Citation for published version:
Lindgren, F & Rue, H 2015, 'Bayesian Spatial Modelling with R-INLA', *Journal of Statistical Software*, vol. 63, no. 19.

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Publisher Rights
CC BY

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 02. Oct. 2026



Journal of Statistical Software

January 2015, Volume 63, Issue 19. *http://www.jstatsoft.org/*

Bayesian Spatial Modelling with R-INLA

Finn Lindgren
University of Bath

Håvard Rue
Norwegian University of
Science and Technology

Abstract

The principles behind the interface to continuous domain spatial models in the R-**INLA** software package for R are described. The integrated nested Laplace approximation (INLA) approach proposed by Rue, Martino, and Chopin (2009) is a computationally effective alternative to MCMC for Bayesian inference. INLA is designed for latent Gaussian models, a very wide and flexible class of models ranging from (generalized) linear mixed to spatial and spatio-temporal models. Combined with the stochastic partial differential equation approach (SPDE, Lindgren, Rue, and Lindström 2011), one can accommodate all kinds of geographically referenced data, including areal and geostatistical ones, as well as spatial point process data. The implementation interface covers stationary spatial models, non-stationary spatial models, and also spatio-temporal models, and is applicable in epidemiology, ecology, environmental risk assessment, as well as general geostatistics.

Keywords: Bayesian inference, Gaussian Markov random fields, stochastic partial differential equations, Laplace approximation, R.

Traditionally, Markov models in image analysis and spatial statistics have been largely confined to discrete spatial domains, such as lattices and regional adjacency graphs. However, as discussed in Lindgren *et al.* (2011), one can express a large class of random field models as solutions to continuous domain stochastic partial differential equations (SPDEs), and write down explicit links between the parameters of each SPDE and the elements of precision matrices for weights in a discrete basis function representation. As shown by Whittle (1963), such models include those with Matérn covariance functions, which are ubiquitous in traditional spatial statistics, but in contrast to covariance based models it is far easier to introduce non-stationarity into the SPDE models. This is because the differential operators act locally, similarly to local increments in Gibbs-specifications of Markov models, and only mild regularity conditions are required. The practical significance of this is that classical Gaussian random fields can be merged with methods based on the Markov property, providing continuous domain models that are computationally efficient, and where the parameters can be

specified locally without having to worry about positive definiteness of covariance functions.

The fundamental building block of such Gaussian Markov random field (GMRF) models, as implemented in R-**INLA**, is a high-dimensional basis representation, with simple local basis functions. This is in contrast to fixed rank Kriging (Cressie and Johannesson 2008) that typically uses a smaller number of global basis functions, and predictive process methods (Banerjee, Gelfand, Finley, and Sang 2008). See Wikle (2010) of for an overview of such low-rank representation methods. A numerical comparison of the error introduced in Kriging calculations was performed by Bolin and Lindgren (2013) for SPDE based GMRF models, covariance tapering, and process convolutions. A non-parametric approach using similar GMRF models is available in the **LatticeKrig** package on CRAN (Nychka, Hammerling, Sain, and Lenssen 2013).

The different methods can also be combined, although the details for doing that within R-**INLA** are beyond the scope of this paper. For example, the global temperature analysis in Lindgren *et al.* (2011) used a combination of a low dimensional global basis, like in fixed rank Kriging, and a small-scale GMRF process, both with priors based on approximations to continuous domain SPDE models. There is considerable overlap between models formulated using fixed rank Kriging and SPDE/GMRF models, and a clear line separating the methods cannot be drawn.

The R-**INLA** software package currently has direct support for stationary and non-stationary locally isotropic SPDE/GMRF models on compact subsets of $\mathbb{R}$, $\mathbb{R}^2$, and on $\mathbb{S}^2$, as well as separable space-time models. Some non-separable space-time models, non-stationary fully anisotropic models, as well as models on $\mathbb{R}^3$ and other user-defined domains are also partially supported by the internal implementation, but have not yet been added to the basic interface. Consequently, auto-regressive models (see e.g., Cameletti, Lindgren, Simpson, and Rue 2013) are fully supported, but anisotropic advection-diffusion models (see e.g., Sigrist, Künsch, and Stahel 2015) are limited to non-advective models and require advanced user interaction, but support non-stationary anisotropy if only the strength of the non-isotropic diffusion is unknown.

The following sections present the basic ingredients of the link between continuous domains and Markov models and related simulation free Bayesian inference methods (Section 1), describe the structure of the interface to using such models in the R-**INLA** software package (Section 2 and 3), and discuss planned future development (Section 4). Special emphasis is placed on the abstractions necessary to simplify the practical bookkeeping for the users of the software. For further details on the computational and inferential methods in the R-**INLA** package we refer to Martins, Simpson, Lindgren, and Rue (2013).

1. Spatial modelling and inference

This section describes the basic principles of the continuous domain spatial models and Bayesian inference methods in the R-**INLA** package (Rue, Martino, Lindgren, Simpson, and Riebler 2013b).

1.1. Continuous domain spatial Markov random fields

When building and using hierarchical models with latent random fields it is important to remember that the latent fields often represent real-world phenomena that exist independently

of whether they are observed in a given location or not. Thus, we are not building models solely for *discretely observed data*, but for approximations of *entire processes* defined on continuous domains. For a spatial field $x(\boldsymbol{s})$, while the data likelihood typically depends only on the values at a finite set of locations, $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\}$, the model itself defines the joint behaviour for all locations, typically $\boldsymbol{s} \in \mathbb{R}^2$ or $\boldsymbol{s} \in \mathbb{S}^2$ (a sphere/globe). In the case of lattice data, the discretisation typically happens in the observation stage, such as integration over grid boxes (e.g., photon collection in a camera sensor). Often, this is approximated by point-wise evaluation, but there is nothing apart from computational challenges preventing other observation models.

As discussed in the introduction, an alternative to traditional covariance based modelling is to use SPDEs, but carry out the practical computations using Gaussian Markov random field (GMRF) representations. This is done by approximating the full set of spatial random functions with weighted sums of simple basis functions, which allows us to hold on to the continuous interpretation of space, while the computational algorithms only see discrete structures with Markov properties. Beyond the main paper Lindgren *et al.* (2011), this is further discussed by Simpson, Lindgren, and Rue (2012a,b).

Stationary Matérn fields

The simplest model for $x(\boldsymbol{s})$ currently implemented in R-**INLA** is the SPDE/GMRF version of the stationary Matérn family, obtained as the stationary solutions to

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega,$$

where Δ is the Laplacian, κ is the spatial scale parameter, α controls the smoothness of the realisations, τ controls the variance, and Ω is the spatial domain. The right-hand side of the equation, $\mathcal{W}(\boldsymbol{s})$, is a Gaussian spatial white noise process. As noted by Whittle (1954, 1963), the stationary solutions on $\mathbb{R}^d$ have Matérn covariances,

$$\mathsf{COV}(x(\mathbf{0}), x(\boldsymbol{s})) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa\|s\|)^\nu K_\nu(\kappa\|s\|). \tag{1}$$

The parameters in the two formulations are coupled so that the Matérn smoothness is $\nu = \alpha - d/2$ and the marginal variance is

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

From this we can identify the exponential covariance with $\nu = 1/2$. For $d = 1$, this is obtained with $\alpha = 1$, and for $d = 2$ with $\alpha = 3/2$.

From spectral theory one can show that integer values for α gives continuous domain Markov fields (Rozanov 1982), and these are the easiest for which to provide discrete basis representations. In R-**INLA**, the default value is $\alpha = 2$, but $0 \leq \alpha < 2$ are also available, though not as extensively tested. For the non-integer α values the approximation method introduced in the authors' discussion response in Lindgren *et al.* (2011) is used. Historically, Whittle (1954) argued that $\alpha = 2$ was a more natural basic choice for $d = 2$ models than the fractional $\alpha = 3/2$ alternative. Note that fields with $\alpha \leq d/2$ have $\nu \leq 0$ and that such fields have no point-wise interpretation, although they have well-defined integration properties. In particular, this means that the case $d = 2$, $\alpha = 1$, which on a regular lattice discretisation

corresponds to the common CAR(1) model, needs to be interpreted with care (Besag 1981; Besag and Mondal 2005), especially when used in combination with irregular discretisation domains.

The models discussed in Lindgren *et al.* (2011) and implemented in R-**INLA** are built on a basis representation

$$x(\boldsymbol{s}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{s}) x_k, \tag{3}$$

where $\psi_k(\cdot)$ are deterministic basis functions, and the joint distribution of the weight vector $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ is chosen so that the distribution of the functions $x(\boldsymbol{s})$ approximates the distribution of solutions to the SPDE on the domain. To obtain a Markov structure, and to preserve it when conditioning on local observations, we use piecewise polynomial basis functions with compact support. The construction is done by projecting the SPDE onto the basis representation in what is essentially a Finite Element method.

To allow easy and explicit evaluation, for two-dimensional domains we use piece-wise linear basis functions defined by a triangulation of the domain of interest. For one-dimensional domains, B-splines of degrees 1 (piecewise linear) and 2 (piecewise quadratic) are supported. This yields sparse matrices $\boldsymbol{C}$, $\boldsymbol{G}_1$, and $\boldsymbol{G}_2$ such that the appropriate precision matrix for the weights is given by

$$\boldsymbol{Q} = \tau^2(\kappa^4 \boldsymbol{C} + 2\kappa^2 \boldsymbol{G}_1 + \boldsymbol{G}_2)$$

for the default case $\alpha = 2$, so that the elements of $\boldsymbol{Q}$ have explicit expressions as functions of κ and τ. Assigning the Gaussian distribution $\boldsymbol{x} \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1})$ now generates continuously defined functions $x(\boldsymbol{s})$ that are approximative solutions to the SPDE (in a stochastically weak sense).

The simplest internal representation of the parameters in the model interface is $\log(\tau) = \theta_1$ and $\log(\kappa) = \theta_2$, where θ_1 and θ_2 are assigned a joint normal prior distribution. Since τ and κ have a joint influence on the marginal variances of the resulting field, it is often more natural to construct the parameter model using the standard deviation σ and *range* ρ, where $\rho = (8\nu)^{1/2}/\kappa$ is the distance for which the correlation functions have fallen to approximately 0.13, for all $\nu > 1/2$. Another commonly used definition for the range is as the distance at which the correlation is 0.05. The alternative definition used in R-**INLA** has the advantage of explicit dependence on ν. Translating this into τ and κ yields

$$\log \tau = \frac{1}{2} \log \left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}} \right) - \log \sigma - \nu \log \kappa, \tag{4}$$

$$\log \kappa = \frac{\log(8\nu)}{2} - \log \rho. \tag{5}$$

Suppose we want a parameterisation

$$\log \sigma = \log \sigma_0 + \theta_1, \tag{6}$$

$$\log \rho = \log \rho_0 + \theta_2, \tag{7}$$

where σ_0 and ρ_0 are base-line standard deviation and range values. We then substitute $\log\sigma$ and $\log\rho$ into Equation 4 and 5, giving the internal parameterisation

$$\begin{aligned}
\log\kappa_0 &= \frac{\log(8\nu)}{2} - \log\rho_0, \\
\log\tau_0 &= \frac{1}{2}\log\left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}}\right) - \log\sigma_0 - \nu\log\kappa_0, \\
\log\tau &= \log\tau_0 - \theta_1 + \nu\theta_2, \\
\log\kappa &= \log\kappa_0 - \theta_2,
\end{aligned}$$

where now the θ_1 and θ_2 parameters jointly control the τ-parameter.

Non-stationary fields

There is a vast range of possible extensions to the stationary SPDE described in the previous section, including non-stationary versions (see Lindgren *et al.* 2011; Bolin and Lindgren 2011, for examples). In the current version of the package, a non-stationary model defined via spatially varying $\kappa(\boldsymbol{s})$ and $\tau(\boldsymbol{s})$ is available for the case $\alpha = 2$. The SPDE is defined as

$$(\kappa(\boldsymbol{s})^2 - \Delta)(\tau(\boldsymbol{s})x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega,$$

and $\log\kappa(\boldsymbol{s})$ and $\log\tau(\boldsymbol{s})$ are defined as linear combinations of basis functions,

$$\begin{aligned}
\log(\tau(\boldsymbol{s})) &= b_0^\tau(\boldsymbol{s}) + \sum_{k=1}^{p} b_k^\tau(\boldsymbol{s})\theta_k, \\
\log(\kappa(\boldsymbol{s})) &= b_0^\kappa(\boldsymbol{s}) + \sum_{k=1}^{p} b_k^\kappa(\boldsymbol{s})\theta_k,
\end{aligned}$$

where $\{\theta_1, \ldots, \theta_p\}$ is a common set of internal representation parameters, and $b_k^\tau(\cdot)$ and $b_k^\kappa(\cdot)$ are spatial basis functions, some of which, for each k may be identically zero for either τ or κ. The precision matrix for the discrete field representation weights is a simple modification of the stationary one, with the parameter fields (evaluated at the mesh discretisation points) entering via diagonal matrices:

$$\begin{aligned}
\boldsymbol{T} &= \mathrm{diag}(\tau(\boldsymbol{s}_k)), \\
\boldsymbol{K} &= \mathrm{diag}(\kappa(\boldsymbol{s}_k)), \\
\boldsymbol{Q} &= \boldsymbol{T}(\boldsymbol{K}^2\boldsymbol{C}\boldsymbol{K}^2 + \boldsymbol{K}^2\boldsymbol{G}_1 + \boldsymbol{G}_1^\top\boldsymbol{K}^2 + \boldsymbol{G}_2)\boldsymbol{T}.
\end{aligned}$$

Just as in the stationary case, the model can be reparameterised using Equation 4 and 5, where

$$\begin{aligned}
\log(\sigma(\boldsymbol{s})) &= b_0^\sigma(\boldsymbol{s}) + \sum_{k=1}^{p} b_k^\sigma(\boldsymbol{s})\theta_k, \\
\log(\rho(\boldsymbol{s})) &= b_0^\rho(\boldsymbol{s}) + \sum_{k=1}^{p} b_k^\rho(\boldsymbol{s})\theta_k,
\end{aligned}$$

and $\sigma(\boldsymbol{s})$ and $\rho(\boldsymbol{s})$ are the *nominal* local standard deviations and correlation ranges. There are no explicit expressions for the actual values, since they depend on the entire parameter

functions in a non-trivial way. For given values of $\boldsymbol{\theta}$, the marginal variances can be efficiently calculated using the discretised GMRF representation, see Section 2.3.

Given the offsets, $b_0^\sigma(\boldsymbol{s})$ and $b_0^\kappa(\boldsymbol{s})$, and basis functions. $b_k^\sigma(\boldsymbol{s})$ and $b_k^\kappa(\boldsymbol{s})$, for the $\log(\sigma(\boldsymbol{s}))$ and $\log(\rho(\boldsymbol{s}))$ parameter fields, the internal model representation can be constructed using the following identities:

$$b_0^\kappa(\boldsymbol{s}) = \frac{\log(8\nu)}{2} - b_0^\rho(\boldsymbol{s}), \tag{8}$$

$$b_k^\kappa(\boldsymbol{s}) = -b_k^\rho(\boldsymbol{s}), \tag{9}$$

$$b_0^\tau(\boldsymbol{s}) = \frac{1}{2}\log\left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}}\right) - b_0^\sigma(\boldsymbol{s}) - \nu b_0^\kappa(\boldsymbol{s}), \tag{10}$$

$$b_k^\tau(\boldsymbol{s}) = -b_k^\sigma(\boldsymbol{s}) - \nu b_k^\kappa(\boldsymbol{s}). \tag{11}$$

The constant $\Gamma(\nu)/(\Gamma(\alpha)(4\pi)^{d/2})$ is $1/2$ and $1/4$ for $d = 1$, $\alpha = 1$ and 2. For $d = 2$ and $\alpha = 2$ it is $1/(4\pi)$. There is experimental support for constructing basis functions that reduces the influence of the range on the variance for cases where the basis functions for $\log \rho(\boldsymbol{s})$ have rapid changes.

Boundary effects

When constructing solutions to the SPDEs on bounded domains, boundary conditions are imposed, but how to construct practical and proper stochastic boundary conditions for these models is an open research problem. In the current version of the package, all 2D models are restricted to deterministic Neumann boundaries (zero normal-derivatives), as this is easy to construct, has well defined physical interpretation in terms of *reflection*, and has an effect on the covariances that is easy to quantify. As a rule of thumb, the boundary effect is negligible at a distance ρ from the boundary, and the variance is inflated near the boundary by a factor 2 along straight boundaries, and by a factor 4 near right-angled corners. In practice one can therefore avoid the boundary effect by extending the domain of interest by a distance at least ρ, as well as avoid sharp corners. The built-in mesh generation routines (see Section 2.1) are designed to do this. For one-dimensional models, the boundaries can also be defined as Dirichlet (value zero at the boundary), free, or cyclic.

Space-time models

While no space-time models are currently implemented explicitly, it is possible to construct such models using general code features. The most important method is to construct a Kronecker product model. Starting from a basis representation

$$x(\boldsymbol{s}, t) = \sum_k \psi_k(\boldsymbol{s}, t) x_k,$$

where each basis function is the product of a spatial and a temporal basis function, $\psi_k(\boldsymbol{s}, t) = \psi_i^{\boldsymbol{s}}(\boldsymbol{s})\psi_j^t(t)$, the space-time SPDE

$$\frac{\partial}{\partial t}(\kappa(\boldsymbol{s})^2 - \Delta)^{\alpha/2}\left(\tau(\boldsymbol{s})x(\boldsymbol{s}, t)\right) = \mathcal{W}(\boldsymbol{s}, t), \quad (\boldsymbol{s}, t) \in \Omega \times \mathbb{R}$$

generates a precision matrix for the weight vector $\boldsymbol{x}$ as $\boldsymbol{Q} = \boldsymbol{Q}_t \otimes \boldsymbol{Q}_{\boldsymbol{s}}$, where $\boldsymbol{Q}_{\boldsymbol{s}}$ is the precision for the previous purely spatial model, and $\boldsymbol{Q}_t$ is the precision corresponding to a

one-dimensional random walk. Any temporal GMRF model can be used in this construction, and Section 3 contains examples for how to specify such models in R-**INLA**. See Cameletti *et al.* (2013) for a case-study using a Kronecker model based on a temporal AR(1) process, including the full R code (although note that the interface has evolved slightly since the case-study was implemented).

Kronecker models generate separable covariance functions, which are simple but often unrealistic. The internal representation of the SPDE precision structures however also permits construction of non-separable models, as long as the unknown parameters appear in the appropriate places. Non-separable models that can be constructed in this way include special cases of the stochastic heat equation. Wrapper functions for constructing such models are expected to be added in the future, as well as extensions for advection-diffusion models.

1.2. Bayesian inference

The R-**INLA** package (Rue *et al.* 2013b) implements the integrated nested Laplace approximation (INLA) method introduced by Rue *et al.* (2009). This method performs direct numerical calculation of posterior densities in a large latent Gaussian (LGM) sub-class of Bayesian hierarchical models, avoiding time-consuming Markov chain Monte Carlo simulations. The implementation covers models of the following form,

$$(\boldsymbol{\theta}) \sim p(\boldsymbol{\theta}) \tag{12}$$

$$(\boldsymbol{x} \mid \boldsymbol{\theta}) \sim \mathsf{N}(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}) \tag{13}$$

$$\eta_i = \sum_j c_{ij} x_j \tag{14}$$

$$(y_i \mid \boldsymbol{x}, \boldsymbol{\theta}) \sim p(y_i \mid \eta_i, \boldsymbol{\theta}) \tag{15}$$

$\boldsymbol{\theta}$ are (hyper)parameters, $\boldsymbol{x}$ is a latent Gaussian field, $\boldsymbol{\eta}$ is a linear predictor based on known covariate values c_{ij}, and $\boldsymbol{y}$ is a data vector,

The basic principle is to approximate the posterior density for $(\boldsymbol{\theta} \mid \boldsymbol{y})$ using a Gaussian approximation $\widetilde{p}(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$ for the posterior for the latent field, evaluated at the posterior mode, $\boldsymbol{x}^*(\boldsymbol{\theta}) = \underset{\boldsymbol{x}}{\operatorname{argmax}}\, p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$,

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \left.\frac{p(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})} \approx \left.\frac{p(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})}{\widetilde{p}(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})},$$

which is called a Laplace approximation. This allows approximate evaluation of the (unnormalised) posterior density for $\boldsymbol{\theta}$ at any point. The algorithm uses numerical optimisation to find the mode of the posterior. The marginal posteriors for each θ_k and x_j are then calculated using numerical integration over $\boldsymbol{\theta}$, with another Laplace approximation involved in the latent field marginal posterior calculations:

$$p(\theta_k \mid \boldsymbol{y}) \approx \int \widetilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y})\, \mathrm{d}\boldsymbol{\theta}_{-k},$$

$$p(x_j \mid \boldsymbol{y}) \approx \int \widetilde{p}(x_j \mid \boldsymbol{\theta}, \boldsymbol{y})\, \widetilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y})\, \mathrm{d}\boldsymbol{\theta}$$

For more details and information, see Martins *et al.* (2013).

The linear predictor

An important aspect of the package interface is how to specify the connection between the latent field $\boldsymbol{x}$ and the linear predictor $\boldsymbol{\eta}$. R-**INLA** uses a formula syntax similar to the standard one used for linear model estimation with `lm()`. The main difference is that random effects of all kinds, including smooth nonlinear effects, structured graph effects and spatial effects, are specified using terms `f()`, where the user specifies the properties of such effects. The first argument of each `f()` specifies what element of the latent effect should apply to each observation, either as a scalar location of covariate value (for smooth nonlinear effects) or as a direct component index.

The linear predictor $\eta_i = \text{fixed}_i\beta + f(\text{time}_i) + \text{random}_i$ can be constructed using the formula

```
~ -1 + fixed + f(time, model = "rw2") + f(random, model = "iid")
```

Let $\boldsymbol{z}^{(1)}$ and $\boldsymbol{z}^{(2)}$ denote the covariate values for the `fixed` and `time` effects, and let $\boldsymbol{z}^{(3)}$ denote random effect indices. We can then rewrite the linear model from Equation 14 using mapping functions $h_k(\cdot)$ that denote the mapping from covariates or indices to the actual latent value for formula component k,

$$\begin{aligned} h_1(z_i^{(1)}) &= z_i^{(1)}\beta \\ h_2(z_i^{(2)}) &= \text{smooth effect evaluated at } z_i^{(2)} \\ h_3(z_i^{(3)}) &= \text{random effect component number } z_i^{(3)} \\ \eta_i &= \sum_k h_k(z_i^{(k)}) \end{aligned}$$

The latent field is the joint vector of all latent Gaussian variables, including the linear covariate effect coefficient β. For missing values in the z-vectors, the h functions are defined to be zero.

Since this construction only allows each observation to directly depend on a single element from each $h_k(\cdot)$ effect, this does not cover the case when an effect is defined using a basis expansion such as Equation 3. To solve this, R-**INLA** can apply a second layer of linear combinations to the $\boldsymbol{\eta}$ predictor,

$$\boldsymbol{\eta}^* = \boldsymbol{A}\boldsymbol{\eta}, \tag{16}$$

where $\boldsymbol{A}$ is a user-defined sparse matrix. This allows the SPDE models to be treated as indexed random effects, and the mapping between the basis weights and function values is done by placing appropriate $\psi_j(\boldsymbol{s})$ values in the $\boldsymbol{A}$ matrix. Whenever an $\boldsymbol{A}$ matrix is used, the elements of the $\boldsymbol{\eta}^*$ vector are the linear predictor values used in the general observation model in Equation 15, instead of $\boldsymbol{\eta}$. This is further formalised in Section 2.5.

See Section 2.2 for how to construct the part of the $\boldsymbol{A}$ matrix needed for an SPDE model, and Section 2.5 for how to set up the joint matrices needed for general models.

2. R interface

The R-**INLA** interface to the SPDE models described in the previous section is divided into five basic categories: (1) Mesh construction, (2) space mapping, (3) SPDE model construction, (4) plotting, and (5) INLA input and output structure bookkeeping.

Due mostly to the the complexity of building the binary executables that form the computational backbone of the R-**INLA** package, it is not available to install from Comprehensive R Archive Network (CRAN), but can still be easily installed and upgraded within R (R Core Team 2014) via the repositories `http://www.math.ntnu.no/inla/R/stable` and `http://www.math.ntnu.no/inla/R/testing`, or directly from its web page (Rue *et al.* 2013b). The non-testing version is updated less frequently than the testing version. The package website `http://www.r-inla.org/` contains more documentation, as well as a discussion forum. The recommended way to access the full source code is to clone the repository located at `http://code.google.com/p/inla/` (Rue, Martino, Lindgren, Simpson, and Riebler 2013a). On GNU/Linux systems, the `Makefile` supplied in the supplementary material can be used to download the code and build a binary R package.

The major challenge when designing a general software package for practical use of the SPDE/GMRF models is that of *bookkeeping*, i.e., how to assist the user in keeping track of the links between continuous and discrete representations, in a way that frees the user from having to know the details of the implementation and internal storage. To solve this, a bit of abstraction is needed to avoid cluttering the interface with those details. Thus, instead of visibly keeping track of mappings between triangle mesh node indices and data locations, the user can use sparse matrices to encode these relationships, and wrapper functions are provided to manipulate these matrices and associated index and covariate vectors in ways suitable for the intended usage.

2.1. Mesh construction

The basic tools for building basis function representations are provided by the low level function `inla.mesh.create()` and the three high level functions `inla.nonconvex.hull(loc, ...)`, `inla.mesh.2d()` and `inla.mesh.1d()`. The latter function defines B-spline basis representations in one dimension (see Section 3.1 for an example). The remainder of this section gives a brief introduction to mesh generation for two-dimensional domains.

The aim is to create the triangulated mesh on top of which the SPDE/GMRF representation is to be built. The example in Figure 1 illustrates a common usage case, which is to have semi-randomly scattered observation locations in a region of space such that there is no physical boundary, just a limited observation region. When dealing with only covariances between data points, this distinction is often unimportant, but here it becomes a possibly vital part of the model, since the SPDE will exhibit boundary effects. In the R-**INLA** implementation, Neumann boundaries are used, which increases the variance near the boundary. If we intend to model a stationary field across the entire domain of observations, we must therefore extend the model domain far enough so that the boundary effects don't influence the observations. However, note that the reverse is also true: if there *is* a physical boundary, the boundary effects may actually be desirable. The function `inla.mesh.2d()` allows us to create a mesh with small triangles in the domain of interest, and use larger triangles in the extension used to avoid boundary effects. This minimises the extra computational work needed due to the extension.

```
R> m <- 50
R> points <- matrix(runif(m * 2), m, 2)
R> mesh <- inla.mesh.2d(loc = points, cutoff = 0.05, offset = c(0.1, 0.4),
+    max.edge = c(0.05, 0.5))
```

The `cutoff` parameter is used to avoid building many small triangles around clustered input locations, `offset` specifies the size of the inner and outer extensions around the data locations, and `max.edge` specifies the maximum allowed triangle edge lengths in the inner domain and in the outer extension. The overall effect of the triangulation construction is that, if desired, one can have smaller triangles, and hence higher accuracy of the field representation, where the observation locations are dense, larger triangles where data is more sparse (and hence provides less detailed information), and large triangles where there is no data and spending computational resources would be wasteful. However, note that there is neither any guarantee nor any requirement that the observation locations are included as nodes in the mesh. If one so desires, the mesh can be designed from different principles, such as lattice points with no relation to the precise measurement locations. This emphasises the decoupling between the continuous domain of the field model and the discrete data locations.

A new feature is the option to compute a non-convex covering to use as boundary information,

```
R> bnd <- inla.nonconvex.hull(points, convex = 0.12)
```

with the resulting domain boundary shown in Figure 1b. The boundary is then supplied to `inla.mesh.2d()`,

```
R> mesh <- inla.mesh.2d(boundary = bnd, cutoff = 0.05, max.edge = c(0.1))
```

resulting in the non-convex mesh shown in Figure 1c. Just as before, the SPDE edge effects can be moved outside the domain of interest using an extension with larger triangles, shown in Figure 1d:

```
R> mesh <- inla.mesh.2d(boundary = bnd, cutoff = 0.05, offset = c(1, 0.5),
+    max.edge = c(0.1, 0.5) )
```

For geostatistical problems with global data, one can work directly on a spherical mesh. Any spatial coordinates must first be converted into 3D Cartesian coordinates. For longitudes and latitudes this can be done with `inla.mesh.map()`, and the result can be used with `inla.mesh.2d()`:

```
R> loc.cartesian = inla.mesh.map(loc.longlat, projection = "longlat")
R> mesh2 = inla.mesh.2d(loc = loc.cartesian, ...)
```

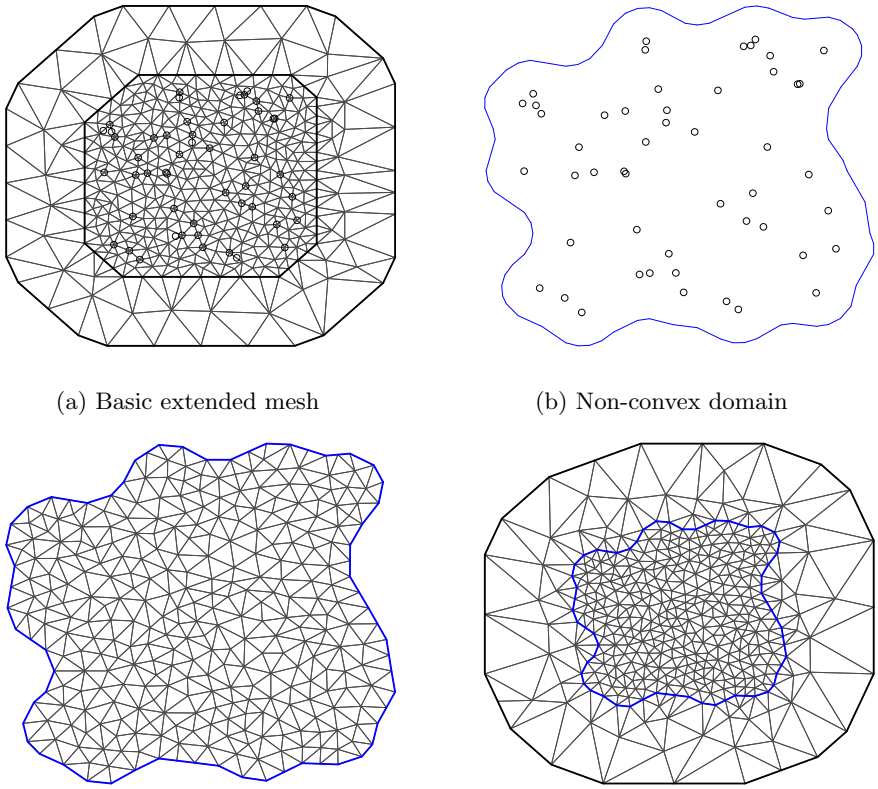
Alternatively, a semi-regular mesh can be constructed using the more low-level command

```
R> mesh2 = inla.mesh.create(globe = 10)
```

where the `globe` parameter specified the number of sub-segments to use, when subdividing an icosahedron. The points are adjusted to lie on constant latitude circles. See Figure 3 for an example of how to plot fields defined on spherical meshes.

2.2. Mapping between meshes and continuous space

One of the most important features is the `inla.spde.make.A()` functions, which computes the sparse weight matrices needed to map between the internal representation of weights for basis functions and the values of the resulting functions and fields. The basic syntax is



(a) Basic extended mesh

(b) Non-convex domain

(c) Non-convex mesh

(d) Extended non-convex mesh

Figure 1: Illustrations of meshes constructed based on a common set of domain definition points but different mesh generation parameters.

```
R> A <- inla.spde.make.A(mesh, loc = points)
```

which produces a matrix with $A_{ij} = \psi_j(\boldsymbol{s}_i)$ for all points $\boldsymbol{s}_i$ in `points`.

Models in R-**INLA** can have several replicates, as well as being *grouped*, which corresponds to Kronecker product models. In order to obtain the correct `A` matrix for such models, the user can specify indices for the parameters `group` and `repl`. A recently introduced feature also allows specifying a one dimensional `group.mesh`, which is then interpreted as defining a Kronecker product basis, such as for the space-time models mentioned in Section 1.1, and an example is given in Section 3.2.

2.3. SPDE model construction

As the theory and practice evolves, new internal SPDE representation models are implemented. The current development focuses on models with internal name `spde2`. To find out what the currently available user-accessible models are, use the function `inla.spde.models()`.

Defining an SPDE model object can now be as simple as

```
R> spde <- inla.spde2.matern(mesh, alpha = 2)
```

but in practice we need to also specify the prior distribution for the parameters, and/or modify the parameterisation to suit the specific situation. This is true in particular when the models are used as simple smoothers, as there is then rarely enough information in the likelihood to fully identify the parameters, giving more importance to the prior distributions.

Using the theory from Section 1.1, the empirically derived range expression $\rho = \sqrt{8\nu}/\kappa$ allows for construction of a model with known range and variance $(= 1)$ for $(\theta_1, \theta_2) = (0, 0)$, via

```
R> sigma0 <- 1
R> size <- min(c(diff(range(mesh$loc[, 1])), diff(range(mesh$loc[, 2]))))
R> range0 <- size/5
R> kappa0 <- sqrt(8)/range0
R> tau0 <- 1/(sqrt(4 * pi) * kappa0 * sigma0)
R> spde <- inla.spde2.matern(mesh, B.tau = cbind(log(tau0), -1, +1),
+    B.kappa = cbind(log(kappa0), 0, -1), theta.prior.mean = c(0, 0),
+    theta.prior.prec = c(0.1, 1))
```

Here, `sigma0` is the field standard deviation and `range0` is the spatial range for $\boldsymbol{\theta} = \mathbf{0}$, and `B.tau` and `B.kappa` are matrices storing the parameter basis functions introduced in Section 1.1. For stationary models, only the first matrix row needs to be supplied. In this example, the prior median for the spatial range is chosen heuristically to be a fifth of the approximate domain diameter.

Setting suitable priors for $\boldsymbol{\theta}$ in these models generally is difficult problem. The heuristic used above is to specify a fairly vague prior for θ_1 which controls the variance, with σ_0^2 being the median prior variance, and a larger prior precision for θ_2. When `range0` is a fifth of the domain size, the precision 1 for θ_2 gives an approximate 95% prior probability for the range being shorter than the domain size. Experimental helper functions for constructing parameterisations and priors are included in the package.

Models with range larger than the domain size are usually indistinguishable from intrinsic random fields, which can be modelled by fixing κ to zero (or rather some small positive value) with `B.tau = cbind(log(tau0), 1)` and `B.kappa = cbind(log(small), 0)`. Note that the sum-to-zero constraints often used for lattice based intrinsic Markov models is inappropriate due to the irregular mesh structure, and a *weighted* sum-to-zero constraint is needed to reproduce such models. The option `constr = TRUE` to the `inla.spde.matern()` call can be used to apply an integrate-to-zero constraint instead, which uses the triangle geometry to calculate the needed weights. Further integration constraints can be specified using `extraconstr.int = list(A = A, e = e)` option, which implements constraints of the form

$$\boldsymbol{A}\boldsymbol{C}\boldsymbol{x} = \boldsymbol{e},$$

where $\boldsymbol{C}$ is the sparse matrix with elements $C_{ij} = \int_\Omega \psi_i(\boldsymbol{s})\psi_j(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}$. Non-integration constraints can be supplied with `extraconstr`, and all constraints will be passed on automatically to the `inla()` call later.

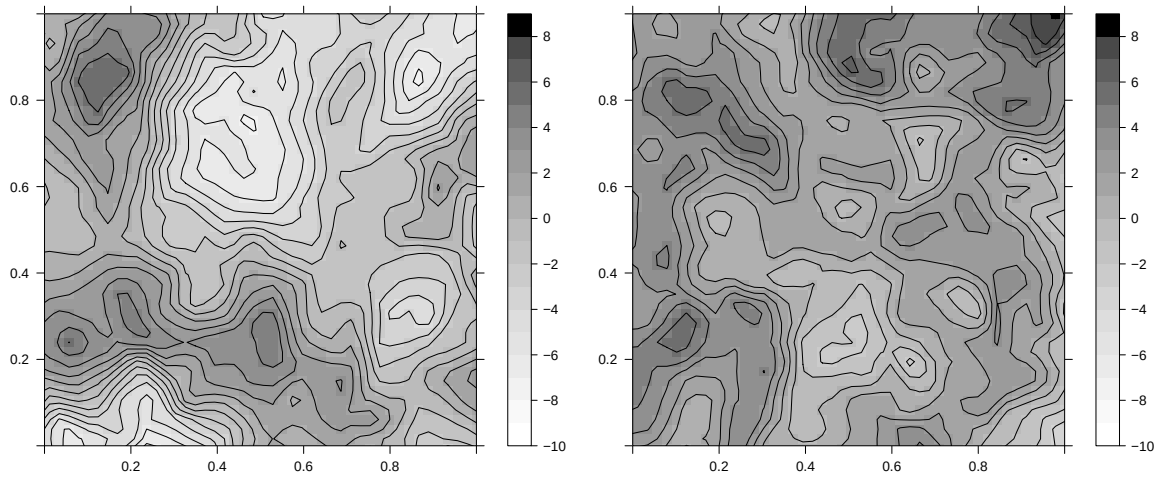


Figure 2: The two random field samples, with only the domain of interest, $[0,1]\times[0,1]$, shown.

Properties and sampling

There are several helper functions for querying properties about `spde` model objects, the most important one being `inla.spde.precision()`. To obtain the precision for the constructed model, with standard deviation a factor 3 larger than the prior median value, and range equal to the prior median, use

```
R> Q <- inla.spde.precision(spde, theta=c(log(3), 0))
```

The following code then generates two samples from the model,

```
R> x <- inla.qsample(n = 2, Q)
```

and the resulting fields are shown in Figure 2. To take any constraints specified in the `spde` object into account when sampling, use

```
R> x <- inla.qsample(n = 2, Q, constr = spde$f$extraconstr)
```

Obtaining covariances is a much more costly operation, but the function `inla.qinv(Q)` can quickly calculate all covariances between neighbours (in the Markov sense), including the marginal variances. Finally, `inla.qsolve(Q,b)` uses the internal R-**INLA** methods for solving a linear system involving $\boldsymbol{Q}$.

2.4. Plotting

Mesh structure

The interface supports a `plot()` function aimed at plotting the basic structure of a triangulation mesh. By specifying `rgl = TRUE`, the **rgl** (Adler, Murdoch, and others 2014) plotting system is used, which is useful in particular for spherical domains. Variations of the following commands were used to produce Figure 1:

```
R> plot(mesh)
R> plot(mesh, rgl = TRUE)
R> lines(mesh$segm$bnd, mesh$loc, add = FALSE)
```

Spatial fields

For plotting fields defined on meshes, one option is to use the `rgl` option, which supports specifying colour values for the nodes, producing an interpolated field plot, and optionally draw triangle edges and vertices in the same plot:

```
R> plot(mesh, rgl = TRUE, col = x[, 1],
+    color.palette = function(n) grey.colors(n, 1, 0),
+    draw.edges = FALSE, draw.segments = TRUE, draw.vertices = FALSE)
```

The more common option is to explicitly evaluate the field on a regular lattice, and use any matrix-based plotting mechanism, such as `image()`:

```
R> proj <- inla.mesh.projector(mesh, dims = c(100, 100))
R> image(proj$x, proj$y, inla.mesh.project(proj, field = x[, 1]))
```

All the figures showing fields have been drawn using a wrapper around the `levelplot()` from **lattice** (Sarkar 2008), which is available in the supplementary material.

The `inla.mesh.project/or()` functions are here used to map between the basis function weights for the mesh nodes and points on a regular grid, by default a 100×100 lattice covering the mesh domain. The functions also support several types of projections for spherical domains,

```
R> mesh2 <- inla.mesh.create(globe = 10)
R> proj2a <- inla.mesh.projector(mesh2, projection = "longlat",
+    dims = c(361, 181))
R> proj2b <- inla.mesh.projector(mesh2, projection = "mollweide",
+    dims = c(361, 181))
```

with projected fields shown in Figure 3.

2.5. Advanced predictor manipulation

To aid the user in setting up an appropriate $\boldsymbol{A}$ for a whole model, the function `inla.stack()` can be used. The function is meant to hide most of the tedious vector manipulation that is necessary, and from the user's point of view can be seen as implementing a few abstract operations on effect and predictor definitions. Section 2.6 contains a practical example showing the use of this feature, while here we describe the abstract operations.

First, all effects are conceptually joined into a compact matrix-like notation:

$$\begin{aligned} \boldsymbol{Z} &= \left[\boldsymbol{z}^{(1)} \quad \dots \quad \boldsymbol{z}^{(K)}\right], \\ \boldsymbol{\eta} &= H(\boldsymbol{Z}) = \sum_k h_k(\boldsymbol{z}^{(k)}), \\ \boldsymbol{\eta}^* &= \boldsymbol{A}\, H(\boldsymbol{Z}), \end{aligned}$$

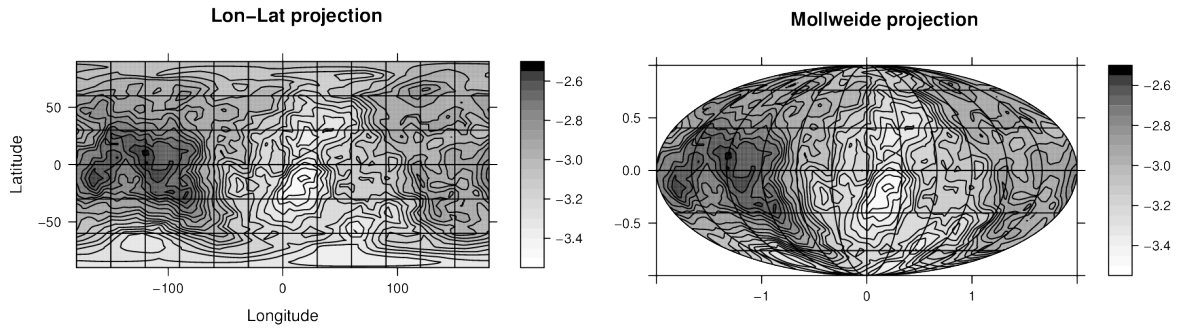


Figure 3: Projections of a sample from an SPDE model on a spherical domain. The left panel uses longitude-latitude projection, and the right hand panel uses the equal area Mollweide projection.

where $h_k(\cdot)$ are the effect mapping functions defined in Section 1.2 and specified using a model `formula`. The effects are treated as *named* vectors, regardless of the ordering. Any effect known to H but not present in a particular $\boldsymbol{Z}$ is treated as "no effect", which is the same effect as when providing `NA` values.

The first operation is to construct *sums* of predictors (shown here only for two predictors):

$$\begin{aligned}
\boldsymbol{\eta}^* &= \boldsymbol{A}_1\, H(\boldsymbol{Z}_1) + \boldsymbol{A}_2\, H(\boldsymbol{Z}_2) = \widetilde{\boldsymbol{A}}\, H(\widetilde{\boldsymbol{Z}}),\\
\widetilde{\boldsymbol{Z}} &= \begin{bmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{bmatrix},\\
\widetilde{\boldsymbol{A}} &= \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \end{bmatrix}.
\end{aligned}$$

The joining of the $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ effect collections is performed by matching vector names, adding `NA` for any missing components.

The second operation is to *join* predictors in sequence (only two predictors shown):

$$\begin{aligned}
\boldsymbol{\eta}_1^* &= \boldsymbol{A}_1\, H(\boldsymbol{Z}_1),\\
\boldsymbol{\eta}_2^* &= \boldsymbol{A}_2\, H(\boldsymbol{Z}_2),\\
\widetilde{\boldsymbol{\eta}}^* &= (\boldsymbol{\eta}_1^*, \boldsymbol{\eta}_2^*) = \widetilde{\boldsymbol{A}}\, H(\widetilde{\boldsymbol{Z}}),\\
\widetilde{\boldsymbol{Z}} &= \begin{bmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{bmatrix},\\
\widetilde{\boldsymbol{A}} &= \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 \end{bmatrix}.
\end{aligned}$$

As a post-processing step for both operations, the new covariate matrix $\widetilde{\boldsymbol{Z}}$ and matrix $\widetilde{\boldsymbol{A}}$ are analysed to detect any duplicate rows in $\widetilde{\boldsymbol{Z}}$ or any all-zero columns in $\widetilde{\boldsymbol{A}}$, and those are by default removed in order to minimise the internal size of the model representation.

The syntax for a *sum* operation is

```
R> stack <- inla.stack(data = list(...), A = list(A1, A2, ...),
+    effects = list(list(...), list(...), ...), tag = ...)
```

where each `A` matrix has an associated list of effects. A *join* operation is performed by supplying two or more previously generated stack objects,

```
R> stack <- inla.stack(stack1, stack2, ...)
```

Any vectors specified in the `data` list, most importantly the response variable vector itself, should be the same length as the predictor itself (scalars are replicated to the appropriate length). With the help of the name-`tag` it also keeps track of the indices needed to map from the original inputs into the resulting *stacked* representation. See Section 2.6 for an illustrating example of using `inla.stack()`.

Note that $H(\cdot)$ is conceptually defined by the model `formula`, which needs to mention every covariate component present in $\boldsymbol{Z}$ and that is meant to be used.

2.6. Bayesian inference

In this section we will look at a simple example of how to use the SPDE models in latent Gaussian models when doing direct Bayesian inference based on the INLA method described in Section 1.2.

Consider a simple Gaussian linear model involving two independent realisations (replicates) of a latent spatial field $x(\boldsymbol{s})$, observed at the same m locations, $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\}$, for each replicate. For each $i = 1, \ldots, m$,

$$\begin{aligned} y_i &= \beta_0 + c_i\beta_c + x_1(\boldsymbol{s}_i) + e_i, \\ y_{i+m} &= \beta_0 + c_{i+m}\beta_c + x_2(\boldsymbol{s}_i) + e_{i+m}, \end{aligned}$$

where c_i is an observation-specific covariate, e_i is measurement noise, and $x_1(\cdot)$ and $x_2(\cdot)$ are the two field replicates. Note that the intercept, β_0, can be interpreted as a spatial covariate effect, constant over the domain.

We use the basis function representation of $x(\cdot)$ to define a sparse matrix of weights $\boldsymbol{A}$ such that $x(\boldsymbol{s}_i) = \sum_j A_{ij}x_j$, where $\{x_j\}$ is the joint set of weights for the two replicate fields. If we only had one replicate, we would have $A_{ij} = \psi_j(\boldsymbol{s}_i)$. The matrix can be generated by `inla.spde.make.A()`, which locates the points in the mesh and organises the evaluated values of the basis functions for the two replicates:

```
R> A <- inla.spde.make.A(mesh, loc = points, index = rep(1:m, times = 2),
+    repl = rep(1:2, each = m))
```

For each observation, `index` gives the corresponding index into the matrix of measurement locations, and `repl` determines the corresponding replicate index. In case of missing observations, one can either keep this `A` matrix while setting the corresponding elements of the data vector `y` to `NA`, or omit the corresponding elements from `y` as well as from the `index` and `repl` parameters above. Also note that the row-sums of $\boldsymbol{A}$ are 1, since the piece-wise linear basis functions we use sum to 1 at each location.

Rewriting the observation model in vector form gives

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{1}\beta_0 + \boldsymbol{c}\beta_c + \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e} \\ &= \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{1}\beta_0) + \boldsymbol{c}\beta_c + \boldsymbol{e} \end{aligned}$$

Using the helper functions, we can generate data using our two previously simulated model replicates,

```
R> x <- as.vector(x)
R> covariate <- rnorm(m * 2)
R> y <- 5 + covariate * 2 + as.vector(A %*% x) + rnorm(m * 2) * 0.1
```

The `formula` in `inla()` defines a linear predictor $\boldsymbol{\eta}$ as the sum of all effects, and an `NA` in a covariate or index vector is interpreted as *no effect*. To accommodate predictors that involve more than one element of a random effect, one can specify a sparse matrix of weights defining an arbitrary linear combination of the elements of $\boldsymbol{\eta}$, giving a new predictor vector $\boldsymbol{\eta}^*$. The linear predictor output from `inla()` then contains the joint vector $(\boldsymbol{\eta}^*, \boldsymbol{\eta})$. To implement our model, we separate the spatial effects from the covariate by defining

$$\boldsymbol{\eta}_e = \begin{bmatrix} \boldsymbol{x} + \mathbf{1}\beta_0 \\ \boldsymbol{c}\beta_c \end{bmatrix},$$

and construct the predictor as

$$\begin{aligned} \boldsymbol{\eta}_e^* &= \boldsymbol{A}(\boldsymbol{x} + \mathbf{1}\beta_0) + \boldsymbol{c}\beta_c \\ &= \begin{bmatrix} \boldsymbol{A} & \boldsymbol{I} \end{bmatrix} \boldsymbol{\eta}_e = \boldsymbol{A}_e \boldsymbol{\eta}_e \end{aligned}$$

so that now $\mathsf{E}(\boldsymbol{y} \mid \boldsymbol{\eta}_e) = \boldsymbol{\eta}_e^*$. The bookkeeping required to describe this to `inla()` involves concatenating matrices and adding `NA` elements to the covariates and index vectors:

$$\begin{aligned} \boldsymbol{A}_e &= \begin{bmatrix} \boldsymbol{A} & \boldsymbol{I}_{2m} \end{bmatrix} \\ \texttt{field0} &= (1, \ldots, n, 1, \ldots, n) \\ \texttt{field} &= (\texttt{field0}, \texttt{NA}, \ldots, \texttt{NA}) \\ \texttt{intercept} &= (1, \ldots, 1, \texttt{NA}, \ldots, \texttt{NA}) \\ \texttt{cov} &= (\texttt{NA}, \ldots, \texttt{NA}, c_1, \ldots, c_{2m}) \end{aligned}$$

Doing this by hand with `Matrix::cBind()`, `c()`, and `rep()` quickly becomes tedious and error-prone, so one can instead use the helper function `inla.stack()`, which takes blocks of data, weight matrices, and effects and joins them, adding `NA` where needed. Identity matrices and constant covariates can be abbreviated to scalars, with a complaint being issued if the input is inconsistent or ambiguous.

We also need to keep track of the two field replicates, and use `inla.spde.make.index()`, which gives a list of index vectors for indexing the full mesh and its replicates (it can also be used for indexing Kronecker product `group` models, e.g., in simple multivariate and spatio-temporal models). The code

```
R> mesh.index <- inla.spde.make.index(name = "field", n.spde = spde$n.spde,
+    n.repl = 2)
```

generates a list `mesh.index` with three index vectors,

$$\begin{aligned} \texttt{field} &= (1, \ldots, n, 1, \ldots, n), \\ \texttt{field.repl} &= (1, \ldots, 1, 2, \ldots, 2), \\ \texttt{field.group} &= (1, \ldots, 1, 1, \ldots, 1). \end{aligned}$$

The predictor information for the observed data can now be collected, using

```
R> st.est <- inla.stack(data = list(y = y), A = list(A, 1),
+    effects = list(c(mesh.index, list(intercept = 1)),
+      list(cov = covariate)), tag = "est")
```

where the `tag` identifier can later be used for identifying the correct indexing into the `inla()` output. As discussed in Section 2.5, each "`A`" matrix must have an associated list of "`effects`", in this case `A`:(`field`, `field.repl`, `field.group`, `intercept`) and `1`:(`cov`). The `data` list may contain anything associated with the "left hand side" of the model, such as exposure `E` for Poisson likelihoods. By default, duplicates in the effects are identified and replaced by single copies (`compress = TRUE`), and effects that do not affect $\boldsymbol{\eta}^*$ are removed completely (`remove.unused = TRUE`), so that each column of the resulting `A` matrix has a least one non-zero element.

If we want to obtain the posterior prediction of the combined spatial effects at the mesh nodes, $x(\boldsymbol{s}_i) + \beta_0$, we can define

$$\begin{aligned}\boldsymbol{\eta}_p &= \boldsymbol{x} + \mathbf{1}\beta_0\\ \boldsymbol{\eta}_p^* &= \boldsymbol{I}\boldsymbol{\eta}_p = \boldsymbol{A}_p\boldsymbol{\eta}_p\end{aligned}$$

and construct the corresponding information stack with

```
R> st.pred <- inla.stack(data = list(y = NA), A = list(1),
+    effects = list(c(mesh.index, list(intercept = 1))), tag = "pred")
```

We can now join the estimation and prediction stack into a single stack,

```
R> stack <- inla.stack(st.est, st.pred)
```

and the result is simplified by removing duplicated effects:

$$\boldsymbol{\eta}_s^* = \begin{bmatrix}\boldsymbol{A}_e & \mathbf{0}\\ \mathbf{0} & \boldsymbol{A}_p\end{bmatrix}\begin{bmatrix}\boldsymbol{\eta}_e\\ \boldsymbol{\eta}_p\end{bmatrix} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}+\mathbf{1}\beta_0\\ \boldsymbol{c}\beta_c\\ \boldsymbol{x}+\mathbf{1}\beta_0\end{bmatrix} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{I}\\ \boldsymbol{I} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}+\mathbf{1}\beta_0\\ \boldsymbol{c}\beta_c\end{bmatrix} = \boldsymbol{A}_s\boldsymbol{\eta}_s$$

In this simple example, the second block row of $\boldsymbol{A}_s$ (generating $\boldsymbol{x}+\mathbf{1}\beta_0$) is not strictly needed, since the same information would be available in $\boldsymbol{\eta}_s$ itself if we specified `remove.unused = FALSE` when constructing `stack.pred` and `stack`, but in general such special cases can be hard to keep track of.

We are now ready to do the actual estimation. Note that we must explicitly remove the default intercept from the $\boldsymbol{\eta}$-model, since that would otherwise be applied twice in the construction of $\boldsymbol{\eta}^*$, and the constant covariate `intercept` is used instead:

```
R> formula <- y ~ -1 + intercept + cov +
+    f(field, model = spde, replicate = field.repl)
R> inla.result <- inla(formula, data = inla.stack.data(stack, spde = spde),
+    family = "normal", control.predictor = list(A = inla.stack.A(stack),
+      compute = TRUE))
```

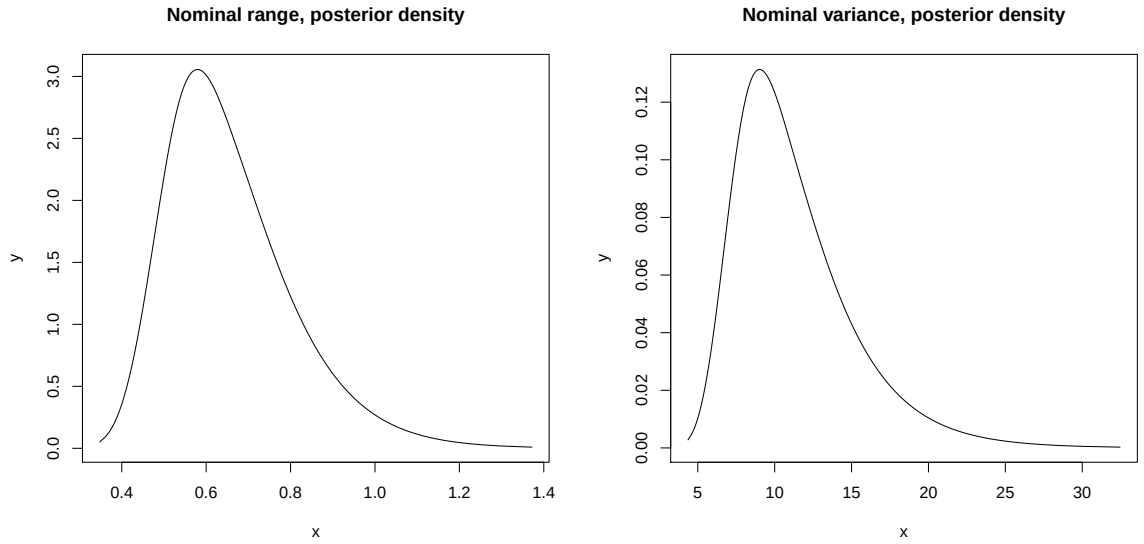


Figure 4: Posterior densities for nominal range and variance. The true values were 0.44 and 9.

The function `inla.stack.data()` produces the list of variables needed to evaluate the formula and `inla.stack.A()` extracts the $\boldsymbol{A}_s$ matrix.

Since the SPDE-related contents of `inla.result` can be hard to interpret, the helper function `inla.spde2.result()` can be used to extract the relevant parts and transform them into more user-friendly information, such as posterior densities for range and variance instead of raw distributions for θ, as shown in Figure 4:

```
R> result <- inla.spde2.result(inla.result, "field", spde)
R> plot(result[["marginals.range.nominal"]][[1]], type = "l",
+    main = "Nominal range, posterior density")
```

The posterior means and standard deviations for the latent fields can be extracted and plotted as follows, where `inla.stack.index()` provides the necessary mappings between the `inla()` output and the original data stack specifications:

```
R> index <- inla.stack.index(stack, "pred")$data
R> linpred.mean <- inla.result[["summary.linear.predictor"]]$mean
R> linpred.sd <- inla.result[["summary.linear.predictor"]]$sd
R> image(proj$x, proj$y, inla.mesh.project(proj,
+    linpred.mean[index[mesh.index$field.repl == 1]]))
R> image(proj$x, proj$y, inla.mesh.project(proj,
+    linpred.sd[index[mesh.index$field.repl == 1]]))
```

Figure 5 shows the posterior mean and standard deviation fields as drawn with `levelplot()` instead of `image()`.

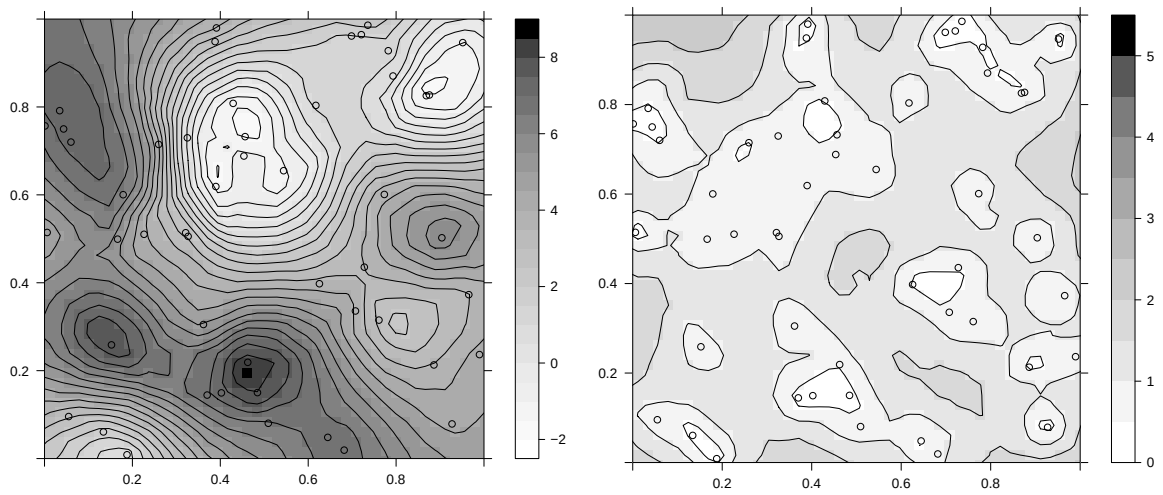


Figure 5: Posterior mean (left) and standard deviation (right) for $\beta_0 + x_1(\boldsymbol{s})$.

3. Further examples

The package website (Rue *et al.* 2013b) has several tutorials and case-studies, including an extensive collection of examples for the SPDE models. Here we give only three brief examples showing fundamental concepts.

3.1. Non-Gaussian data: Tokyo rainfall

The Tokyo rainfall data set contains daily rainfall indicator counts (`y`) for a period of two years, and can be modelled using a seasonally varying Binomial model, $y_t \sim \mathsf{Bin}(n_t, p_t)$, $t = 1, \ldots, 366$. Here, $n_t = 2$ for all days except one (where $n_t = 1$) and p_t is a slowly varying periodic function giving the rainfall probability for each day of the year. We will model p_t as a logit-transformed Gaussian process.

This is a simple example for using a 1D SPDE model with a 2nd order B-spline basis representation. The desired model is an intrinsic 2nd order random walk, but since there is no `inla.spde2.intrinsic()` wrapper function yet, the details are set up explicitly.

First, a 2nd order B-spline basis mesh with 24 basis functions and cyclic boundary conditions is defined, and an appropriate `spde` model is constructed. The mesh is specified to have cyclic boundary conditions,

```
R> data("Tokyo")
R> knots <- seq(1, 367, length = 25)
R> mesh <- inla.mesh.1d(knots, interval = c(1, 367), degree = 2,
+    boundary = "cyclic")
```

and the prior median for τ is calculated to give a specified prior median standard deviation when κ is fixed to a small value,

```
R> sigma0 <- 1
R> kappa0 <- 1e-3
R> tau0 <- 1/(4 * kappa0^3 * sigma0^2)^0.5
```

and finally an `inla.spde2` object is created:

```
R> spde <- inla.spde2.matern(mesh, constr = FALSE,
+    B.tau = cbind(log(tau0), 1), B.kappa = cbind(log(kappa0), 0),
+    theta.prior.prec = 1e-4)
```

Next, the data is organised using `inla.stack()`:

```
R> A <- inla.spde.make.A(mesh, loc = Tokyo$time)
R> time.index <- inla.spde.make.index("time", n.spde = spde$n.spde)
R> stack <- inla.stack(data = list(y = Tokyo$y, link = 1, Ntrials = Tokyo$n),
+    A = list(A), effects = list(time.index), tag = "est")
R> formula <- y ~ -1 + f(time, model = spde)
R> data <- inla.stack.data(stack)
R> result <- inla(formula, family = "binomial", data = data,
+    Ntrials = data$Ntrials, control.predictor = list(
+    A = inla.stack.A(stack), link = data$link, compute = TRUE))
```

Finally, the results can be plotted, as shown in Figure 6:

```
R> time <- 1:366
R> index <- inla.stack.index(stack, "est")$data
R> plot(Tokyo$time, Tokyo$y / Tokyo$n, xlab = "Day", ylab = "Probability")
R> lines(time, result$summary.fitted.values$mean[index])
R> lines(time, result$summary.fitted.values$"0.025quant"[index], lty = 2)
R> lines(time, result$summary.fitted.values$"0.975quant"[index], lty = 2)
```

3.2. Kronecker product models for space-time

The following illustrates the principles for how to set up a Kronecker product space-time model where both space and time are treated continuously, with a Matérn model in space and a stationary AR(1) model on coefficients for 2nd order B-splines in time. Assumed inputs are

- `y`: Measurements.
- `station.loc`: Coordinates for measurement stations.
- `station.id`: For each measurement, which station index?
- `time`: For each measurement, what time?

The time interval for the model is $[0, 100]$.

```
R> knots <- seq(0, 100, length = 11)
R> mesh1 <- inla.mesh.1d(loc = knots, degree = 2, boundary = "free")
R> mesh2 <- inla.mesh.2d(...)
R> spde <- inla.spde2.matern(mesh2, alpha = 2, ...)
R> index <- inla.spde.make.index("space", n.spde = spde$n.spde,
```

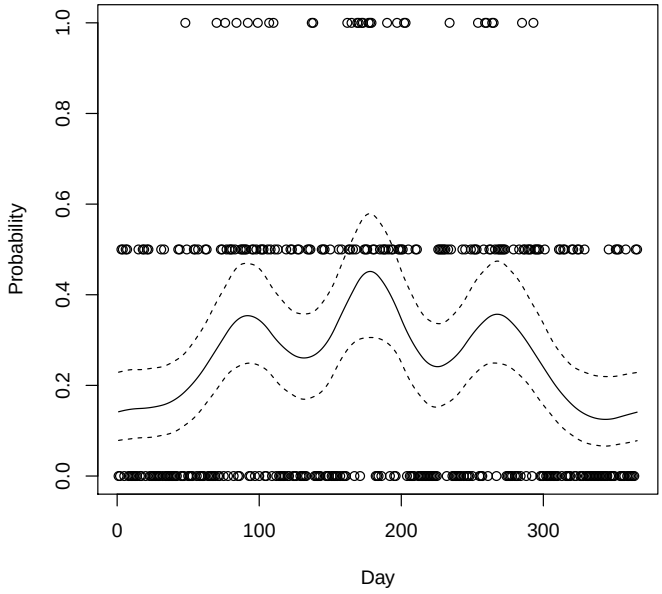


Figure 6: Empirical and model-based Binomial probability estimates for the Tokyo rainfall data set, with 95% posterior predictive bounds. The empirical probability estimates are the proportion of observed rainfall days for each day of the year.

```
+    n.group = mesh1$m)
R> formula <- y ~ -1 + f(space, model = spde, group = space.group,
+    control.group = list(model = "ar1"))
R> A <- inla.spde.make.A(mesh2, loc = station.loc, index = station.id,
+    group = time, group.mesh = mesh1)
R> stack <- inla.stack(data = list(y = y), A = list(A),
+    effects = list(index))
```

3.3. Multiple observational likelihood models

For problems with data from more than one observational likelihood model, `inla.stack()` can help in structuring the inputs, greatly reducing the amount of hand-tooled `rep()` statements needed to setup the input data structures. The following example shows how to setup a model where `y1` are Binomial and `y2` are Poisson, but the formula connects them via the linear predictor effects:

```
R> stack1 <- inla.stack(data = list(Y = cbind(y1, NA), link = 1, N = N1),
+    ...)
R> stack2 <- inla.stack(data = list(Y = cbind(NA, y2), link = 2, E = E2),
+    ...)
R> stack <- inla.stack(stack1, stack2)
R> data <- inla.stack.data(stack)
R> result <- inla(Y ~ ..., data = data, family = c("binomial", "poisson"),
+    Ntrials = data$N, E = data$E, control.predictor = list(
+      A = inla.stack.A(stack), link = data$link, compute = TRUE))
```

4. Future development

The R-**INLA** package is in constant development, with new models added as they are needed and developed. The current work for the SPDE models is focusing on construction of parameter basis functions and priors for non-stationary model parameters, as well as implementing extensions to non-separable space-time models and more flexible boundary conditions. An associated package **excursions** for computing level excursion sets with joint excursion probabilities, as well as credible regions for contour curves, is available from CRAN (Bolin and Lindgren 2015).

As the size of spatial and spatio-temporal models and data sets grows, iterative matrix methods and other approximation techniques for more complex models are also being investigated, with the long-term goal of replacing the core of R-**INLA** to more easily handle such challenges.

Acknowledgments

The authors wish to thank their collaborators David Bolin, Michela Cameletti, Janine Illian, Johan Lindström, Thiago Martins, Daniel Simpson, Sigrunn Sørbye, Elias Krainski, and Ryan Yue, who have all contributed with ideas and suggestions for the development of the spatial model interface. We are also grateful to the editors and reviewers for their thoughtful comments on the manuscript.

References

Adler D, Murdoch D, others (2014). ***rgl**: 3D Visualization Device System (OpenGL).* R package version 0.93.1098, URL http://CRAN.R-project.org/package=rgl.

Banerjee S, Gelfand AE, Finley AO, Sang H (2008). "Gaussian Predictive Process Models for Large Spatial Datasets." *Journal of the Royal Statistical Society B*, **70**(4), 825–848.

Besag J (1981). "On a System of Two-Dimensional Recurrence Equations." *Journal of the Royal Statistical Society B*, **43**(3), 302–309.

Besag J, Mondal D (2005). "First-Order Intrinsic Autoregressions and the De Wijs Process." *Biometrika*, **92**(4), 909–920.

Bolin D, Lindgren F (2011). "Spatial Models Generated by Nested Stochastic Partial Differential Equations, with an Application to Global Ozone Mapping." *The Annals of Applied Statistics*, **5**(1), 523–550.

Bolin D, Lindgren F (2013). "A Comparison between Markov Approximations and Other Methods for Large Spatial Data Sets." *Computational Statistics & Data Analysis*, **61**, 7–21.

Bolin D, Lindgren F (2015). "Excursion and Contour Uncertainty Regions for Latent Gaussian Models." *Journal of the Royal Statistical Society B*, **77**(1), 85–106.

Cameletti M, Lindgren F, Simpson D, Rue H (2013). "Spatio-Temporal Modeling of Particulate Matter Concentration through the SPDE Approach." *AStA Advances in Statistical Analysis*, **97**(2), 109–131.

Cressie NAC, Johannesson G (2008). "Fixed Rank Kriging for Very Large Spatial Data Sets." *Journal of the Royal Statistical Society B*, **70**(1), 209–226.

Lindgren F, Rue H, Lindström J (2011). "An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach." *Journal of the Royal Statistical Society B*, **73**(4), 423–498.

Martins TG, Simpson D, Lindgren F, Rue H (2013). "Bayesian Computing with **INLA**: New Features." *Computational Statistics & Data Analysis*, **67**, 68–83.

Nychka D, Hammerling D, Sain S, Lenssen N (2013). ***LatticeKrig**: Multiresolution Kriging Based on Markov Random Fields*. URL http://CRAN.R-project.org/package=LatticeKrig.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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

Rue H, Martino S, Chopin N (2009). "Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations." *Journal of the Royal Statistical Society B*, **71**, 319–392.

Rue H, Martino S, Lindgren F, Simpson D, Riebler A (2013a). *Code Repository for R-**INLA**: Approximate Bayesian Inference Using Integrated Nested Laplace Approximations*. URL http://code.google.com/p/inla/.

Rue H, Martino S, Lindgren F, Simpson D, Riebler A (2013b). *R-**INLA**: Approximate Bayesian Inference Using Integrated Nested Laplace Approximations*. Trondheim, Norway. URL http://www.r-inla.org/.

Sarkar D (2008). ***lattice**: Multivariate Data Visualization with R*. Springer-Verlag, New York.

Sigrist F, Künsch HR, Stahel WA (2015). "Stochastic Partial Differential Equation Based Modelling of Large Space-Time Data Sets." *Journal of the Royal Statistical Society B*, **77**(1), 3–33.

Simpson D, Lindgren F, Rue H (2012a). "In Order to Make Spatial Statistics Computationally Feasible, We Need to Forget about the Covariance Function." *Environmetrics*, **23**(1), 65–74.

Simpson D, Lindgren F, Rue H (2012b). "Think Continuous: Markovian Gaussian Models in Spatial Statistics." *Spatial Statistics*, **1**, 16–29.

Whittle P (1954). "On Stationary Processes in the Plane." *Biometrika*, **41**(3/4), 434–449.

Whittle P (1963). "Stochastic Processes in Several Dimensions." *Bulletin of the International Statistical Institute*, **40**, 974–994.

Wikle CK (2010). "Low-Rank Representations for Spatial Processes." In AE Gelfand, PJ Diggle, M Fuentes, P Guttorp (eds.), *Handbook of Spatial Statistics*, pp. 107–118. Chapman & Hall/CRC, Boca Raton.

Affiliation:

Finn Lindgren
Department of Mathematical Sciences
University of Bath
Claverton Down
BA2 7AY, Bath, United Kingdom
E-mail: f.lindgren@bath.ac.uk
URL: http://people.bath.ac.uk/fl353/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 63, Issue 19 *Submitted:* 2013-06-05
January 2015 *Accepted:* 2014-09-30