a constant trend) against the MIPs (Fig. 20(b)) shows high efficiencies on the MIP criterion for the computationally extremely cheap compound- $D$ -optimal design.

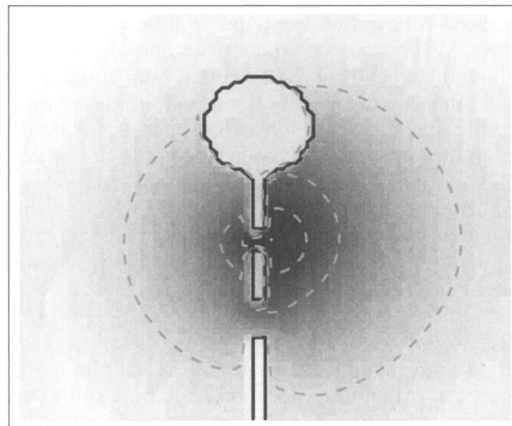
Summarizing, we believe that we have demonstrated that the relationships between the two linked approaches can go far beyond mere estimation issues.

**Alessandro Ottavi and Daniel Simpson** (Norwegian University of Science and Technology, Trondheim)

The Markov property, which is so important in temporal and discrete models, has long been absent in continuous spatial models, and the authors are to be congratulated for reminding us that it exists and is useful. In particular, the connection between stochastic partial differential equations and Gaussian random fields gives a practical method for specifying the *local* properties of a random field. Within this context, we are interested in using local effects to model non-stationarity. In particular, we are interested in non-stationarity induced by local topography, especially in the context of Bayesian smoothing.

Constructing Bayesian smoothers over complicated regions (i.e. regions that are not  $\mathbb{R}^2$ ) is a difficult problem and most successful efforts have involved, in some way, partial differential operators (Wood *et al.*, 2008; Ramsay, 2002). We note in passing that the techniques described in the paper under consideration yield an efficient Bayesian extension of the FELSPLINE method of Ramsay (2002), which does not suffer the large sample size drawbacks of soap film smoothing (Wood *et al.*, 2008). In fact, as these models are defined over an *irregular* tessellation, we can represent any hard physical boundary (almost) exactly and we can control either the (deterministic) value of the field on, or the flow through, the boundary by using standard finite element techniques.

Whereas a river, lake or mountain may provide a hard physical boundary, in other situations the constraints may simply impede or discourage movement. Such *soft* constraints are difficult to deal with properly and we propose to model them by locally deforming the physical space into the third dimension (see Sampson and Guttorp (1992)). This deformed space can then be directly tessellated and non-stationary SPDE models can be built on it. A caricature deformation is provided in Fig. 21, which shows the covariance centred at one point. The covariance can be clearly seen to wrap around the deformation in a sensible way.



**Fig. 21.** Covariance function centred at the darkest point of the non-stationary random field induced by a deformation: the contours of the covariance function can be clearly seen to distort locally around the object, whereas the far away covariances remain almost unchanged

**Omiros Papaspiliopoulos** (Universitat Pompeu Fabra, Barcelona) and **Emilio Porcu** (University of Castilla la Mancha and University of Göttingen)

We congratulate the authors for this beautiful paper. We have some suggestions that may be considered by the authors for a more general view of the problem.

- (a) Main result 1 does not hold when the parameter  $\nu$  is not an integer. A more general setting would allow us to index the fractional Sobolev space associated with the Gaussian random field (GRF) with a Matérn covariance function and thus allow us to study the properties of such a GRF in terms of fractal dimension. Such information seems to be lost with the approach proposed by the authors. In this respect, the tapering approach proposed in Gneiting (2002) seems to be more promising.
- (b) The Matérn model is very popular also because it is very handy; having a closed form for the related spectrum, it has been widely used, for instance, to apply Yadrenko's (1983) theory for the equivalence of GRF measures. But there are other covariance functions, such as the generalized Cauchy (Gneiting and Schlather, 2004) and the Dagum (Berg *et al.*, 2008) functions, which allow the separation of the fractal dimension and the Hurst effect of the associated GRF. This is a significant advantage for statistical estimation, and such a property is not offered by the Matérn covariance, which has light tails. The generalized Cauchy model

$$C_{\alpha,\gamma}(\mathbf{u}, \mathbf{v}) := (1 + \|\mathbf{v} - \mathbf{u}\|^\alpha)^{-\gamma}, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^d, \quad (46)$$

is positive definite on  $\mathbb{R}^d$  for all  $d \in \mathbb{N}$ , for  $0 < \alpha \leq 2$  and  $0 < \gamma$ , whereas the Dagum model admits the expression

$$K_{\beta,\gamma}(\mathbf{u}, \mathbf{v}) := 1 - \left( \frac{\|\mathbf{v} - \mathbf{u}\|^\beta}{1 + \|\mathbf{v} - \mathbf{u}\|^\beta} \right)^\gamma, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^d, \quad (47)$$

for which sufficient conditions of positive definiteness on  $\mathbb{R}^d$  for all  $d$  are  $\beta\gamma \leq 1$  and  $\beta < 1$ .

Recently, Ruiz-Medina *et al.* (2011) have shown that, for a GRF with a generalized Cauchy or with a Dagum covariance function, we have a local pseudodifferential representation, in the mean-square sense, of the type

$$(-\Delta)^\rho X(\mathbf{u}) = \mathcal{W}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^d, \quad (48)$$

where  $-\Delta$  denotes the negative Laplacian operator and  $\mathcal{W}$  is Gaussian white noise and where  $\rho$  is identically equal to  $(d + \alpha)/4$  and  $(d + \gamma\beta)/4$  for the generalized Cauchy and the Dagum functions respectively. It would be interesting to consider whether some type of Gaussian Markov random-field approximation applies to these GRFs.

**Marc Saez** (University of Girona and Consortium for Biomedical Research Network in Epidemiology and Public Health, Barcelona) and **Jorge Mateu** (University Jaume I, Castellón)

We first congratulate the authors for this excellent and innovative work, providing an authoritative review of the 'big  $n$  problem'. We indeed believe that the paper will become a seminal paper in the context of computational spatial statistics. We focus on comments on some aspects that we believe merit further discussion; in particular on how the stochastic partial differential equation generalizes to non-separable space-time models. Whereas in other extensions (the case of manifolds or non-stationary fields) the authors adequately extend the basic approach, in the non-separable space-time case the explanations given on how to obtain a Gaussian Markov random-field representation are not self-contained. It is not clear to us what the resulting system of coupled temporal stochastic differential equations is, or how the system can be discretized, even though the authors suggest using an Euler method. Can the authors be more specific in this respect? Another question related to this extension is to know which regularity conditions are violated when the noise process is not white in time. There is no doubt that this assumption is too restrictive, and we would like to know whether it could be relaxed in some way.

In relation to the aspects that deserve more discussion, we first would like to refer to the edge effects and boundary conditions and, secondly, to the issues of spatial scale. It is known that, at least in small databases or spatial data with a large spatial scale, edge effects can have an effect on the spatial pattern near the edge. It would be necessary, therefore, to investigate how to incorporate a varying edge in the representation. When considering covariates, it is also important to investigate the optimal spatial scale, i.e. the resolution, to avoid overfitting. The resolution of the spatial effect should not be less than the spatial scale of the covariates because otherwise the spatial effect, and not the covariates, explains the data. Since there is a trade-off between the accuracy of the Gaussian Markov random-field representation and the

number of vertices, further research is needed on the optimal number of triangles, i.e. the resolution, in conjunction, if possible, with the presence of varying edge effects.

**Alexandra M. Schmidt** (*Universidade Federal do Rio de Janeiro*)

The idea of approximating continuously indexed Gaussian fields through Gaussian Markov random fields is extremely appealing because of the sparseness of the precision structures of Gaussian Markov random fields. Even with current computational power, this is a real gain, especially in the case of large data sets. Moreover, the approach proposed provides a wide class of spatial models, allowing for the construction of Matérn fields on the sphere, and also non-stationary and oscillating covariance structures.

I shall concentrate my comments on the non-isotropic models aspect of the model proposed. Schmidt and O'Hagan (2003) proposed a Bayesian approach to the deformation idea of Sampson and Guttorp (1992). The idea is to assign a Gaussian process prior to the deformation function  $f(\cdot)$  that maps the original locations in  $R^2$  into the latent space (also in  $R^2$ ). The smoothness of the deformation is defined by the smoothness of the covariance function of  $f(\cdot)$ . Schmidt and O'Hagan (2003) used a Kronecker product between a covariance matrix describing the covariance between the axes that define the latent space and the correlation between the monitored locations in the original space. This correlation is modelled through a squared exponential correlation function, which is equivalent to assuming a Matérn correlation function for which  $\nu \rightarrow \infty$ . This assumption leads to an infinitely mean-squared differentiable function and a smooth mapping. The idea behind this structure is to obtain a deformation that is not very different from the original configuration but is still sufficiently flexible to capture the non-stationarity in the data. In Section 3.4, the authors obtain a non-stationary stochastic partial differential equation (SPDE) that reproduces the deformation method assuming Matérn covariances. This is done by fixing  $\alpha = 2$  which is equivalent to  $\nu = 1$ , when  $d = 2$ . Therefore, the original Matérn correlation function of the mapping function  $f(\cdot)$  assumes  $f(\cdot)$  is only one time mean-square differentiable. This might lead to undesirable folding of the configuration of points in the latent space. The authors also claim that one advantage is that the parameters of the resulting SPDE do not depend directly on the deformation itself. Although this might be an advantage for inference, it is important to recover the posterior distribution of  $f(\cdot)$  to understand the non-stationarity in the data. More recently, Schmidt *et al.* (2011) extended the mapping onto  $R^d$  for  $d > 2$ , by making use of covariates to define the other dimensions of the latent space. Increasing the dimension of the latent space helps to avoid the foldings mentioned above. They also discuss a simpler version of the model which also makes use of covariates. It is appealing that the SPDE provides another possibility for exploring covariates in the spatial covariance structures.

**Daniel Simpson** (*Norwegian University of Science and Technology, Trondheim*)

I congratulate Lindgren, Rue and Lindström on their outstanding contribution to the spatial statistics literature. In the simplest case, when the field is stationary and isotropic, it is easy to write down the Green function of the partial differential operator and, therefore, to derive the convolution representation of the corresponding Matérn field. When  $\nu = \alpha - d/2$  is the smoothness parameter and  $\kappa$  is the range parameter, this representation is

$$x(s) = \frac{\sigma^2}{\Gamma(\eta + d/2)(4\pi)^{d/2}\kappa^{2\eta}2^{\eta-1}} \int_{\mathbb{R}^d} (\kappa\|s - t\|)^\eta K_\eta(\kappa\|s - t\|) dW(t),$$

where  $\eta = (\nu - d/2)/2$  may be negative. It is tempting to approximate this by a finite sum using the method introduced by Higdon (1998). *This does not work!* The kernel convolution method, besides being less efficient than the method in the paper under discussion, produces posterior fields with noticeable artefacts (Bolin and Lindgren, 2009; Simpson *et al.*, 2010).

Of course, the paper considers much more than the simplest case with great success. The finite element method introduced succeeds in these complicated cases because it *directly* approximates the required random function in a stable consistent manner. When considering a bounded domain in  $\mathbb{R}^d$ , the proof of the convergence result (equation (11) in the paper) boils down to an application of Céa's lemma, which says that the error in the Galerkin approximation to an elliptic partial differential equation is bounded above by the error of the best approximation to the solution over the approximation space chosen. The space of piecewise linear functions is extremely well studied and its approximation properties have formed the backbone of finite element theory since the middle of last century.

I am extremely excited that the authors have extended the spatial statistics toolbox to include the finite element method, which is the workhorse of applied mathematics, physics and engineering. My excitement stems mostly from the fact that, given Gaussian data, the mean of the posterior random field can be

approximated in *linear* time by using multigrid methods. For non-Gaussian data, the maximum *a posteriori* estimate can be computed efficiently by solving a non-linear partial differential equation (see Hegland (2007) and Griebel and Hegland (2010), who considered this in the context of density estimation). The obvious challenge for the computational statistics community is to compute uncertainty in linear time. If this can be done, it will be possible to solve truly huge problems.

**Alfred Stein** (*Twente University, Enschede*)

This paper is a fine contribution to spatial statistics. We may have been aware that Gaussian Markov random fields and the Gaussian fields are connected, but the paper provides a solid explanation on the basis of mathematical theory. It raises several items for discussion.

- (a) The choice of Gaussianity is particular and I doubt whether equally elegant results can be obtained for non-Gaussian random fields, in particular for spatial count data. Also, the explicit link exists for two, relatively wide classes of covariance functions: the Matérn class and periodic covariance functions. However, there are more permissible covariance functions, like spherical covariances. It is not immediately obvious how the current procedure can be expanded. Finally, multivariate spatial data are natural expansions of the univariate case presented; for example a bivariate Gaussian random field may have an explicit link with the bivariate Gaussian Markov random field in an almost trivial way, but it would be good to hear the authors say the final word on it.
- (b) Applying the method apparently requires a triangularization of the area. I cannot believe that it is always as simple and straightforward as in the examples now. Moreover, results may be sensitive to a particular choice. In an image analysis, it is natural to have a square grid of pixel values, whereas traditional geostatistics have data irregularly spread over an area, and a map on a square grid contains a visualized output map as an essential result. It is not entirely clear yet what the effects are of such choices in the final results.
- (c) This brings me then to my final comment, namely the issue of quality of data. In the current setting, the quality of data is entirely governed by a choice of the (Gaussian) distribution. For many data that is not enough, as locations may be uncertain, or refer to an aggregated area, whereas in other studies the data themselves are poorly defined. Yet, scientists collect and analyse them, simply because they are the best available to tell a scientific story—they are fit for use. A good example refers to data that are obtained from interpreting a soil layer and that are poorly defined just because the soil classes are poorly defined. Scientists rely on fuzzy approaches here. I wonder how the theory presented could proceed in such circumstances.

As these are all promising aspects for future research, for now I can only compliment the authors on their achievements.

**Paul Switzer** (*Stanford University*)

I have a concern regarding spatial models based on a direct specification of local conditional distributions, although these enjoy substantial computational advantages compared with approaches that use the unconditional Gaussian field (GF) covariance function for local estimation. Nevertheless, the conditional specification is not completely satisfactory, for example because it says nothing about how to construct models in a consistent way for different grid spacings. Your approach to this problem is insightful—estimate parameters of the unconditional covariance by using available data and then find a local conditional model that is approximately consistent with this estimated GF.

If you are starting with a GF then a device used to sidestep the large  $n$  problem for local inference is to condition arbitrarily only on observations within a neighbourhood by relying on the ‘screen’ effect of the closest data. Data weights for filtering or interpolation are determined by applying the GF specification restricted to the estimation location and a small number of neighbouring observations, via kriging. Whereas kriged maps will typically resemble maps of posterior means, this may not be so for local estimation precision which is more model dependent.

GF parameter estimates should be about the same regardless of the grid mesh that is used to sample a realization of a GF, i.e. the parameter  $\nu$  in the Matérn covariance does not depend on how the GF is sampled. This would seem to imply that the Gaussian Markov random-field representation would have the same order neighbourhood regardless of the grid spacing that is used to sample the GF.

The covariance of an isotropic GF for a two-dimensional field is the same as the one that we would obtain if the two-dimensional field were restricted to a one-dimensional linear transect. So the Matérn

parameter  $\nu$  would be the same whether estimated from one-dimensional data or two-dimensional data, although the implied neighbourhood order would be different for  $d = 1$  and  $d = 2$ .

Finally, I am curious how local estimation is affected by the choice of triangulation that is imposed on the observation domain for irregularly spaced data. Your suggestion, to have smaller triangles where there are more data makes sense but can anything more be said? For  $d = 1$  how should we think about triangulation for unevenly dense observations?

**Kamil Turkman** (*University of Lisbon*)

I congratulate the authors for this excellent work which will have significant consequences in modelling spatial data.

The objective of the paper is to find ways to treat the data as point referenced and to use a Gaussian field (GF) as the model, and to do the computations as if the data are discretely indexed by using an appropriate Gaussian Markov random-field (GMRF) model which represents the GF in the best possible way. The authors stress the word representation rather than approximation, perhaps indicating that this GMRF, although it is the best representation, may not approximate GFs to a desired level.

To switch from one model to another in this manner during the different phases of the modelling, we must assume that the likelihoods under the GMRF and GF models do not differ or deviate much. Indeed, this point was clearly expressed in Rue and Held (2005), where the GMRF representation was chosen by minimizing a metric of discrepancy of the form

$$\|\Sigma_1 - \Sigma_0\| = \sum_{i,j} \{r_1(i,j) - r_0(i,j)\}^2 w_{ij},$$

where  $\Sigma_1$  and  $\Sigma_0$  denote respectively the covariance matrices of the GF and GMRF over the observed data locations, and  $r_1(i,j)$  and  $r_0(i,j)$  are respectively the elements of  $\Sigma_1$  and  $\Sigma_0$ . The weights are chosen as being inversely proportional to distances. In contrast, the normal comparison lemma and its refinements (Piterbarg, 1996) suggest that the metric of discrepancy should be based on

$$\sum_{i>j} |r_1(i,j) - r_0(i,j)|,$$

since

$$F_1(x) - F_0(x) = \sum_{i>j} \{r_1(i,j) - r_0(i,j)\} \int_0^1 F_h(x|X_{hi} = x_i, X_{hj} = x_j) \phi_h(x_i, x_j) dh \tag{49}$$

where  $F_1(x)$  and  $F_2(x)$  are the density functions of the vectors  $X_1$  and  $X_0$  corresponding to GF and GMRF processes at the data locations,  $X_h = X_1\sqrt{1-h} + X_0\sqrt{h}$  for some,  $h \in (0, 1)$ ,  $\phi_h(x_i, x_j)$  is the joint density function of  $X_{hi}$  and  $X_{hj}$  and  $F_h(x|\cdot)$  is the conditional distribution function of  $X_h$ .

From this equality, a sharp inequality can be found:

$$|F_1(x) - F_0(x)| \leq \sum_{i>j} |r_1(i,j) - r_0(i,j)| \int_0^1 \phi_h(x_i, x_j) dh,$$

In this paper, the optimal GMRF representation for a specific class of GFs and for a given irregular grid configuration is obtained in a different norm. However, it is not clear what sort of discrepancy this representation induces on the covariances, and hence on the likelihoods under GF and GMRF models and its consequence on the inference. The worrying point is that, in accordance with the normal comparison lemma, the total error of approximation is additive in the L1 covariance errors.

**Christopher K. Wikle** (*University of Missouri, Columbia*) and **Mevin B. Hooten** (*Utah State University, Logan*)

First, we congratulate the authors on another important contribution in what is quickly becoming a renaissance on approximate Bayesian methods. Given that our interest in correlated random fields has been primarily from a continuous spatial perspective rather than a discrete one, this latest paper is quite intriguing. The general concept of thinking about common forms of dependence as a solution to a spectrally defined stochastic partial differential equation (SPDE) is clever and similar in spirit to other general ideas we have been fond of in the past.

As we have spent a considerable amount of time thinking about dynamic spatiotemporal models and the origin of spatial processes, we see this paper as a potentially very valuable contribution. Indeed, the notion of discretizing SPDEs to form the basis of Markovian statistical models from both the spectral (e.g. Wikle *et al.* (2001)) and physical space (e.g. Wikle (2003), Wikle and Hooten (2006, 2010) and Hooten and Wikle (2007)) perspectives has been a primary focus of our own research. A heuristic summary of the relationship



between such approximations and their theoretical counterparts can be found in Cressie and Wikle (2011). A key component of these presentations is the full exploitation of the hierarchical framework that allows us to place dependent random processes on the parameters that control the dynamical interactions (i.e. the parameters in the SPDE), for which no analytical solution exists. We have focused on building models true to the aetiology of the underlying processes—which is, more often than not, in the (Markovian) dynamical evolution. Furthermore, it is important to remember that real world processes are non-linear. The development of analytical covariance functions for the governing SPDEs of such processes are typically intractable, yet the motivating Markovian models suggested by the discretization (either in physical or spectral space) are quite tractable and flexible (for example, see Hooten and Wikle (2007)). Wikle and Hooten (2010) discussed a general form for ‘quadratic interaction models’ and showed the connection to various classes of spatiotemporal PDEs. Wikle and Holan (2011) extend this to higher order interactions.

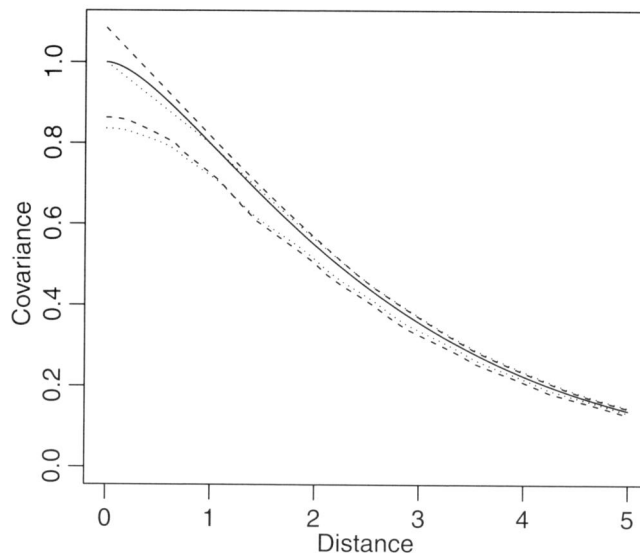
From a computational perspective, we have found that the algorithms of Rue and Martino (2009) are very computationally efficient when a solution can be found. Thus, for very large and well-behaved continuous physical systems where sufficient data exist and correlation structures are smooth, the approach outlined by Lindgren and his colleagues may be quite valuable for computational reasons alone.

The authors replied later, in writing, as follows.

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

#### Triangulations

As pointed out by Cooley, Hoeting and Brown, it is not necessary to place the triangulation vertices at observation points. Indeed, the *observation matrix* in the global temperature example in Section 4.2 was introduced for this very reason. For a given triangulation, the matrix can be used to extract any observable linear combination of field values, allowing observations in arbitrary locations as well as regional averages. For point observations, each row of the matrix contains three non-zero elements, one for each corner of the triangle containing the point, and the sparsity structure of the posterior precision matrix is unaffected. There is also no requirement to use a regular grid for such models. The triangulation implementation in `R-INLA` has a parameter for a minimum required distance between data-located vertices. In the temperature example this was set to 10 km, allowing the vertex placement to follow the data density, without



**Fig. 22.** Target covariance (—), linear interpolation covariance envelopes (·····) and GMRF covariance envelopes (-----): the envelopes show the minimum and maximum covariances for each pairwise point distance within the domain; the linear interpolation is accurate at grid nodes, but the GMRF approximation is closer to the target overall

generating excessively small triangles. An example where the triangulation was chosen completely independently of the data locations is given by Bolin and Lindgren (2011b). Crujeiras and Prieto add that the triangulation could be chosen adaptively on the basis of local approximation error estimates, and we agree that this can potentially be useful for non-stationary anisotropic models.

*Approximation properties*

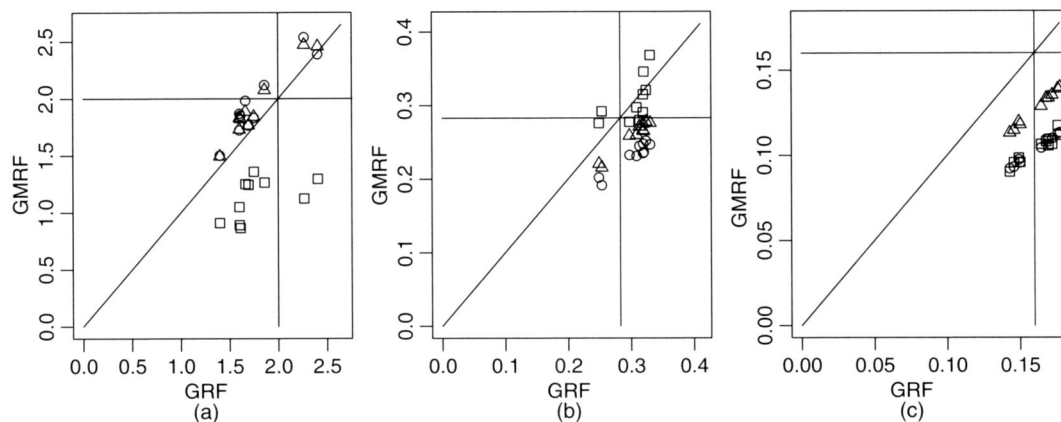
Ippoliti, Martin and Bhansali note that the stochastic partial differential equation (SPDE)–Gaussian Markov random-field (GMRF) precision coefficients do not match those obtained by inverting the covariance matrix of a field sampled on a regular grid. However, the approximation is not aiming to approximate the field only at the vertices of the triangulation, but also at the intermediate points obtained via linear interpolation. If the same interpolation method is used with the sampled covariances, the overall field covariances are underestimated, whereas the SPDE–GMRF approach gives an overall closer approximation. The upper and lower envelopes of all the pairwise covariances for the two settings are shown in Fig. 22, together with the target covariance function. Bolin and Lindgren (2009) investigated how this effect influences the kriging results, comparing with tapering and kernel convolutions, as well as alternative choices of basis functions in the GMRF construction, including wavelets and *B*-splines.

*Kernel methods*

Kernel convolution methods, as mentioned by Furrer, Furrer and Nychka, are useful theoretical tools, but can in practice be cumbersome and computationally intensive (Bolin and Lindgren, 2009). The kernel generated by the SPDE operator for a Matérn field takes the shape of another Matérn covariance with different shape parameter. The kernels are singular for  $\alpha \leq d$ , and non-differentiable for  $\alpha \leq d + 2$ , so the commonly used discrete kernel sums result in kriging and parameter estimation artefacts and do not yield either the correct pointwise distribution or correct distributions for regional averages, unless the range is large compared with the kernel placement distances. Also, since the kernels have non-compact support, they yield dense matrices for the posterior precisions. Using compactly supported kernels as suggested by Mateu is similar to moving average processes in time series analysis and is problematic unless the data are densely and evenly located on the domain of interest, whereas the GMRF models are counterparts to auto-regressive models, that are much more flexible tools for approximating general dependence structures.

*Parameter estimation*

Both when estimating parameters and calculating kriging interpolations, the results are influenced by the choice of boundary conditions. The easiest method for avoiding these effects is to extend the triangulation beyond the study region by an amount that is sufficiently large to cover the correlation range of the field, since this allows the boundary effects to drop off to virtually zero before reaching the data locations that influence the likelihood. As seen in Fig. 23(a), this eliminates the bias in the maximum likelihood estimates



**Fig. 23.** Estimated (a)  $\sigma^2$ , (b)  $\kappa$  and (c)  $\tau^{-2} \alpha \sigma^2 \kappa^2$  for an  $h = 1$  lattice GMRF ( $\square$ ), an  $h = 1$  lattice with irregular extension ( $\circ$ ) and an  $h = \frac{1}{2}$  lattice with irregular extension, all with Neumann boundary conditions, plotted against the corresponding estimates based on a sampled covariance matrix: the estimations were computed on 10 samples from a  $\nu = 1$  stationary Matérn field, with exact observation at a  $20 \times 20$  lattice

of the field variances  $\sigma^2$ . The bias for the rescaling parameter  $\tau^{-2}$  is reduced when the triangulation resolution is increased, as seen in Fig. 23(b). The parameter  $\sigma^2\kappa^{2\nu}$  that was used in the comparison by Lee and Kaufman is equivalent to  $\tau^{-2}$  in the SPDE models used in the paper. We believe that the highly variable results in their comparison is explained by noting that the precision matrix was chosen by simply deleting rows and columns, which is equivalent to conditioning, in this case leading to approximate Dirichlet boundary conditions. To compensate for the resulting small variances near the boundary of such a model, the variance parameter needs to be greatly overestimated. The combined comparisons show that Neumann boundaries are safer, and that extending the boundary further reduces the bias.

#### *Boundaries*

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

A common alternative to Neumann conditions is to let  $\partial_n^2 = 0$ , which is easily accomplished for regular grids by replacing the two-dimensional grid Laplacian with the one-dimensional Laplacian along the boundaries. For a unit square domain, the resulting null space is spanned by the four functions  $1, u_1, u_2$  and  $u_1u_2$  but the rank deficiency is only 3, since the last function is not piecewise linear. Although this gives approximately the standard polynomially intrinsic models on  $\mathbb{R}^2$ , the construction also hints at a more general method for more general SPDE models currently under investigation. The idea is to start with a fully intrinsic precision for the interior of the domain and to add appropriate penalties generated by SPDE models within the boundary. For one-dimensional models, this eliminates the boundary effects entirely, and the higher dimension cases show promising results.

#### *Model checking*

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

#### *Priors*

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

#### *Simultaneous auto-regressions*

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

*Fractional operators*

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

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

where  $p$  is the degree of the polynomial and  $b_i$  are coefficients in a strictly positive polynomial, and the corresponding discretized precision matrix can be obtained as

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

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

$$\int_{\mathbb{R}^d} \{ \tilde{S}(k) - S(k) \}^2 \omega(k) dk \propto \int_{\kappa^2}^{\infty} \left\{ z^\alpha - \sum_{i=0}^p b_i (z - \kappa^2)^i \right\}^2 z^{-2p-1-\lambda} dz.$$

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

$$\begin{pmatrix} b_0 \\ b_1 \kappa^2 \end{pmatrix} = \kappa^{2\alpha} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/(2+\lambda) & 1/(1+\lambda) \\ 1/(1+\lambda) & 1/\lambda \end{pmatrix}^{-1} \begin{pmatrix} 1/(2+\lambda-\alpha) \\ 1/(1+\lambda-\alpha) \end{pmatrix},$$

$$\begin{pmatrix} b_0 \\ b_1 \kappa^2 \\ b_2 \kappa^4 \end{pmatrix} = \kappa^{2\alpha} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/(4+\lambda) & 1/(3+\lambda) & 1/(2+\lambda) \\ 1/(3+\lambda) & 1/(2+\lambda) & 1/(1+\lambda) \\ 1/(2+\lambda) & 1/(1+\lambda) & 1/\lambda \end{pmatrix}^{-1} \begin{pmatrix} 1/(4+\lambda-\alpha) \\ 1/(3+\lambda-\alpha) \\ 1/(2+\lambda-\alpha) \end{pmatrix}$$

respectively.

The limiting case  $\lambda \rightarrow \infty$  is equivalent to Taylor approximation at  $k = \mathbf{0}$  and gives  $(b_0, b_1) = \kappa^{2\alpha-2} (\kappa^2, \alpha)$  for  $\lceil \alpha \rceil = 1$  and  $(b_0, b_1, b_2) = \kappa^{2\alpha-4} \{ \kappa^4, \alpha \kappa^2, \alpha(\alpha-1)/2 \}$  for  $\lceil \alpha \rceil = 2$ . These limiting approximations provide good agreement for integrals of the field over regions but, for better behaviour of the short distance covariances,  $\lambda$  needs to be chosen more carefully. For a given measure of deviation between the desired and approximate covariance functions, the optimal  $\lambda$  can be determined numerically, as a function of  $\alpha$ . For fractional  $\alpha$  between 0 and 2 on  $\mathbb{R}^2$ , the parsimonious choice  $\lambda = \alpha - \lceil \alpha \rceil$  approximately minimizes the maximal absolute difference between the covariance functions. As noted by Cooley and Hoeting,  $\alpha$  is in practice often chosen only from the integers and half-integers, and we obtain  $(b_0, b_1) = \kappa^{-1} (3\kappa^2/4, 3/8)$  for  $\alpha = \frac{1}{2}$  and  $(b_0, b_1, b_2) = \kappa^{-1} (15\kappa^4/16, 15\kappa^2/8, 15/128)$  for  $\alpha = \frac{3}{2}$ . Combined with the recursive construction for  $\alpha > 2$ , this provides GMRF approximations for all positive integers and half-integers. This includes the exponential covariance in  $\mathbb{R}^2$ , which corresponds to  $\alpha = \frac{3}{2}$ . The resulting covariance is shown in Fig. 24, together with the covariance from the spectral Taylor approximation. Further investigations are needed to determine how well the measurement noise model can incorporate the resulting deviation in small scale variation that is introduced by the approximation. Also shown in Fig. 24 is the covariance for a model with  $\alpha = 2$ , showing the same qualitative behaviour at zero, but different mid-range behaviour.

*Long-range dependence*

As discussed by Bhattacharya *et al.* (1983), apparent long-range dependence in data cannot be distinguished from a non-stationary mean or trend. An alternative to constructing covariance functions with such behaviour is therefore to use a two-stage model, where the local behaviour is treated separately from the long-range behaviour. In practical situations, spatially varying basis functions are often used to capture large-scale variations, leaving the rest for a spatial field component. This can easily be extended to allow the basis weights to differ between realization of the field, in effect increasing the spectral density near zero. For identifiability reasons, intrinsic models can be preferable, but alternatives such as conditioning on zero integral for the field can also be used, and are implemented in R-INLA. As suggested by Fearnhead, another even more general approach is to model the observed field as the sum of several latent fields with

different range. Care must be taken to handle the near non-identifiability of such models, noting that each individual latent component may be of less interest than their sum.

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

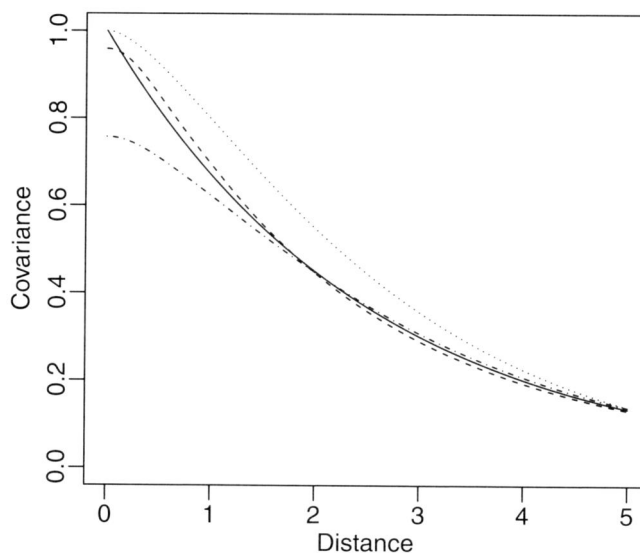
#### Deformations

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

#### General extensions

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

Functional data can be treated either directly in the general observation model, or by incorporating desired basis functions into the finite element basis itself, leading to block matrices in the precisions. In a



**Fig. 24.** Covariances-based spectral approximation: desired exponential covariance (—,  $d = 2$ ,  $\alpha = \frac{3}{2}$  and range 5), Taylor approximation (---), parsimonious weighting (-----,  $\lambda = \frac{1}{2}$ ) and theoretical covariance for  $\alpha = 2$  (· · · · ·)

setting with a local set of temporal basis functions  $\{\xi_k^1, \dots, \xi_k^m\}$  for each spatial triangulation vertex  $k$ , the resulting elements of the joint  $\mathbf{K}$ -matrix take the form

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

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

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

## References in the discussion

- Åberg, S. and Podgórski, K. (2010) A class of non-Gaussian second order random fields. *Extremes*, 1–36.
- Åberg, S., Podgórski, K. and Rychlik, I. (2009) Fatigue damage assessment for a spectral model of non-Gaussian random loads. *Probab. Engng Mech.*, **24**, 608–617.
- Abramowitz, M. and Stegun, I. A. (eds) (1965) *Handbook of Mathematical Functions*. New York: Dover Publications.
- Ainsworth, M. and Oden, J. T. (2000) *A posteriori* error estimation in finite element analysis. In *Pure and Applied Mathematics*. New York: Wiley.
- Anderes, E. B. and Stein, M. L. (2011) Local likelihood estimation for nonstationary random fields. *J. Multiv. Anal.*, **102**, 505–520.
- Balgovind, R., Dalcher, A., Ghil, M. and Kalnay, E. (1983) A stochastic-dynamic model for the spatial structure of forecast error statistics. *Monthly Weath. Rev.*, **111**, 701–722.
- 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.
- Barndorff-Nielsen, O. and Halgreen, C. (1977) Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions. *Probab. Theor. Reltd Flds*, **38**, 309–311.
- Berg, C., Mateu, J. and Porcu, E. (2008) The Dagum family of isotropic correlation functions. *Bernoulli*, **14**, 1134–1149.
- Berger, J. O., De Oliveira, V. and Sansó, B. (2001) Objective Bayesian analysis of spatially correlated data. *J. Am. Statist. Ass.*, **96**, 1361–1374.
- Bermúdez, A., Hervella-Nieto, L., Prieto, A. and Rodríguez, R. (2010) Perfectly Matched Layers for time-harmonic second order elliptic problems. *Arch. Computnl Meth. Engng*, **17**, 77–107.
- Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B*, **36**, 192–236.
- Besag, J. (1981) On a system of two-dimensional recurrence equations. *J. R. Statist. Soc. B*, **43**, 302–309.
- Besag, J. and Mondal, D. (2005) First-order intrinsic autoregressions and the de Wijs process. *Biometrika*, **92**, 909–920.
- Bhattacharya, R. N., Gupta, V. K. and Waymire, E. (1983) The Hurst effect under trends. *J. Appl. Probab.*, **20**, 649–662.
- Bolin, D. (2009) Computationally efficient methods in spatial statistics, applications in environmental modeling. *Licentiate Thesis*. Lund University, Lund.
- Bolin, D. and Lindgren, F. (2009) Wavelet Markov models as efficient alternatives to tapering and convolution fields. *Mathematical Sciences Preprint 2009:13*. Lund University, Lund.
- Bolin, D. and Lindgren, F. (2011a) Spatial wavelet Markov models are more efficient than covariance tapering and process convolutions. To be published.
- Bolin, D. and Lindgren, F. (2011b) Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Ann. Appl. Statist.*, **5**, 523–550.
- Bookstein, F. L. (1989) Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattn Anal. Mach. Intell.*, **11**, 567–585.
- Cameletti, M., Ignaccolo, R. and Bande, S. (2011) Comparing air quality statistical models. *Working Paper*. (Available from <http://arxiv.org/pdf/1011.1845v2>.)
- Carr, P., Geman, H., Madan, D. B. and Yor, M. (2003) Stochastic volatility for Lévy processes. *Math. Finan.*, **13**, 345–382.
- Challener, P., McNeill, D. and Gattiker, J. (2010) Assessing the probability of rare climate events. In *The Oxford Handbook of Applied Bayesian Analysis* (eds A. O'Hagan and M. West), pp. 403–430. Oxford: Oxford University Press.
- Christensen, G. E., Rabbitt, R. D. and Miller, M. I. (1996) Deformable templates using large deformation kinematics. *IEEE Trans. Im. Process.*, **5**, 1435–1447.
- Clark, P. K. (1973) A subordinated stochastic process model with finite variance. *Econometrica*, **41**, 135–155.
- Coifman, R. R. and Maggioni, M. (2006) Diffusion wavelets. *Appl. Computnl Harm. Anal.*, **21**, 53–94.

- Conti, S. and O'Hagan, A. (2010) Bayesian emulation of complex multi-output and dynamic computer models. *J. Statist. Plannng Inf.*, **140**, 640–651.
- Crainiceanu, C. M., Diggle, P. J. and Rowlingson, B. (2008) Bivariate binomial spatial modeling of loa loa prevalence in tropical Africa. *J. Am. Statist. Ass.*, **103**, 21–37.
- 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 Wikle, C. K. (2011) *Statistics for Spatio-temporal Data*. Hoboken: Wiley.
- Das, B. (2000) Global covariance modeling: a deformation approach to anisotropy. *PhD Dissertation*. Department of Statistics, University of Washington, Seattle.
- Diggle, P. J., Menezes, R. and Su, T. (2010) Geostatistical inference under preferential sampling (with discussion). *Appl. Statist.*, **59**, 191–232.
- Diggle P. J., Moyeed, R. A. and Tawn, J. A. (1998) Model-based geostatistics (with discussion). *Appl. Statist.*, **47**, 299–350.
- 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.
- Eliazar, I. and Klafter, J. (2004) Spatial gliding, temporal trapping, and anomalous transport. *Physica D*, **187**, 30–50.
- Eliazar, I. and Klafter, J. (2005) Lévy Ornstein-Uhlenbeck, and subordination: spectral vs. jump description. *J. Statist. Phys.*, **119**, 165.
- Fotopoulos, S. B., Jandhyala, V. K. and Khapalova, E. (2010) Exact asymptotic distribution of change-point mle for changes in the mean of Gaussian sequences. *Ann. Appl. Statist.*, **4**, 1081–1104.
- Furrer, E. M. (2008) Asymptotic behavior of a continuous approximation to the kriging weighting function. *Technical Note NCAR/TN476+STR*. National Center for Atmospheric Research, Boulder.
- Furrer, E. M. and Nychka, D. W. (2007) A framework to understand the asymptotic properties of Kriging and splines. *J. Kor. Statist. Soc.*, **36**, 57–76.
- Furrer, E. M., Nychka, D. W., Piret, G. and Furrer, R. (2011) An asymptotic framework under the Matérn covariance model. To be published.
- Gelfand, A. E. and Banerjee, S. (2010) Multivariate spatial process models. In *Handbook of Spatial Statistics* (eds A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp), pp. 495–515. Boca Raton: Chapman and Hall–CRC.
- Giraldo, R., Delicado, P. and Mateu, J. (2011) Ordinary kriging for function-valued spatial data. *Environ. Ecol. Statist.*, to be published.
- Givoli, D. (1999) Recent advances in the DtN FE method. *Arch. Computnl Meth. Engng*, **6**, 71–116.
- Gneiting, T. (2002a) Nonseparable, stationary covariance functions for space-time data. *J. Am. Statist. Ass.*, **97**, 590–600.
- Gneiting, T. (2002b) Compactly supported correlation functions. *J. Multiv. Anal.*, **83**, 493–502.
- Gneiting, T., Kleiber, W. and Schlather, M. (2010) Matérn cross-covariance functions for multivariate random fields. *J. Am. Statist. Ass.*, **105**, 1167–1177.
- Gneiting, T. and Schlather, M. (2004) Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Rev.*, **46**, 269–282.
- Godwin, D. J. (2000) Deformations in shape space. *PhD Dissertation*. University of Leeds, Leeds.
- Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.
- Griebel, M. and Hegland, M. (2010) A finite method for density estimation with Gaussian process priors. *SIAM J. Numer. Anal.*, **47**, 4759–4792.
- Gumprecht, D., Müller, W. G. and Rodríguez-Díaz, J. M. (2009) Designs for detecting spatial dependence. *Geograph. Anal.*, **41**, 127–143.
- Guttorp, P. and Gneiting, T. (2006) Studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika*, **93**, 989–995.
- Hegland, M. (2007) Approximate maximum a posteriori with Gaussian process priors. *Constr. Approximn*, **26**, 205–224.
- Hesthaven, J. S. and Warburton, T. (2008) *Nodal Discontinuous Galerkin Methods*. New York: Springer.
- 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 modeling. 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.
- Höhle, M. (2009) Additive-multiplicative regression models for spatio-temporal epidemics. *Biometr. J.*, **51**, 961–978.
- Hooten, M. B. and Wikle, C. K. (2007) A Hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environ. Ecol. Statist.*, **15**, 59–70.
- Illian, J. B. and Hendrichsen, D. K. (2010) Gibbs point processes with mixed effects. *Environmetrics*, **21**, 341–353.
- Illian, J. B. and Rue, H. (2010) A toolbox for fitting complex spatial point process models using integrated Laplace transformation (INLA). *Technical Report*. Trondheim University, Trondheim.

- Illian, J. B., Sørbye, S. H., Rue, H. and Hendrichsen, D. K. (2011) Fitting a log gaussian cox process with temporally varying effects—a case study. To be published.
- Irvine, K., Gitelman, A. I. and Hoeting, J. A. (2007) Spatial designs and properties of spatial correlation: effects on covariance estimation. *J. Agric. Biol. Environ. Statist.*, **12**, 450–469.
- Jandhyala, V. K., Liu, P. and Fotopoulos, S. B. (2011) Estimation of an unknown change-point occurring in the mean and covariance matrix of a multivariate Gaussian process with application to annual mean radiosonde temperature deviations at south and north polar zones. Submitted to *J. Clim.*
- Jun, M. and Stein, M. L. (2007) An approach to producing spacetime covariance functions on spheres. *Technometrics*, **49**, 468–479.
- Jun, M. and Stein, M. L. (2008) Nonstationary covariance models for global data. *Ann. Appl. Statist.*, **2**, 1271–1289.
- Keller, J. B. and Givoli, D. (1989) Exact nonreflecting boundary conditions. *J. Computnl Phys.*, **82**, 172–192.
- Kennedy, M. and O'Hagan, A. (2001) Bayesian calibration of computer models (with discussion). *J. R. Statist. Soc. B*, **63**, 425–464.
- Kent, J. T. and Mardia, K. V. (1994) The link between kriging and thin-plate splines. In *Probability, Statistics and Optimisation* (ed. F. P. Kelly), pp. 325–339. New York: Wiley.
- Kongsgård, H. W. (2011) The genetics of conflict: low level interaction between conflict events. (Available from <http://ssrn.com/abstract=1768198>.)
- Kumar, A., Meerschaert, M. M. and Vellaisamy, P. (2011) Fractional normal inverse Gaussian diffusion. *Statist. Probab. Lett.*, **81**, 146–152.
- Le, N. D. and Zidek, J. V. (2006) *Statistical Analysis of Environmental Space-time Processes*. New York: Springer.
- LeVeque, R. J. (2002) *Finite Volume Methods for Hyperbolic Problems*. Cambridge: Cambridge University Press.
- Li, B., Nychka, D. W. and Ammann, C. M. (2010) The value of multi-proxy reconstruction of past climate (with discussions and rejoinder). *J. Am. Statist. Ass.*, **105**, 883–911.
- Lindgren, G., Bolin, D. and Lindgren, F. (2010) Non-traditional stochastic models for ocean wave—Lagrange models and nested SPDE models. *Eur. Phys. J. Specl Top.*, **185**, 209–224.
- Loeppky, J. L., Sacks, J. and Welch, W. J. (2009) Choosing the sample size of a computer experiment: a practical guide. *Technometrics*, **51**, 366–376.
- Majumdar, A., Gelfand, A. E. and Banerjee, S. (2005) Spatio-temporal change-point modeling. *J. Statist. Plannng Inf.*, **130**, 149–166.
- Mardia, K. V. (2007) Should geostatistics be model-based? In *Proc. Conf. Geomathematics and GIS Analysis of Resources, Environment and Hazards, Beijing*, pp. 4–9.
- Mardia, K. V., Kent, J. T., Hughes, G. and Taylor, C. C. (2010) Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika*, **96**, 975–982.
- Martin, R. J. (2000) Exact Gaussian maximum likelihood and simulation for regularly-spaced observations with Gaussian correlations. *Biometrika*, **87**, 727–730.
- Martin, R. J. and Dwyer, J. (1994) Approximations to the covariance properties of processes averaged over irregular spatial regions. *Commun. Statist. Theor. Meth.*, **23**, 913–945.
- Mateu, J., Fernandez-Aviles, G. and Montero, J. M. (2011) On a class of non-stationary, compactly supported spatial covariance functions. *Stochast. Environ. Res. Risk Assessmnt*, to be published.
- Møller, J., Syversveen, A. and Waagepetersen, R. (1998) Log Gaussian Cox processes. *Scand. J. Statist.*, **25**, 451–482.
- Møller, J. and Waagepetersen, R. P. (2004) *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall–CRC.
- Møller, J. and Waagepetersen, R. P. (2007) Modern statistics for spatial point processes (with discussion). *Scand. J. Statist.*, **34**, 643–711.
- Montegranario, H. and Espinosa, J. (2007) A regularization approach for surface reconstruction from point clouds. *Appl. Math. Computn*, **188**, 583–595.
- Müller, W. G. and Stehlik, M. (2010) Compound optimal spatial designs. *Environmetrics*, **21**, 354–364.
- Narcowich, F. J. (1995) Generalized Hermite interpolation and positive definite kernels on a Riemannian manifold. *J. Math. Anal. Applic.*, **190**, 165–193.
- North, G. R., Wang, J. and Genton, M. G. (2011) Correlation models for temperature fields. Submitted to *J. Clim.*
- Nychka, D. (1995) Splines as local smoothers. *Ann. Statist.*, **23**, 1175–1197.
- Obukhov, O. M. (1947) Statistically homogenous random fields on the globe. *Usp. Mat. Nauk.*, **2**, 196–198.
- Pardo-Igúzquiza, E. and Dowd, P. A. (1997) AMLE3D: a computer program for the statistical inference of covariance parameters by approximate maximum likelihood estimation. *Comput. Geosci.*, **7**, 793–805.
- Pardo-Igúzquiza, E., Mardia, K. V. and Chica-Olmo, M. (2008) MLMATERN: a computer program for maximum likelihood inference with the spatial Matern covariance model. *Comput. Geosci.*, **35**, 1139–1150.
- Piterbarg, V. I. (1996) *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. Providence: American Mathematical Society.
- Ramsay, T. (2002) Spline smoothing over difficult regions. *J. R. Statist. Soc. B*, **64**, 307–319.
- Roberts, G. O. and Tweedie, R. L. (1996) Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli*, **2**, 341–363.



- Rosanol, J. A. (1977) Markov random fields and stochastic partial differential equations. *Math. USSR Sbor.*, **32**, 515.
- Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton: 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.
- Ruiz-Medina, M. D., Porcu, E. and Fernández-Pascual, R. (2011) The Dagum and auxiliary covariance families: towards reconciling two-parameter models that separate fractal dimension and the Hurst effect. *Probab. Engng Mech.*, **26**, 259–268.
- Sampson, P. D. (2010) Constructions for nonstationary spatial processes. In *Handbook of Spatial Statistics* (eds A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp), pp. 119–130. Boca Raton: CRC Press.
- Sampson, P. D. and Guttorp, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Statist. Ass.*, **87**, 108–119.
- Santner, T., Williams, B. and Notz, W. (2003) *The Design and Analysis of Computer Experiments*. New York: Springer.
- Schmidt, A. M., Guttorp, P. and O’Hagan, A. (2011) Considering covariates in the covariance structure of spatial processes. *Environmetrics*, **22**, 487–500.
- Schmidt, A. M. and O’Hagan, A. (2003) Bayesian inference for non-stationary spatial covariance structures via spatial deformations. *J. R. Statist. Soc. B*, **65**, 743–758.
- Simpson, D., Lindgren, F. and Rue, H. (2010) In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Preprint Statistics 16/2010*. Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim.
- Sokolov, I. M. (2000) Lévy flights from a continuous-time process. *Phys. Rev. E*, **63**, article 011104.
- Stein, M. L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.
- Stein, M. L. (2005) Nonstationary spatial covariance functions. *Technical Report 21*. Center for Integrating Statistical and Environmental Science, University of Chicago, Chicago. (Available from <http://www.stat.uchicago.edu/cises/research/cises-tr21.pdf>.)
- Sun, Y., Li, B. and Genton, M. G. (2011) Geostatistics for large datasets. In *Advances and Challenges in Space-time Modelling of Natural Events* (eds J. M. Montero, E. Porcu and M. Schlather). New York: Springer. To be published.
- Vapnik, B. (1998) *Statistical Learning Theory*. New York: Wiley.
- Vecchia, A. V. (1988) Estimation and model identification for continuous spatial process. *J. R. Statist. Soc. B*, **50**, 297–312.
- Wegener, J. (2010) Noise convolution models: fluids in stochastic motion, non-Gaussian tempo-spatial fields, and a notion of tilting. *Doctoral Thesis*. Lund University, Lund.
- Wendland, H. (2005) *Scattered Data Approximation*. New York: Cambridge University Press.
- Whittle, P. (1954) On stationary processes in the plane. *Biometrika*, **41**, 434–449.
- Wikle, C. K. (2003) Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, **84**, 1382–1394.
- Wikle, C. K. and Holan, S. H. (2011) Polynomial nonlinear spatio-temporal integro-difference equation models. *J. Time Ser. Anal.*, to be published.
- Wikle, C. K. and Hooten, M. B. (2006) Hierarchical Bayesian spatio-temporal models for population spread. In *Applications of Computational Statistics in the Environmental Sciences: Hierarchical Bayes and MCMC Methods* (eds J. S. Clark and A. Gelfand), pp. 145–169. Oxford: Oxford University Press.
- Wikle, C. K. and Hooten, M. B. (2010) A general science-based framework for nonlinear spatiotemporal dynamical models. *Test*, **19**, 417–451.
- Wikle, C. K., Milliff, R. F., Nychka, D. and Berliner, L. M. (2001) Spatiotemporal hierarchical Bayesian modeling: tropical ocean surface winds. *J. Am. Statist. Ass.*, **96**, 382–397.
- Wiktorsson, M. (2002) Simulation of stochastic integrals with respect to Lévy processes of type G. *Stoch. Processes Appl.*, **101**, 113–125.
- Wood, S. N., Bravington, M. V. and Hedley, S. L. (2008) Soap film smoothing. *J. R. Statist. Soc. B*, **70**, 931–955.
- Wu, Y.-C., Liao, S.-L. and Shyu, S.-D. (2009) Closed-form valuations of basket options using a multivariate normal inverse Gaussian model. *Insur. Math. Econ.*, **44**, 95–102.
- Yadrenko, M. (1983) *Spectral Theory of Random Fields*. New York: Optimization Software.
- Zhang, H. (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Am. Statist. Ass.*, **99**, 250–261.
- Zhu, Z. and Stein, M. L. (2006) Spatial sampling design for prediction with estimated parameters. *J. Agric. Biol. Environ. Statist.*, **11**, 24–44.
- Zimmerman, D. L. (2006) Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, **17**, 635–652.