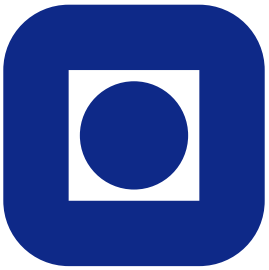
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

In order to make spatial statistics computationally feasible, we need to forget about the covariance function

by

Daniel Simpson, Finn Lindgren and Håvard Rue

PREPRINT
STATISTICS NO. 16/2010



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL
http://www.math.ntnu.no/preprint/statistics/2010/S16-2010.pdf
Håvard Rue has homepage: http://www.math.ntnu.no/∼hrue
E-mail: hrue@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

In order to make spatial statistics computationally feasible, we need to forget about the covariance function

Daniel Simpson, Finn Lindgren & Håvard Rue*
Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim, Norway

November 24, 2010

Abstract

Gaussian random fields (GRFs) are the most common way of modelling structured spatial random effects in spatial statistics. Unfortunately, their high computational cost renders the direct use of GRFs impractical for large problems and approximations are commonly used. In this paper we compare two approximations to GRFs with Matérn covariance functions: the kernel convolution approximation and the Gaussian Markov random field representation of an associated stochastic partial differential equation. We show that the second approach is a natural way to tackle the problem and is better than methods based on approximating the kernel convolution.

1 Introduction

Problems that have a non-negligible spatial component are ubiquitous in environmental statistics. In contrast to traditional statistical modelling, practical problems in spatial statistics are, by and large, *computational* in nature. Most applications feature large sets of data collected at irregular locations, which necessitates the development of fast and efficient methods for computing both the point estimates and the associated uncertainties of the parameters in the model.

Throughout this paper, we will consider the following generic scenario. Assume that we have observed some data $\{y_i\}_{i=1}^N$ at some spatial locations $\{s_i\}_{i=1}^N$ and that we have a hierarchical model for this observation process (Diggle and Ribeiro, 2006)

$$\begin{aligned} y_i|x_i, \boldsymbol{\theta} &\overset{i.i.d.}{\sim} \pi(y_i|x_i, \boldsymbol{\theta}) \\ \mathbf{x} &\sim \mathrm{N}\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta})\right) \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}), \end{aligned}$$

*Corresponding author. Email: hrue@math.ntnu.no

where $\boldsymbol{\theta}$ is a vector of model parameters; $\boldsymbol{\mu}(\boldsymbol{\theta})$ is a model for the mean of the underlying, unobserved spatial process $\mathbf{x}$; $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the covariance matrix of $\mathbf{x}$; and $\pi(\boldsymbol{\theta})$ is a prior on the vector of parameters. In most applications, we are interested in the posterior distribution of parameters $(\boldsymbol{\theta}, \mathbf{x})$ given the data $\mathbf{y}$. An application of Bayes' formula shows that the posterior is given by

$$\pi(\boldsymbol{\theta}, \mathbf{x}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \prod_{j=1}^{N} \pi(y_j|x_j, \boldsymbol{\theta}).$$

In most situations, the posterior will not be of a standard form and it is necessary to investigate it numerically.

The standard method for specifying the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is to define it as

$$\Sigma_{ij} = c(s_i, s_j),$$

where $c(s_i, s_j)$ is a covariance function, which forces the covariance matrix to be positive semi-definite. In general, finding useful covariance functions is difficult and typically it is assumed that the covariance between two points only depends on the distance between the points (or in the nomenclature of spatial statistics, that $\mathbf{x}$ is a stationary, isotropic random field). The class of Gaussian random fields that we are focussing on in this paper are the Matérn random fields, which are stationary, isotropic random fields with covariance functions given by[1]

$$c_\nu(s_i, s_j) = \frac{\sigma^2}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}2^{\nu-1}} \left(\kappa \left\|s_i - s_j\right\|\right)^\nu K_\nu \left(\kappa \left\|s_i - s_j\right\|\right) \tag{1}$$

where $\kappa, \sigma^2, \nu > 0$. It can be shown that the parameter ν defines the smoothness of the random field, whereas κ is a scale parameter and σ^2 is a variance parameter. Matérn random fields are popular models in many areas of applied statistics where the ability to specify the smoothness of the field is useful. Furthermore, the Matérn family includes two of the most popular covariance models—the exponential covariance function, which occurs when $\nu = 1/2$, and the Gaussian covariance function, which is the limit case as $\nu \to \infty$.

The major problem with using non-compactly supported covariance functions (such as the Matérn covariance functions) is that the resulting covariance matrices are completely dense and, therefore, *inference with* (1) *has a computational complexity of* $\mathcal{O}(N^3)$*, where* N *is the number of data points.* This makes direct use of the Matérn covariance function manifestly unsuitable for practical inference on moderate to large datasets. This is, of course, not a new problem in spatial statistics and a number of approximations have been developed to overcome this problem such as covariance tapering (Furrer et al., 2006) and predictive process modelling (Banerjee et al., 2008).

Since the 1980s, there has been a great deal of work done on Gaussian Markov random fields (GMRFs), especially over graphs. In practical situations, this Markov

[1]The scaling that we are using for the Matérn covariance function is slightly non-standard. It has been chosen to reflect the solution to the stochastic partial differential equation defined below.

property essentially forces the precision matrix $\mathbf{Q} = \mathbf{\Sigma}^{-1}$ to be *sparse*. The sparsity of the precision matrix allows for computation to be performed in $\mathcal{O}(n^{3/2})$ operations, where n is the number of vertices in the graph, if the problem is two-dimensional (Rue and Held, 2005). This reduction in computational complexity allows for the solution of really quite large spatial models (for instance, a two-dimensional second order conditional auto-regressive (CAR) model with $\sim 200,000$ nodes). Furthermore, the Markov property allows us to use fast approximate inference methods, such as INLA (Rue et al., 2009), to compute with these models.

The Markov random fields that we are considering in this paper are the stationary solutions to stochastic partial differential equations (SPDEs) of the form

$$(\kappa^2 - \Delta)^{\alpha/2} x(s) = \sigma W(s), \qquad s \in \mathbb{R}^2, \tag{2}$$

where $\Delta = \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}$ is the Laplacian on $\mathbb{R}^2$; $W(s)$ is white noise (which in two dimensions can be thought of as the 'derivative' of the Brownian sheet); α is a positive integer related to the smoothness parameter in (1) by $\nu = \alpha - d/2$, where d is the dimension of the space; and κ and σ are the scale and variance parameters from the Matérn covariance function. In order to assure that $x(s)$ is an ordinary random field, we assume that $\alpha > d/2$, which forces the associated smoothness parameter ν to be positive. We note here that the model (2) can be easily extended in a number of ways to give non-stationary Gaussian random fields on the sphere and other more general spatial domains.

The major aim of this paper is to demonstrate that, for this class of models, methods based on the SPDE formulation aresignificantly faster for practical computation than methods based on the covariance function. For comparison, we will consider a *mathematically equivalent* method based on a convolution (or moving average) representation of the random field

$$x(s) = \int_{\mathbb{R}^2} k(s, s')\, dW(s'), \tag{3}$$

where $k(s, s') = c_\eta(s, s')$ is the Matérn covariance function with smoothness parameter $\eta = (\alpha - d)/2$. In Section 2, we will show that this convolution representation is equivalent to the solution of the SPDE (2). Method based on approximations to (3) are very popular in spatial statistics, as the kernel function $k(s, s')$ can be quite general and can be used to generate non-stationary, anisotropic random fields.

In the remainder of the paper, we will firstly review the link between the SPDE (2) and the convolution representation (3) and discuss the practical problems that are encountered when using convolution representations as a basis for a computational method. In Section 3 we will review the methods of Lindgren et al. (2010) for constructing a GMRF directly from (2). In the following section we will compare the accuracy and performance of the SPDE and convolution field methods. Finally, we will discuss the various extensions of the SPDE method and give reasons apart from computational efficiency that SPDEs are *good models* in spatial statistics.

2 Stochastic partial differential equations and convolution fields

In order to demonstrate the link between convolution representation of a Matérn field and the SPDE (2), we need to look at precisely what we mean by a *solution* to a SPDE[2]. There are many different concepts of a solution, but in this paper we will be mainly interested in the solution defined by Walsh (1986). In order to define a solution, we need a class of test functions that can be used to 'pick out' enough features of the equation to completely define it. The standard choice is to take the test functions to be the class of smooth functions that go to zero at infinity. We then define a solution to be any random field $x(s)$ that satisfies

$$\int_{\mathbb{R}^2} x(s)(\kappa^2 - \Delta)^{\alpha/2}\phi(s)\, ds = \sigma \int_{\mathbb{R}^2} \phi(s)\, dW(s), \tag{4}$$

for *every* smooth test function $\phi(s)$ (Walsh, 1986). The integral on the right hand side of (4) is, by definition, a Gaussian random variable with mean zero and variance $\int_{\mathbb{R}^2} |\phi(s)|^2\, ds$. The link between (4) and convolution fields can be seen by a careful choice of the test function $\phi(s)$: take $\phi(s)$ to be the solution to

$$(\kappa^2 - \Delta)^{\alpha/2}\phi(s) = \psi(s),$$

for some smooth test function $\psi(s)$. It follows from the basic properties of $(\kappa^2 - \Delta)^{\alpha/2}$ that $\phi(s)$ is a smooth function. If we use this $\phi(s)$ in (4), we get

$$\int_{\mathbb{R}^2} x(s)\psi(s)\, ds = \int_{\mathbb{R}^2} (\kappa^2 - \Delta)^{-\alpha/2}\psi(s)\, dW(s). \tag{5}$$

In order to make it the rest of the way to the convolution representation, we need to know what $(\kappa^2 - \Delta)^{-\alpha/2}\psi(s)$ looks like. Fortunately, as we are working in $\mathbb{R}^2$, Fourier transform methods can be used to find an integral representation of the solution, namely

$$(\kappa^2 - \Delta)^{-\alpha/2}\psi(s) = \int_{\mathbb{R}^2} c_\eta(s,t)\psi(t)\, dt,$$

where $c_\eta(s,t)$ is the Matérn covariance function (1) with smoothness parameter $\eta = (\alpha - d)/2$. The link between the Matérn covariance function and fractional PDEs was noted by Whittle (1963). Plugging this integral representation into (5) and changing the order of integration, we get

$$\int_{\mathbb{R}^2} x(s)\psi(s)\, ds = \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^2} c_\eta(s,t)\, dW(t) \right) \psi(s)\, ds,$$

for every smooth test functions $\psi(s)$. This is the weak form of the equation

$$x(s) = \int_{\mathbb{R}^2} c_\eta(s,t)\, dW(t), \tag{6}$$

which is our desired result, namely equation (3).

[2]This section is, necessarily, more technical than the other parts of this paper. The message is that the solution of (2) can be written as (6). This representation forms the basis for the computational method described in Section 2.1.

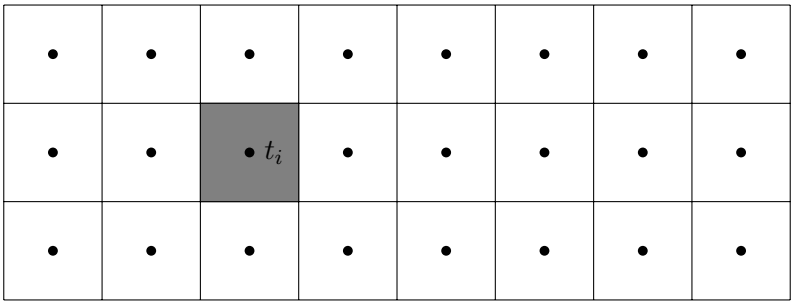


Figure 1: The typical design for the kernel approximation. Highlighted in grey is the box B_i and the centre point is t_i.

2.1 Computing with convolution representations

The typical way to use the convolution field representation (6) is to replace the integral by a sum

$$x(s) \approx \sum_{i=1}^{n} c_\eta(s, t_i)\xi_i$$

where t_i are the midpoints of the boxes B_i and $\xi_i \sim \mathrm{N}(0, |B_i|)$ are independent and n is the number of boxes (see Figure 1) (Xia and Gelfand, 2005). Unfortunately, this typical approach does not work for general Matérn fields—to see this we note that the kernel function $c_\eta(t, s)$ is singular if $\eta = (\alpha - d)/2 \leq 0$. This can occur as we have only assumed that $\alpha > d/2$ (which ensures the covariance itself isn't singular). As α is related to the smoothness of the random field (Bolin and Lindgren, To appear), this says that the typical convolution approach is only possible for random fields that have more than $d/2$ mean square continuous derivatives.

This problem can be rectified by using a more appropriate discretisation of (6). One possibility is to take

$$x(s) \approx \sum_{i=1}^{n} \left(\frac{1}{|B_i|} \int_{B_i} c_\eta(s, t)\, dt \right) \xi_i,$$

where the integral *smooths out* the singularity in the kernel function. Clearly, this is only a reasonable approach if the integral can be computed cheaply. In one dimension, where the ith box is $B_i = [t_{i,L}, t_{i,R}]$, this integral can be evaluated exactly in terms of modified Struve functions as

$$\frac{1}{t_{i,R} - t_{i,L}} \int_{t_{i,L}}^{t_{i,R}} c_\eta(s, t)\, dt = \frac{\sigma^2}{2\kappa^{2\eta} D} \left(s_L I(\eta, \kappa|s_L|) - s_R I(\eta, \kappa|s_R|) \right)$$

where $s_{L/R} = s - t_{i,L/R}$, $D = t_{i,R} - t_{i,L}$, $I(\eta, t) = K_\eta(t)L_{\eta-1}(t) + L_\eta(t)K_{\eta-1}(t)$, and $L_\zeta(t)$ is the modified Struve function with parameter ζ (Gradshteyn and Ryzhik, 1994, Equation 6.561.4). Figure 2 shows a kernel for $\eta = 1$ along with the corresponding smoothed version. We have been unable to analytically compute this integral for a two dimensional field.

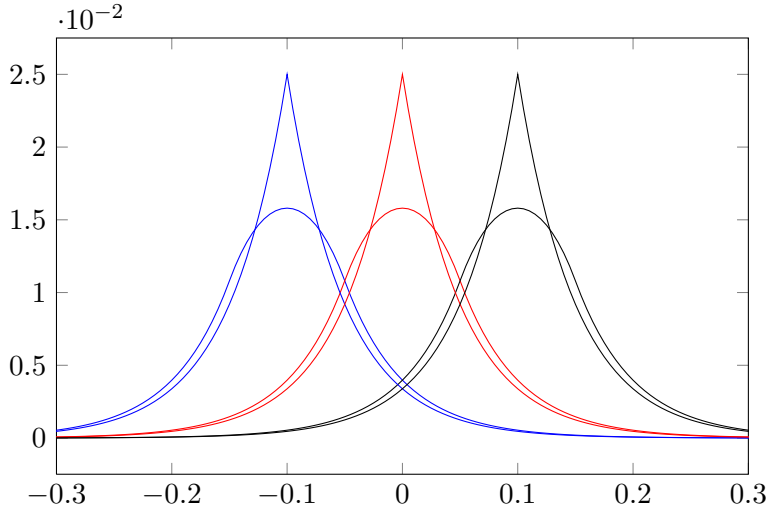


Figure 2: A comparison of the standard kernel function and the smoothed kernel function for $\alpha = 2$ ($\eta = 1/2$) and $\kappa = 20$.

A simple test when assessing the quality of a method for approximating a random function is to investigate how well it can represent simple deterministic functions. Figure 3 shows the best representation in each basis to the function $x(s) \equiv 1$, when $\alpha = 2$, $\kappa = 20$ and $n = 11$. There are two things immediately apparent from this figure: the simple kernel fails to reproduce constant functions, and there are significant edge effects when using kernel methods. The first problem can be partially alleviated by taking more points—when $n = 51$ the error continues to oscillate but its magnitude is reduced to around 1%. The second problem cannot be fixed, however the common work-around is to enlarge the region of interest. It can be seen that the area effected is of the same order of magnitude as the range of the random field (which is around $\sqrt{8\nu}/\kappa \approx 0.17$ Lindgren et al., 2010). The red line in Figure 3 is the best representation of $x(s) \equiv 1$ in the finite element basis (defined in Section 3). This basis, which *does not depend on the parameters of the random field* can exactly represent constant (as well as linear) functions.

The effect of smoothing the kernels can be easily seen in Figure 4. This shows a comparison of the Kriging estimate to a one dimensional function re-construction problem using the typical kernel approximation, the smoothed kernel approximation, and the GMRF representation (defined in the next section). In order to demonstrate the differences between the methods, we have taken only eleven equally spaced knots/basis functions at locations that do not correspond to the data. We have taken $\alpha = 2$, which results in a field with one continuous derivative. The instability of the common kernel approximation is immediately clear from Figure 4(a). We also note that error in the GMRF representation is comparable with the error from the smoothed kernel approximation and is never worse than the common kernel estimator. We also note that, assuming you can even compute it, the smoothed kernel approximation is *significantly*

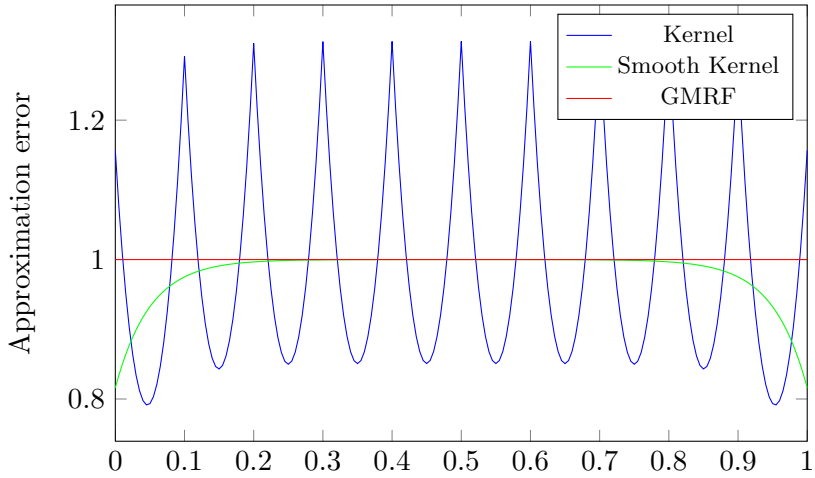


Figure 3: This figure shows the error in the approximations to $x(s) \equiv 1$ for all three sets of basis functions for $\alpha = 2$ and $\kappa = 20$. The blue line shows the simple kernel approximation. The smoothed kernel approximation (green line) behaves much better, although it does demonstrate edge effects. The finite element basis used for the GMRF representation of the SPDE (red line) reproduces constant functions exactly.

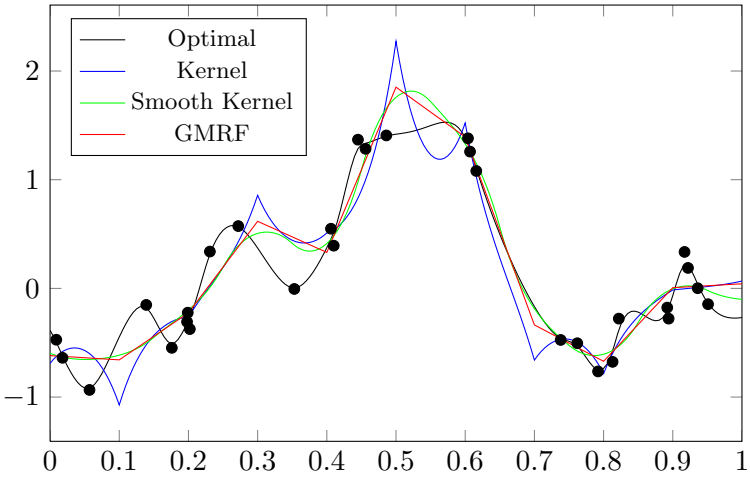
Dimension	Kernel	GMRF
1	$\mathcal{O}(Nn^2 + mn + n^3)$	$\mathcal{O}(N + m + n)$
2	$\mathcal{O}(Nn^2 + mn + n^3)$	$\mathcal{O}(N + m + n^{3/2})$

Table 1: Asymptotic operation counts for the calculation of the n weights for the Kriging estimator $\mathbb{E}(x(s)|Y, \theta) = \sum_{i=1}^{n} w_i \phi_i(s)$, where $\phi_i(s)$ are the basis functions. Here N is the number of data points and m is the number of kriging locations. It is assumed that $N > n$.

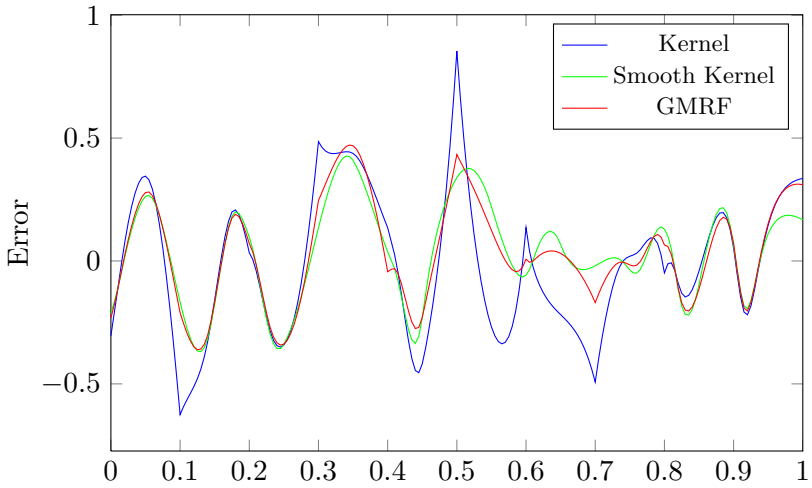
more computationally expensive than the GMRF representation—in terms of computational complexity, $\mathcal{O}(n^3)$ vs $\mathcal{O}(n)$, where n is the number of knots/basis functions. While the reduction in computational complexity in two dimensions is smaller—$\mathcal{O}(n^3)$ vs $\mathcal{O}(n^{3/2})$—it is more significant as n is typically much larger in two dimensions than it is in one (cf. the curse of dimensionality). Table 1 gives the asymptotic operation counts for both methods. It should be noted that if $N \ll n$ the kernel methods may be faster than the SPDE method.

3 GMRF representations of SPDEs

In this section we will outline the method of Lindgren et al. (2010) for constructing efficient GMRF representations of (2) for integer values of α. In the interests of clarity, we will particularly focus on the case $\alpha = 2$, which corresponds to a second order conditional autoregression. For the general case, we refer the interested reader



(a) Comparison of the Kriging estimates



(b) The deviation from the optimal Kriging estimates

Figure 4: The above figures show the one dimensional Kriging when $\alpha = 2$ and $\kappa = 20$ with 11 knots. The GMRF representation, detailed in Section 3 is almost as accurate as the smoothed kernel approximation and is much cheaper and much more straightforward to compute. The erratic behaviour of the kernel method can be clearly seen.

to Lindgren et al. (2010).

As we cannot represent $\mathbb{R}^2$ on a computer, the first thing that we need to do is restrict our attention to some bounded region D in $\mathbb{R}^2$, usually a rectangle. This rectangle needs to be chosen large enough to avoid any effects from the artificial boundary infecting the solution. Given such a bounded region D, the next step is to cover it in triangles in such a way that the n vertices are well distributed throughout D. It is important to note that the location of these vertices is in no way related to the location of the data points. With this in place, we will construct an approximate solution to (2) using the finite element method.

The key point in the finite element method is to replace the smooth test functions in (4) with N well chosen piecewise linear functions. For the ith vertex in our triangularisation, we define the test function $\phi_i(s)$ to be the piecewise linear function that is equal to one on vertex i and zero on all other vertices. Using these finite element test functions in (4) for $\alpha = 2$ leads to the linear equation

$$\left(\kappa^2 \tilde{\mathbf{C}}_n + \mathbf{G}_n\right) \tilde{\mathbf{x}}_n \stackrel{d}{=} \mathrm{N}(\mathbf{0}, \tilde{\mathbf{C}}_n),$$

where the matrices are given by

$$[\tilde{\mathbf{C}}_n]_{ij} = \int_D \phi_i(s)\phi_j(s)\, ds$$
$$[\mathbf{G}_n]_{ij} = \int_D \nabla\phi_i(s) \cdot \nabla\phi_j(s)\, ds.$$

These integrals can be computed explicitly (see, for example, Lindgren et al. (2010)) and are only non-zero if vertex i is a neighbour of vertex j. Therefore, all of these matrices are *sparse* (and, in fact, symmetric positive semi-definite), which makes them amenable to fast numerical methods.

In order to complete the approximation, we note that

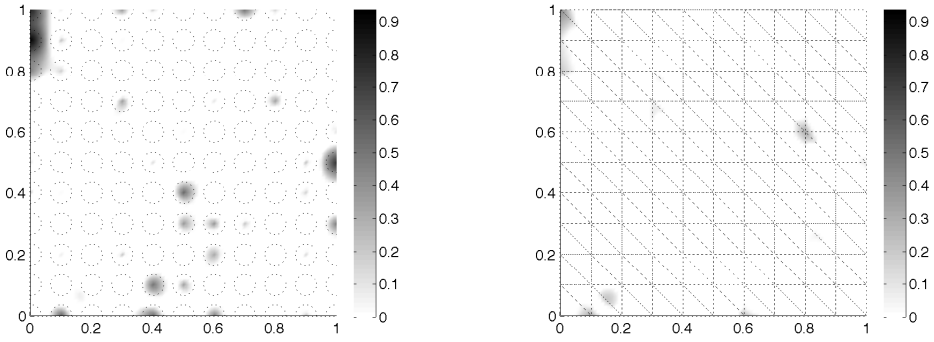
$$\tilde{\mathbf{x}}_n \stackrel{d}{=} \mathrm{N}(\mathbf{0}, \tilde{\mathbf{Q}}^{-1}),$$

where $\tilde{\mathbf{Q}} = \left(\kappa^2 \tilde{\mathbf{C}}_n + \mathbf{G}_n\right)^T \tilde{\mathbf{C}}_n^{-1} \left(\kappa^2 \tilde{\mathbf{C}}_n + \mathbf{G}_n\right)$. Unfortunately, as $\tilde{\mathbf{C}}_n$ is not a diagonal matrix, $\tilde{\mathbf{Q}}_n$ is not sparse, however we follow Lindgren et al. (2010) and replace $\tilde{\mathbf{C}}_n$ by the diagonal matrix $\mathbf{C}_n$ that has on its diagonal the row sums of $\tilde{\mathbf{C}}_n$. The resulting GMRF representation is

$$\mathbf{x}_n \stackrel{d}{=} \mathrm{N}(\mathbf{0}, \mathbf{Q}_n^{-1}), \tag{7}$$

where $\mathbf{Q}_n = \left(\kappa^2 \mathbf{C}_n + \mathbf{G}_n\right)^T \mathbf{C}_n^{-1} \left(\kappa^2 \mathbf{C}_n + \mathbf{G}_n\right)$.

A comparison of the kernel estimate and the GMRF representation for a two-dimensional kriging problem is presented in Figure 5. In this case, we have taken our parameters to be $\kappa = 20$ and $\alpha = 3$, which results in the field having two mean-squared derivatives. It is clear from Figure 5(a) that the error in the kernel estimator is strongly related to the location of the kernels. This does not happen in the GMRF case.



(a) The kernel approximation

(b) The GMRF representation

Figure 5: The above figures show the deviation from the optimal Kriging estimate for a two dimensional problem with $\alpha = 3$ and $\kappa = 20$ with 11 knots in each direction. The dashed circles in 5(a) show the location of the integration points t_i, while the dashed lines in 5(b) show the triangles over which the finite element approximation is calculated. The dark areas in each figure show the regions of large deviation from the optimal estimate.

4 Discussion

The main thesis of this paper is that *methods bases on GMRF approximations to SPDEs are superior to methods based on the corresponding kernel methods*. Of course, such blanket statements need to be unpacked carefully and this section is devoted to doing just that.

The restriction to the SPDE models. To begin with, this statement strongly restricts the class of random field models to models based on SPDEs and, in particular, to the Matérn class of random fields. It has been argued by others, in particular Stein (1999), that the Matérn class is the only class of covariance functions that you need for practical spatial statistics. We have further restricted the smoothness parameter in the Matérn field to be of the form $\nu = \alpha - d/2$ for an *integer* α. We need to critically assess whether or not this is a practical restriction—that is we must, for a given application, ask whether or not we truly need values of ν outside of this class. This is not a simple question to answer: for very large problems, the computational advantages of GMRFs make a very convincing argument for their use over kernel methods and, therefore, for restricting the smoothness parameter. This question is further complicated by the strong relationship between the scale and smoothness parameters (Zhang, 2004). In the end, the restriction on ν is not, in our opinion, a handicap, but rather it reflects the very real fact that there are no universally appropriate methods for hard computational problems.

What do we mean by a superior method? The discussion in the previous paragraph implies a definition of a 'superior' method. When saying that the GMRF rep-

resentation of the SPDE model is superior to the kernel approximation, we are really saying that it is more computationally efficient. This is a reasonable definition under the condition that the two methods compute pretty much the same thing (we will discuss this in the next paragraph)—we have, after all, predicated the entire discussion on the idea that "practical problems in spatial statistics are, by and large, computational in nature". By this criterion, the GMRF method is superior based simply on operation counts. This, along with the mathematical equivalence of the two methods, highlights an important tenet of computational statistics: *equivalent mathematical statements do not lead to equally useful computational algorithms.* Of course, real life is more complicated than operation counts and it is necessary to consider the availability of efficient software for these SPDE methods. This is discussed below.

Is the GMRF representation an *approximation* to the SPDE? There is an important issue that we skirted in Section 3: *does this representation converge to the solution of the SPDE?* The simple answer is yes, but a more detailed answer is in order. If we are to use the GMRF approximation to a Matérn field we would like to be able to control the approximation error. It can be shown using some finite element theory that the error in the approximation can be bounded above by some constant times h^2, where h is the diameter of the largest circle that can be inscribed in a triangle in the mesh (Lindgren et al., 2010). This tells us that, as long as the vertices of the triangles are distributed in a reasonable way over the domain, the approximation will be good. A similar requirement is needed in order to make the kernel method converge (Xia and Gelfand, 2005). Ideally, whenever using a kernel method or a GMRF approximation to an SPDE as a prior model, the sensitivity of the marginal posteriors to the computational mesh should be tested. Realistically however, we expect that the posterior is insensitive to small perturbations in the prior and simply treat the methods as exact. This assumption is necessary as it is common for the size of the prior field to be as large as is computationally feasible, which makes refining the mesh any further impossible.

There is a second way of thinking about the correctness of the GMRF representation that bypasses the inconvenient and technical discussion about convergence. The GMRF representation of the SPDE is (up to the diagonal approximation of $\tilde{\mathbf{C}}$) the *best* approximation to $x(s)$ over the space of piecewise linear functions defined over our triangularisation of D. From this point of view, we can simply define our triangles and feel comfortable in the knowledge that the we have the best possible GMRF representation of the full random field over these triangles.

The approximation properties of the Kernel methods are troubling. Figure 3 shows that the simple kernel approximation can behave very poorly. This implies that great care must be taken when using the kernel methods, especially with respect to the selection of integration points. Furthermore, the quality of these approximations depend strongly on the smoothness and scale parameters which, again, suggests caution is in order. Finite element basis functions do not share this problem and their approximation properties have been very well studied (Brenner and Scott, 2007).

The SPDE *means* something. There is a very strong conceptual advantage that SPDE methods have over kernel methods that we have not mentioned—the differential operator on the left hand side of (2) has a strong physical interpretation. The first advantage of this is pedagogical: it allows for a new point of departure when explaining these models to people with a strong physics, engineering or applied mathematics background. The second advantage is that the aforementioned group of people use partial differential equations to model all manner of phenomena and this knowledge can be incorporated into building more complex models of spatial dependence. This idea can be used to easily incorporate (spatially dependent) drift terms or anisotropy into the model. While this can also be done using convolution kernels, the SPDE methodology mirrors the way in which these processes are usually modelled in physics (Ockendon et al., 2003) and computational biology (Murray, 2003).

The SPDE is independent of geometry. The great theoretical advantage of defining a Gaussian random field via a stochastic partial differential equation is that the form of the SPDEs that we are considering does not depend on the underlying geometry of the physical space. This is a distinct advantage when compared with GRFs defined through covariance functions or kernel methods, in which the physical geometry is of vital importance. Furthermore, we note that the finite element method depends only on the *local* behaviour of the operator and the field and can therefore be applied on a complicated domain or even on a manifold (Lindgren et al., 2010). In particular Bolin and Lindgren (To appear) use the GMRF representation to define a random field on a sphere in the context of environmental modelling. We note that it is possible to define a kernel representation for a Matérn GRF on a sphere using similar arguments to those in Section 2, however the resulting kernel does not have a closed form[3]. It is also worth noting that the comments in the previous paragraph also hold on a manifold, that is the GMRF representation can be used to construct nonstationary, anisotroptic Gaussian random fields over general manifolds.

Is there software? The availability of freely available software is a vital part of the development of new methods in computational statistics. As such, we have implemented the GMRF representation described in this paper as part of the `R`–package `INLA` available from `http://r-inla.org`. Furthermore, the code for all of the examples in this paper, which were computed using `Matlab`, is available in the supplementary material.

References

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

[3]It follows from Fourier analysis that it will be an infinite series of spherical harmonics.

D. Bolin and F. Lindgren. Spatially models generated by nested stochastic partial differential equations. *Annals of Applied Statistics*, To appear.

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

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

R. Furrer, M. G. Genton, and D. Nychka. Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.*, 15(3):502–523, 2006.

I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Academic press New York, fifth edition, 1994.

F. Lindgren, J. Lindström, and H. Rue. An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach. Technical report 5, Department of mathematical sciences, Norwegian University of Science and Technology, 2010.

J.D. Murray. *Mathematical Biology: Spatial models and biomedical applications*. Springer Verlag, 2003. ISBN 0387952284.

J. Ockendon, S. Howison, A. Lacey, and A. Movchan. *Applied partial differential equations*. Oxford University Press, USA, 2003. ISBN 0198527713.

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

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

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

J. Walsh. An introduction to stochastic partial differential equations. *École d'Été de Probabilités de Saint Flour XIV-1984*, pages 265–439, 1986.

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

G. Xia and A.E. Gelfand. Stationary Process Approximation for the Analysis of Large Spatial Datasets. 2005.

H. Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261, 2004.