Spatial Statistics 8 (2014) 39–51

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Scaling intrinsic Gaussian Markov random field priors in spatial modelling

Sigrunn Holbek Sørbye [a,*], Håvard Rue [b]

[a] Department of Mathematics and Statistics, Faculty of Science, University of Tromsø, N-9037 Tromsø, Norway

[b] Department of Mathematical Sciences, Norwegian University of Science and Technology, Norway

ARTICLE INFO

Article history:
Received 25 February 2013
Accepted 25 June 2013
Available online 18 July 2013

Keywords:
Hyperprior
ICAR-models
R-INLA
Latent Gaussian model
Precision
Scaling

ABSTRACT

In Bayesian hierarchical regression models, intrinsic Gaussian Markov random fields (IGMRFs) are commonly applied to model underlying spatial or temporal dependency structures. IGMRFs have a scaled precision matrix that reflects the neighbourhood structure of the model, while the scaling is represented as a random precision parameter. The hyperprior chosen for the precision parameter influences the degree of smoothness of the resulting field and this can have a strong effect on posterior results. We suggest that hyperpriors for the precision parameters should be selected according to the type of IGMRF used. Also, hyperpriors for different types of IGMRFs can be selected to give the same degree of smoothness, a priori. This is achieved by mapping the random precision to the marginal standard deviation of the IGMRF and recalculate hyperpriors used for different models. Also, the parameters of the hyperprior can be interpreted in terms of the marginal standard deviation. The given ideas are demonstrated by analysing two different types of spatial data in R-INLA, including a district-level analysis of survival data and the analysis of a spatial point pattern discretized to a grid.

1. Introduction

Intrinsic conditional auto-regressive (ICAR) models (Besag et al., 1991) are widely used as priors to model underlying dependency structures in Bayesian hierarchical models. Following Rue and Held

* Corresponding author. Tel.: +47 77645504; fax: +47 77644765.
E-mail address: sigrunn.sorbye@uit.no (S.H. Sørbye).

2211-6753/$ – see front matter
http://dx.doi.org/10.1016/j.spasta.2013.06.004

(2005), we will refer to these models as intrinsic Gaussian Markov random fields (IGMRFs). These models contain a precision or scaling parameter and the hyperprior chosen for this parameter influences the smoothness of the resulting estimated field. This implies that posterior estimates and inference can be rather sensitive to the choice of hyperpriors. For example, this is noticed in spatial point pattern analysis where the hyperprior for the precision of a spatially structured effect has to be selected carefully to avoid overfitting (Illian et al., 2012). Despite such sensitivity, choices of hyperpriors in hierarchical models have often been made in a rather ad hoc manner (Berger et al., 2005).

The issue of prior selection in Bayesian analysis has been addressed by a number of authors, see for example Gelman et al. (2004) and Lesaffre and Lawson (2012) for thorough presentations of different classes of priors. Some advises on prior specifications are given in Fong et al. (2010), analysing generalized linear mixed models. Roos and Held (2011) quantify sensitivity to prior assumptions for the same class of models, using sensitivity measures based on the Hellinger distance. Here, we focus solely on hyperprior selection for the precision parameters of IGMRFs. The main motivation for this is that these models are widely applied as priors, among others to model spatial and temporal dependency structures in datasets. Software packages provide default hyperpriors for the precision of these models, but usually these have to be tuned by the user. Using the WinBUGS-software (Lunn et al., 2000), precision parameters are assigned a Gamma(ϵ, ϵ) distribution as a default (where ϵ is a small value). However, this has been seen to result in inappropriate priors for the precision of random effects, especially for continuous data (Lunn et al., 2009).

A primary aim of this paper is to suggest easily implemented guidelines to choose a hyperprior for the precision parameter, according to the specific IGMRF-model used. This is achieved by mapping the precision parameter to the marginal standard deviation of the model, calculated under linear constraints. The IGMRF-models considered here can be seen to penalize local deviation from a constant level, a line or a plane. The marginal standard deviations of the models, integrating out the random precision, give information on how large we allow this local deviation to be. A priori, it seems reasonable that different models should be scaled to have a similar range for this local deviation. This is achieved here by defining an upper limit for the marginal standard deviations which is then linked to the hyperparameters. This also provides an alternative interpretation of the hyperparameters which can be used to tune the parameters. As a result, we obtain hyperpriors that are invariant both to the model used, and also to the resolution of the analysis.

We have chosen to implement the given ideas using the methodology of integrated nested Laplace approximations (INLA), introduced in Rue et al. (2009). INLA provides efficient Bayesian inference for a specific class of hierarchical models named latent Gaussian models, described in Section 2. The INLA-methodology is easily available using the R-interface `R-INLA`, which will be used for the implementation. The specific prior models that will be considered include the first and second-order IGMRFs on a line, used to model smooth non-linear functions of covariates. Spatially structured effects are modelled using either a first-order IGMRF defined on an irregular lattice or a two-dimensional second-order IGMRF defined on a regular lattice. All of these models are described in Sections 3.3 and 3.4 in Rue and Held (2005) and at www.r-inla.org. Currently, the precision parameters of these different models are, by default, assigned the same vague Gamma$(1, 5 \cdot 10^{-5})$ distribution in `R-INLA`. Naturally, these defaults can and should not be used blindly and other hyperpriors can easily be specified by the user.

The plan of this paper is as follows. Section 2 describes the structure of the different IGMRFs considered and the use of these as priors in latent Gaussian models. The main ideas of the paper are presented in Section 3, giving details on the mapping between the precision parameter and the marginal standard deviation. Section 4 demonstrates the given suggestions in the analysis of two real data examples. Discussion and concluding remarks are given in Section 5.

2. IGRMFs as priors in latent Gaussian models

The class of Bayesian hierarchical regression models that we consider here are referred to as latent Gaussian models (Rue et al., 2009), being a subclass of structured additive regression models (Fahrmeir and Tutz, 2001; Gelman et al., 2004). We assume that the observational vector $\boldsymbol{y}$ belongs to

an exponential family, in which the mean $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{y})$ is linked to a structured additive predictor

$$\eta_i = g(\mu_i) = \alpha + \sum_j \beta_j z_{ji} + \sum_k f_k(c_{ki}) + u_i, \quad i = 1, \ldots, n,$$

where all random components are assigned Gaussian priors. More specifically, α denotes an intercept or offset, while $\boldsymbol{z}$ denotes covariates assumed to have a fixed linear effect on the response. The corresponding vector of unknown parameters $\boldsymbol{\beta}$ is assumed to have independent zero-mean Gaussian priors with fixed precisions, while the unstructured random effects in $\boldsymbol{u}$ are assigned i.i.d. zero-mean Gaussian priors with random precision. The set of functions $\{f_k(\cdot)\}$ is included to account for non-linear random effects of continuous covariates, having value c_{ki} for observation i, and these functions can be used to explain temporal, spatial or other underlying dependency structures in the data. We consider the case where one or more of these functions are modelled using IGMRFs. In the following, we let $\boldsymbol{x}$ denote the IGMRF-model assumed for a specific function $f_k(\cdot)$ and the main focus is how to tune the hyperprior for the random precision of this model.

A thorough discussion on intrinsic CAR-models is given in Besag and Kooperberg (1995), comparing these models with standard CAR-models. They conclude that the intrinsic version has certain advantages over the standard model, both conceptually and in practice. Viewed as IGMRFs, the models reflect local deviation from an unspecified overall level. Typically, an IGMRF can be constructed such that this deviation can be seen as a smooth curve in time or a smooth surface in space.

More specifically, we define an IGMRF as a vector having an improper Gaussian density with a sparse precision matrix not of full rank. Following Rue and Held (2005), the order of an IGMRF can be defined as the rank deficiency of its precision matrix. This implies that a zero-mean IGMRF of kth order is specified by

$$\pi(\boldsymbol{x}) = (2\pi)^{-(n-k)/2} (|\boldsymbol{Q}|^*)^{1/2} \exp\left\{-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q}\boldsymbol{x}\right\}, \tag{1}$$

where the precision matrix $\boldsymbol{Q}$ is assumed to be semi-positive definite. $|\boldsymbol{Q}|^*$ denotes the generalized determinant equal to the product of the $n-k$ non-zero eigenvalues of $\boldsymbol{Q}$. The marginal variance for $\boldsymbol{x}$ is given by the diagonal elements of the generalized inverse matrix $\boldsymbol{\Sigma}^* = \boldsymbol{Q}^{-1}$. Different IGMRFs can be described by expressing the precision matrix as $\boldsymbol{Q} = \tau \boldsymbol{R}$, where τ denotes the random precision parameter while the matrix $\boldsymbol{R}$ reflects the specific neighbourhood structure of the model, which can be represented as a labelled graph with nodes and edges.

2.1. IGMRFs on a line

To model smooth functions in one dimension we will apply either a first or a second-order IGMRF defined on a line, also referred to as first and second-order random walk models. The first-order model assumes independent first-order increments

$$\Delta x_i = x_{i+1} - x_i \sim N(0, \tau^{-1}), \quad i = 1, \ldots, n-1$$

where, we for simplicity, assume that the n nodes are equally spaced with distance 1. In order to get a proper joint density (in an $(n-1)$-dimensional subspace), we impose the constraint $\sum_i x_i = 0$, which also gives a finite marginal standard deviation. The joint density for $\boldsymbol{x}$ is then defined by

$$\pi(\boldsymbol{x} \mid \tau) \propto \tau^{(n-1)/2} \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2\right),$$

which gives a tridiagonal precision matrix in (1). This model captures local deviation from an overall constant level (represented by the null space of $\boldsymbol{Q}$). This is beneficial in applications where the underlying mean level is approximately or locally constant, see Section 3.3 in Rue and Held (2005) for details.

The second-order model assumes that the second-order increments

$$\Delta^2 x_i = x_{i+2} - 2x_{i+1} + x_i \sim N(0, \tau^{-1})$$

are independent. This model captures local deviation from a line (Rue and Held, 2005, Section 3.4.1). In order to define a proper joint density for $\boldsymbol{x}$, we need to impose that $\sum_i ix_i = 0$, in addition to the sum-to zero constraint used for first-order models. In this case the density of $\boldsymbol{x}$ is

$$\pi(\boldsymbol{x} \mid \tau) \propto \tau^{(n-2)/2} \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n-2}(x_i - 2x_{i+1} + x_{i+2})^2\right).$$

Expressed as in (1), the precision matrix will be a band matrix with bandwidth 5.

Due to the different structure of the two models, it seems unreasonable to assign the same hyperprior to the precision parameters. Also, as noted in Akerkar et al. (2010), the hyperprior for the precision of random walk models needs to be rescaled if the number of nodes is changed, in order to obtain the same variance as before. Assume that the interval between original nodes x_1 and x_2 is divided into k equally sized subintervals giving new equidistant nodes $x'_1, \ldots, x'_{k+1}$. For the first-order model, the precision using the new nodes is

$$\tau_{\text{new}}^{-1} = \text{Var}(x'_2 - x'_1) = \frac{1}{k}\text{Var}(x'_{k+1} - x'_1) = \frac{1}{k}\text{Var}(x_2 - x_1) = (k\tau)^{-1}. \tag{2}$$

For the second-order model, the precision is

$$\tau_{\text{new}}^{-1} = \text{Var}(x'_3 - 2x'_1 - x'_i) = (k^3\tau)^{-1}, \tag{3}$$

as derived in Lindgren and Rue (2008). These identities can be used directly to rescale an original hyperprior, when the number or distance between nodes is changed.

2.2. IGMRFs for spatially structured effects

To model spatial dependence for regional data, we apply a first-order IGMRF for an irregular lattice (Rue and Held, 2005, Section 3.3.2), often referred to as the Besag-model. This model is defined by the density

$$\pi(\boldsymbol{x} \mid \tau) \propto \tau^{(n-1)/2} \exp\left(-\frac{\tau}{2}\sum_{i\sim j} w_{ij}(x_i - x_j)^2\right), \tag{4}$$

where the numbers w_{ij} denote positive and symmetric weights for all pairs of adjacent nodes. Again, we impose a sum-to-zero constraint to get a proper joint density with a finite marginal variance. The marginal variance of this model will depend on both the shape and the number of nodes of the graph.

In modelling spatial autocorrelation in two dimensions we will use the second-order IGMRF on a regular lattice (Rue and Held, 2005, Section 3.4.2). This model is constructed by assigning specific weights to different first- and second-order neighbours, and the resulting model will approximate a thin plate spline. More specifically, the conditional mean for a specific node can be described by a weight matrix in which the four first-order neighbours in the cardinal directions are given weight 8. Further, the nearest neighbours on the diagonals are given weight -2, while the second-order neighbours in the cardinal directions are given weight -1. The resulting model penalizes local deviation from a plane and again the marginal standard deviation of the model, integrating out the random precision, indicates how large we allow this deviation to be. The marginal standard deviation is calculated under the linear constraints $\sum x_{ij} = \sum ix_{ij} = \sum jx_{ij} = 0$. As with the other IGMRFs, the hyperprior of the precision parameter of this model should be rescaled if the resolution of the analysis is changed. Assume that an original lattice of size $n \times n$ is divided into $kn \times kn$ grid cells. The new marginal variance is given by

$$\tau_{\text{new}}^{-1} = (k^2\tau)^{-1}, \tag{5}$$

see Lindgren et al. (2011), and the hyperprior can be rescaled accordingly.

3. Specifying hyperpriors for the precision of IGMRFs

In lack of good guidelines, hyperpriors are often chosen in a rather casual way. For example, it is quite common to use the exact same hyperprior for precision parameters, even though the underlying prior models are very different. The previous section illustrated that it is unnatural to assign the same hyperprior for the precisions of different IGMRF-models, as the structure and the marginal variances of these models are unequal. In this section, we propose to calculate a reference standard deviation for a given model and use this as a guideline in selecting the parameters of a specific hyperprior.

3.1. Mapping the random precision to the marginal standard deviation

For any fixed precision τ, the marginal standard deviation of the components of a Gaussian vector $\boldsymbol{x}$ can be expressed, as a function of τ, by

$$\sigma_\tau(x_i) = \frac{\sigma_{\{\tau=1\}}(x_i)}{\sqrt{\tau}}, \quad i = 1, \ldots, n.$$

For a given IGMRF $\boldsymbol{x}$ with random precision τ, we propose to calculate a reference standard deviation for fixed $\tau = 1$ and then approximate the marginal standard deviation of all components of $\boldsymbol{x}$ by

$$\sigma(x_i) \approx \frac{\sigma_{\text{ref}}(\boldsymbol{x})}{\sqrt{\tau}}, \quad i = 1, \ldots, n. \tag{6}$$

We choose to calculate the reference standard deviation as a geometric mean as this measure can be seen to represent a typical level for a set of positive numbers. Also, it downweights large observations compared with calculating an ordinary average. The reference standard deviation for $\boldsymbol{x}$ is then calculated by

$$\sigma_{\text{ref}}(\boldsymbol{x}) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log\sigma_{\{\tau=1\}}(x_i)\right) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log(\Sigma^*_{ii})\right) \tag{7}$$

where the values Σ^*_{ii} denote the diagonal elements of the generalized inverse matrix $\boldsymbol{\Sigma}^* = \boldsymbol{Q}^{-1}$ in (1), calculated for $\tau = 1$. Note that the reference standard deviation needs to be computed using the given linear constraints for the different models, as otherwise it would not be finite.

Fig. 1 illustrates the marginal standard deviations using fixed precision $\tau = 1$ for all components of the first and second-order IGMRFs on a line, each defined on a graph with $n = 100$ equidistant nodes. Due to the different structure of these two models, both the shape and the level of these curves are different. The calculated reference standard deviations in (7) become equal to $\sigma_{\text{ref}}(\boldsymbol{x}_{\text{rw1}}) = 3.89$ and $\sigma_{\text{ref}}(\boldsymbol{x}_{\text{rw2}}) = 41.39$, respectively. This illustrates that if we choose the same hyperprior for the precisions of these two models, we allow the local deviation of the second-order model to be much larger than the local deviation for the first-order model.

In order to avoid this, we suggest that the mapping in (6) is used to define an upper limit for the marginal standard deviation. A priori, this indicates how large we allow the local deviation, or the influence of the different random effects in a regression model, to be. We denote the upper limit by U and define it by

$$P(\sigma(x_i) > U) \approx P\left(\frac{\tau}{\sigma^2_{\text{ref}}(\boldsymbol{x})} < \frac{1}{U^2}\right) = \alpha, \tag{8}$$

where α is a fixed small probability. This illustrates that by assigning a hyperprior to

$$\frac{\tau}{\sigma^2_{\text{ref}}(\boldsymbol{x})}, \tag{9}$$

the interpretation of the hyperprior remains the same for different models. This allows us to recalculate hyperpriors between different IGMRF-models and to impose a common upper limit for the standard deviation of different models. The hyperprior will be automatically scaled according to the

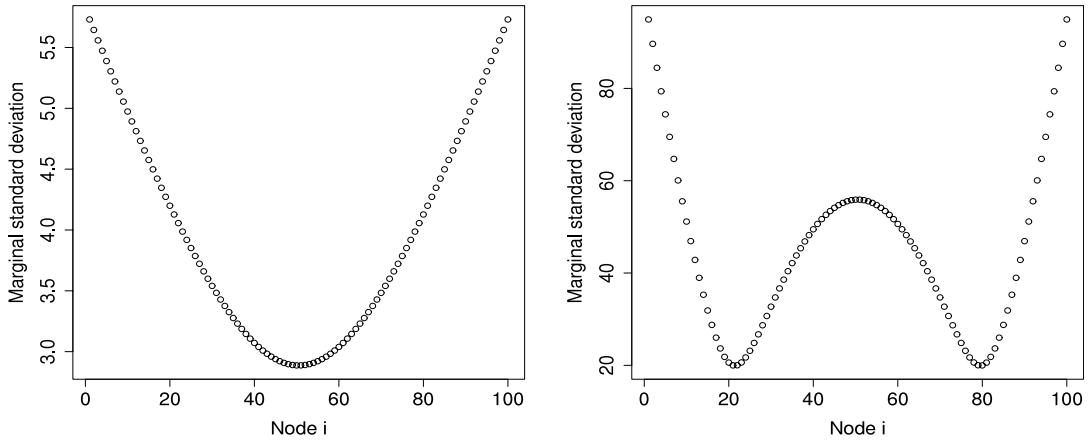

Fig. 1. The marginal standard deviation of a first-order (left panel) and a second-order (right panel) IGMRF on a line, calculated using fixed precision $\tau = 1$.

size and shape of the graph for different IGMRFs as this is captured by the reference standard deviation. Specifically, the calculated reference variances will be in coherence with the theoretical results in Eqs. (2), (3) and (5). Also, parameters for a given hyperprior can be tuned in an intuitive way by varying the upper limit U for a specific value of α.

3.2. Specifications using Gamma hyperpriors

The Gamma distribution represents a natural hyperprior for the precision parameters of IGMRFs, being a conjugate prior. If the precision parameter is assigned a Gamma distribution, $\tau \sim \text{Gamma}(a, b)$, the hyperprior assigned to (9) is simply given by $\text{Gamma}(a, b\sigma^2_{\text{ref}}(\boldsymbol{x}))$. This implies that we can easily account for different marginal variances of IGMRFs, using a common shape parameter a and an inverse-scale parameter adjusted to the model. For example, if we want to recalculate the hyperprior between a first and second-order IGMRF on the line, we can use the same shape parameter for both models and calculate a new inverse-scale parameter

$$b_{\text{rw2}} = b_{\text{rw1}} \frac{\sigma^2_{\text{ref}}(\boldsymbol{x}_{\text{rw1}})}{\sigma^2_{\text{ref}}(\boldsymbol{x}_{\text{rw2}})}.$$

Similarly, we only need to recalculate the inverse-scale parameter to account for different shapes and sizes of the graph for a specific IGMRF.

Applying a Gamma hyperprior, we notice that the upper limit defined in (8) can be expressed as

$$U = \left(\frac{b\sigma^2_{\text{ref}}(\boldsymbol{x})}{F^{-1}(\alpha, a, 1)} \right)^{1/2}, \tag{10}$$

where $F^{-1}(\cdot)$ denotes the inverse cumulative distribution function for a Gamma distributed variable. For a given value of α, we can then interpret the shape and inverse-scale parameters a and b in terms of this upper limit. This is also useful to get an idea about what values of U seem reasonable for a specific application.

4. Examples of spatial modelling

In the following two examples, we explore the ideas of scaling hyperpriors according to the different types of IGMRFs discussed. In the first example, we analyse spatial regional data, imposing a common upper limit for the marginal standard deviations. In the second example, we analyse a spatial point pattern and demonstrate that tuning of the hyperprior for the spatially structured effect is essential to the conclusions made on significance of environmental covariates.

Table 1
Reference standard deviations for different IGMRFs and the upper limits in (10) for the corresponding marginal standard deviations. These are calculated using $\alpha = 0.001$ and a Gamma$(1, 5 \cdot 10^{-5})$ hyperprior for the precisions.

	besag	rw1 (*wbc*)	rw2 (*wbc*)	rw1 (*tpi*)	rw2 (*tpi*)
$\sigma_{\text{ref}}(\boldsymbol{x})$	0.64	8.68	458.08	1.55	2.68
Upper limit U (default prior)	0.14	1.94	102.4	0.35	0.60

4.1. Leukaemia survival data in Northwest England

We first review a dataset analysed in Henderson et al. (2002) concerning spatial variation in survival of adult acute myeloid leukaemia patients in the Northwestern part of England. Data are registered for $n = 1043$ patients, being diagnosed between 1982 and 1998. In addition to survival times of each patient, the given dataset also contains information on sex, age and white blood cell counts (*wbc*) at time of diagnosis. Also, an index named the Townsend deprivation index (*tpi*) is used to measure social deprivation, in which higher and lower values indicate poorer and more wealthy regions, respectively.

Henderson et al. (2002) analysed the given data applying a multivariate frailty approach, including linear predictors for the covariate effects together with possible spatial variation based on 24 districts. In Kneib and Fahrmeir (2007), the given dataset was analysed using a mixed model approach, modelling covariates as penalized splines. Here, we use the same approach as in Martino et al. (2011) and analyse the data in R-INLA, using a piecewise log-constant Cox model. We consider the following model, in which the survival time of the patients are linked to a predictor

$$\eta_i(t) = f_0(t) + \beta_1 z_{1i} + \beta_2 z_{2i} + f_1(wbc_i) + f_2(tpi_i) + f_{spat}(s_i), \quad i = 1, \ldots, n,$$

where $f_0(t)$ denotes a log-baseline function being piecewise constant along the time axis. We include *sex* (z_{1i}) and *age* (z_{2i}) as fixed linear effects and consider both first and second-order IGMRFs to model the effect of the continuous covariates (*wbc* and *tpi*). We also include a spatially structured effect f_{spat} for the 24 districts, using the Besag-model in (4). For identifiability reasons, all functions for the random effects are constrained to sum to 0.

In defining hyperpriors for the three precision parameters $\boldsymbol{\theta} = (\tau_{wbc}, \tau_{tpi}, \tau_{spat})$, we first calculate reference standard deviations for the relevant IGMRFs. Here, the continuous covariates *wbc* and *tpi*, are discretized to have 50 unique values. Due to the wide range of *wbc*, the reference standard deviation for this covariate is much larger than for *tpi*, see Table 1. The table also displays the upper limits in (10) when the precisions are assigned Gamma$(1, 5 \cdot 10^{-5})$ and α is set equal to 0.001.

In order to scale the hyperpriors, we choose to fix the shape parameter $a = 1$ and impose a common upper limit for the marginal standard deviations equal to $U = 0.50$. This value is chosen based on what seems to be a reasonable value looking at the upper limits using the default hyperpriors in Table 1. The adjusted inverse-scale parameters accounting for the different models and resolutions are then given by $b \approx 6.1 \cdot 10^{-4}$ for the Besag-model while

$$b_{wbc,\text{rw1}} \approx 4.0 \cdot 10^{-6}, \qquad b_{wbc,\text{rw2}} \approx 1.5 \cdot 10^{-9},$$
$$b_{tpi,\text{rw1}} \approx 1.1 \cdot 10^{-4} \quad \text{and} \quad b_{tpi,\text{rw2}} \approx 4.0 \cdot 10^{-5}$$

for the two continuous covariates, using either the first or second-order IGMRF. We notice that these parameters are not very different from using a Gamma$(1, 5 \cdot 10^{-5})$ prior, except for the value $b_{wbc,\text{rw2}}$ which becomes very low to account for the large marginal variance of the corresponding second-order IGMRF.

Fig. 2 illustrates the estimated functions of the continuous covariates, using the second-order IGMRFs. We notice that *wbc* is too variable if the hyperprior for the precision of this model is not scaled appropriately. Using the scaled version, the effect of this covariate is seen to be linear, which is coherent with previous analyses (Martino et al., 2011; Sørbye and Rue, 2011). Differences in using either scaled or rescaled hyperpriors for *tpi* are small. Also, there were no significant differences in the estimated spatial effects using the default and the scaled hyperpriors (results not shown).

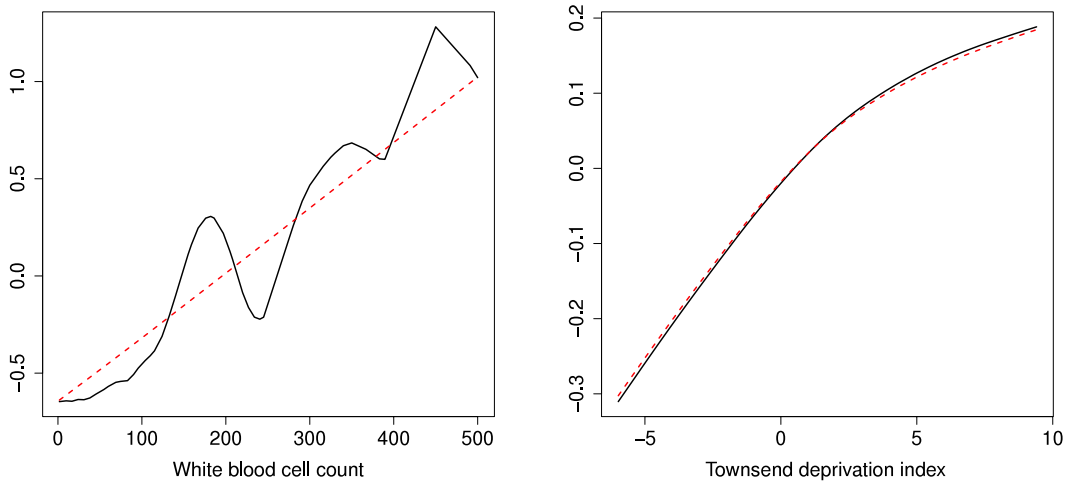

Fig. 2. The estimated effect of white blood cell count (left panel) and the Townsend deprivation index (right panel), using a second-order IGMRF with default hyperprior (black solid line) and scaled hyperprior (red dashed line).

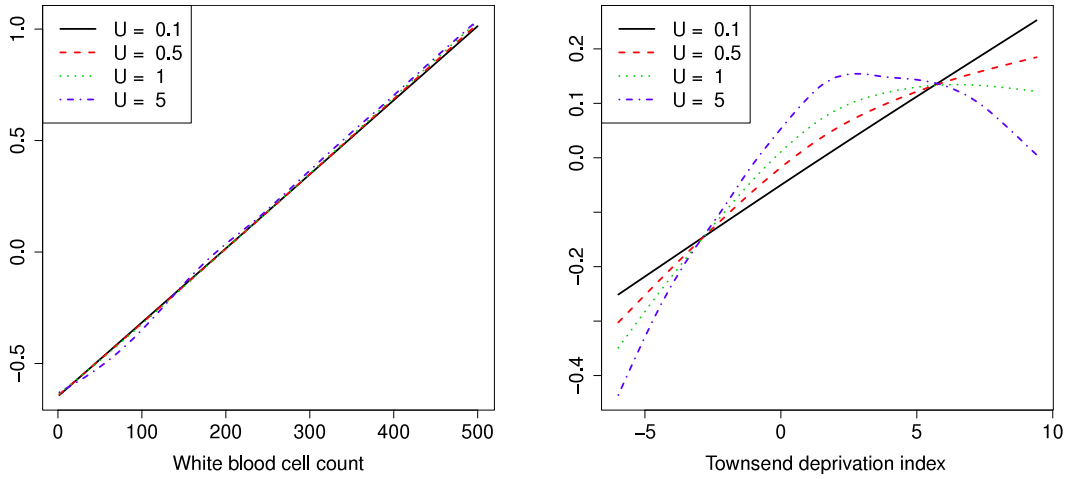

Fig. 3. The estimated effect of *wbc* (left panel) and *tpi* (right panel), using a second-order IGMRF on a line and different upper limits U, as defined in (10), where $\alpha = 0.001$ and $a = 1$.

By varying the upper limit U in (10), we can tune the hyperprior parameters simultaneously for all random effects modelled by IGMRFs. In the current example, we keep the shape parameter fixed at $a = 1$, and recalculate the inverse-scale parameters according to chosen values of U. Fig. 3 displays the resulting estimated effects using a second-order IGMRF in modelling *wbc* and *tpi*. As U is increased, the resulting effect for *wbc* remains linear while the resulting estimated effect of *tpi* becomes more variable. The corresponding estimated spatial effects did not seem to be very sensitive to tuning of the hyperprior.

4.2. Spatial point pattern model including environmental covariates

In this example we analyse the spatial point pattern formed by the species *Protium Tenuifolium*, being one of many tree species registered in a rainforest dataset from Barro Colorado, Panama. The full dataset is derived from a global network of plots coordinated by the Center for Tropical Forest Science (CTFS) and includes a large number of different tree species, observed within a 50-ha plot. In addition to the location of the trees, this dataset also contains information on environmental covariates, including two topography covariates and thirteen different soil covariates.

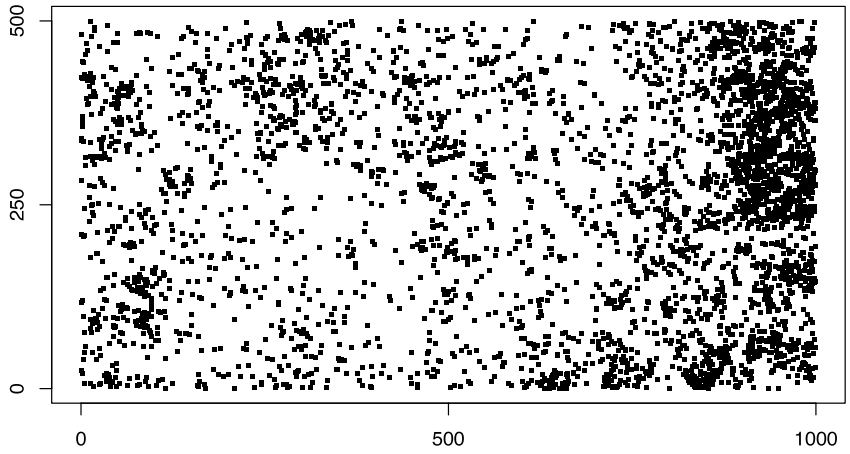

Fig. 4. The point pattern formed by the species *Protium Tenuifolium* observed within a 50-ha plot in Barro Colorado, Panama.

The species under consideration consists of a total of 4294 trees (Fig. 4) and is clearly seen to exhibit spatial inhomogeneity as the point intensity is higher on the right hand side of the plot. An important aim in analysing this type of data is typically to assess the influence of the environmental covariates. In a model, we also need to account for spatial autocorrelation and random variation not explained by the covariates.

4.2.1. Modelling approach

We choose to analyse the given spatial point pattern as a realization of a log-Gaussian Cox process (Møller et al., 1998), in which the point intensity is modelled by

$$\Lambda(s) = \exp\{\eta(s)\},$$

where $\{\eta(s) : s \in^2 \mathbb{R}\}$ is a Gaussian random field. We discretize a given region of interest into n grid cells, $s_1, \ldots, s_n$. Conditioned on the value of the random field $\eta(s)$ within a grid cell s_i, the observed number of points is assumed to follow a Poisson distribution,

$$y_i|\eta(s) \sim \text{Poisson}\left(\int_{s_i} \exp(\eta(s))ds\right) \approx \text{Poisson}(|s_i| \exp(\eta_i)), \quad i = 1, \ldots, n.$$

Here, the grid cell area is $|s_i|$ while η_i is a representative value of the Gaussian field within cell s_i. Further, we assume that η_i can be modelled as a linear predictor

$$\eta_i = \alpha + \sum_{j=1}^{p} \beta_j z_{ji} + f_{spat}(s_i) + u_i, \quad i = 1, \ldots, n, \tag{11}$$

where the vector $\boldsymbol{\beta}$ represents the linear effects of covariates $\{\boldsymbol{z}_j\}_{j=1}^{p}$. The function $f_{spat}(\cdot)$ represents a spatially structured random effect, here modelled by a second-order IGMRF defined on a lattice. The error term $\boldsymbol{u}$ denotes spatially unstructured i.i.d. effects with a random precision, here assigned a Gamma$(1, 5 \cdot 10^{-5})$ hyperprior. In the current analysis, we divide the given region of interest into $n = 50 \times 100$ grid cells, each cell having an area equal to 100 m^2.

The observed environmental covariates are seen to be highly correlated. Based on standard GLM-analysis (omitting the spatially structured and unstructured terms) and backward selection, we select a subset of covariates in which the variance inflation factor (VIF) is less than 5 to avoid problems with multicollinearity. The remaining environmental covariates include terrain slope (*Slope*), together with soil covariates giving content of Aluminium (*Al*), Copper (*Cu*), Iron (*Fe*), Manganese (*Mn*), Phosphorus (*P*), Zinc (*Zn*), Nitrogen (*N*) and also the pH-value (*pH*) of the soil. All of these covariates are log-transformed and standardized prior to the analysis.

The estimated means and the 95% pointwise credible intervals for the effects of the environmental covariates are given Fig. 5. The left panel illustrates that all of the environmental covariates in (11) are

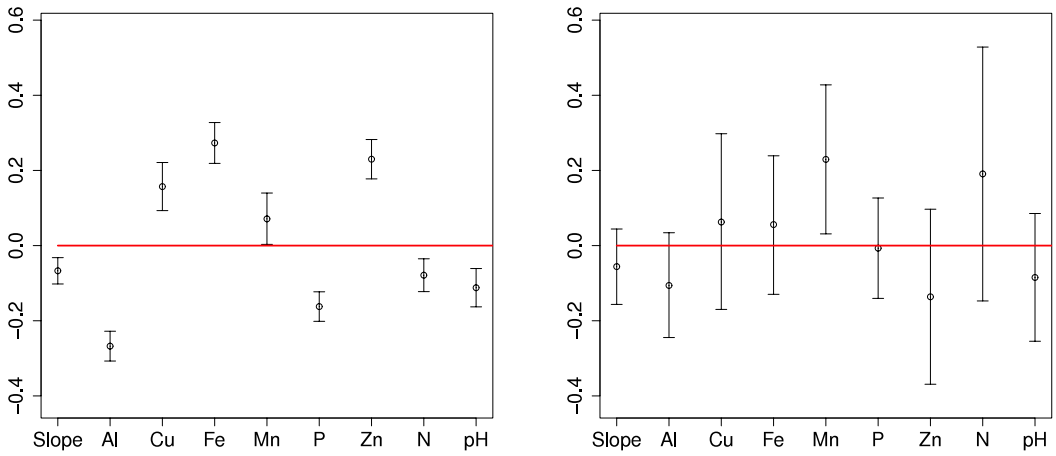

Fig. 5. Estimated coefficients of the environmental covariates and the 95% credible intervals, including only linear effects in (11) (left panel) or the full model analysed using default hyperpriors in R-INLA (right panel).

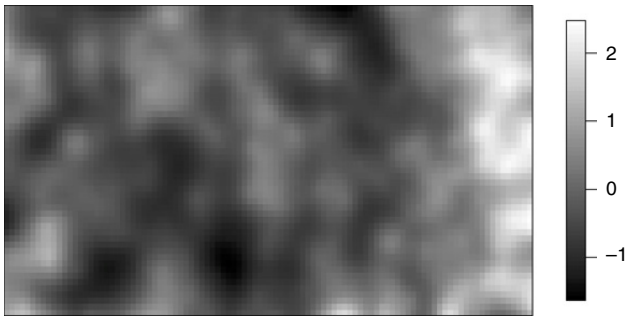

Fig. 6. Estimated spatially structured field for the rainforest species *Protium Tenuifolium*, using a Gamma$(1, 5\cdot 10^{-5})$ hyperprior for the precision of the second-order IGMRF on a lattice.

significant at a 5% level, when the spatially structured and unstructured terms are omitted. However, the variance is clearly underestimated. When the unstructured error term is included to model random variation, the variances increase slightly but only *Slope* becomes non-significant (results not shown). If the spatially structured effect is also included, all of the covariates become non-significant, except for *Mn* (right panel). Here we have used a Gamma$(1, 5 \cdot 10^{-5})$ hyperprior for the precision of the spatial model.

4.2.2. Tuning of the spatially structured effect

The hyperprior chosen for the precision parameter of the spatial term is crucial as it governs the smoothness of the resulting estimated field. If the field is too smooth, spatial autocorrelation is not sufficiently explained (like when the spatial field is omitted). This implies that the variance is underestimated and conclusions on significance of covariates might be invalid. On the other hand, a spatial field with a large variance can give a too detailed spatial effect such that the resulting field is overfitted to the point pattern and potential significant impacts of environmental covariates are not revealed. To investigate this further, we need to tune the hyperprior for the precision of the spatial effect. We have noticed that using the default Gamma$(1, 5\cdot 10^{-5})$ prior in R-INLA, the upper limit for the marginal standard deviation equals $U = 1.85$ (using $\alpha = 0.001$) and this gives a rather detailed spatial effect (Fig. 6). To explain spatial autocorrelation at a larger scale, the spatial effect needs to be more smooth. This can be achieved either by increasing the shape parameter a and/or decreasing the inverse-scale parameter b of the Gamma hyperprior.

Tuning of the hyperprior parameters can be performed in several ways. One option is to use the same strategy as in the previous example, in which the upper limit in (8) is decreased to give a

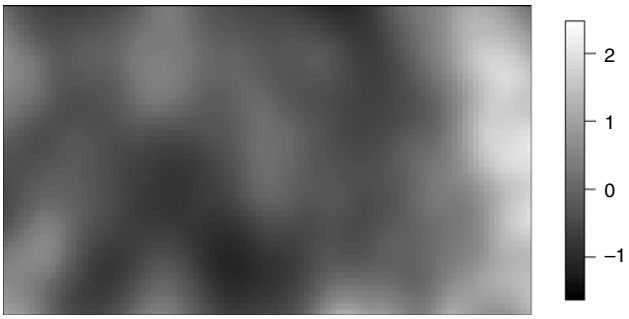

Fig. 7. Estimated spatially structured field for the rainforest species *Protium Tenuifolium*, using a Gamma(25, 0.0451) hyperprior for the precision of the second-order IGMRF on a lattice .

smoother spatial effect. For the given example, we choose to first increase the value of the shape parameter a, as this will explain spatial autocorrelation at larger scales (Beguin et al., 2012; Illian et al., 2012). Notice that using a Gamma hyperprior, the shape parameter a is related to the coefficient of variation c_v as

$$a = c_v^{-2} = \left(\frac{\mathrm{E}(\tau)}{\sqrt{\mathrm{Var}(\tau)}} \right)^2 .$$

This implies that by selecting $a = 1$, the standard deviation of the precision is allowed to be of the same size as its mean. In order to reduce the standard deviation compared with the mean, we increase the shape parameter. Typically, a range of values for a needs to be explored as we do need to be sure that spatial autocorrelation is sufficiently explained. Here, we present results for shape parameter $a = 25$, while the inverse scale parameter is chosen according to an upper limit $U = 0.5$. The resulting estimated spatial effect is illustrated in Fig. 7 while the estimated means and 95% credible intervals for the covariates are given in Fig. 8. The content of Manganese is still seen to have a positive effect on the estimated point pattern intensity. In addition, we find that the soil content of Aluminium has a significantly negative impact. These results are in accordance with Schreeg et al. (2010), where Manganese is said to be an essential plant nutrient, while Aluminium is nonessential for plant growth, possibly having a toxic effect. These conclusions remain when applying a wider range of shape parameters.

5. Concluding remarks

The main focus of the given paper is to illustrate that the marginal standard deviations of commonly applied IGMRF priors can be very different, and we believe that this should be accounted for in assigning hyperpriors to the precision parameters of these models. Using `R-INLA` or other software for Bayesian analysis, the user typically needs to choose and scale these hyperpriors and currently there are no clear guidelines to how this should be done. By mapping the precision to the marginal standard deviation, we can define hyperpriors such that the marginal variances for different models are similar. More specifically, we can easily recalculate a given hyperprior to account for a different model. This implies that in cases where a linear predictor consists of several IGMRFs, we can assign hyperpriors such that these models have comparable influence, a priori.

The ideas presented here do rely on making some subjective choices, including the selection of an upper limit for the marginal standard deviations. An advantage is that the marginal standard deviation of an IGMRF has an intuitive interpretation in terms of local deviation from an unspecified level, here being either a constant, a line or a plane. This interpretation can be used to give a better understanding of different hyperprior choices. Also, the parameters of different hyperpriors can be tuned simultaneously by varying the upper limit. However, the exact value for the upper limit has to be interpreted with care, as this relies on the choice of a specific quantile probability and also on the specific application under consideration.

The given suggestions are easily implemented and can be used to make the selection of hyperpriors less ad hoc compared with making purely subjective choices. In this paper, we have only looked at

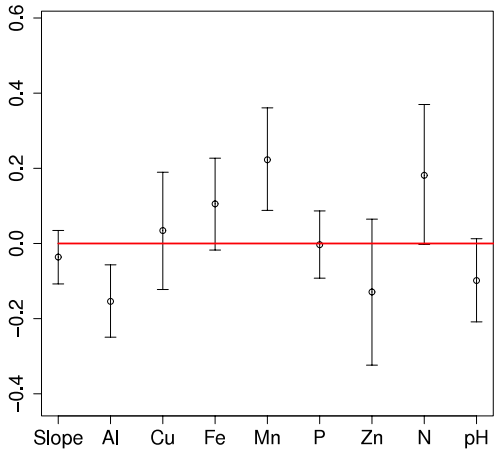

Fig. 8. The estimated mean and 95% credible intervals for the covariates, modelling the rainforest species *Protium Tenuifolium* using a Gamma(25, 0.0451) hyperprior for the precision of the second-order IGMRF on a lattice.

the class of IGMRF-priors due to the specific structure and characteristics of these type of models. However, we believe that guidelines for hyperprior choices of these models are valuable. Especially, these models are used extensively as priors in R-INLA and we aim to implement the given ideas as an easily-available option in this software. Naturally, there are many other commonly applied priors in which guidelines for hyperprior selection would be very welcome. We believe that similar ideas to the ones presented can also be applied for auto-regressive processes. We would also find it very useful to have guidelines for hyperprior choices of the Matérn correlation function, which is often used to model underlying spatial dependency structures. Further research is warranted to explore this further. In general, prior selection remains a complicated field and an important subject of future research.

Acknowledgements

The rainforest dataset has been collected with the support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Celera Foundation, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.

References

Akerkar, R., Martino, S., Rue, H., 2010. Implementing approximate Bayesian inference for survival analysis using integrated nested Laplace approximations. In: Preprint Statistics, Vol. 1. Norwegian University of Science and Technology, pp. 1–38.
Beguin, J., Martino, S., Rue, H., Cumming, S.G., 2012. Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. Methods in Ecology and Evolution 3, 921–929.
Berger, J.O., Strawderman, W., Tang, D., 2005. Posterior propriety and admissibility of hyperpriors in normal hierarchical models. The Annals of Statistics 33, 606–646.
Besag, J., Kooperberg, C., 1995. On conditional and intrinsic autoregressions. Biometrika 82, 733–746.
Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics 43, 1–59.
Fahrmeir, L., Tutz, G., 2001. Multivariate Statistical Modelling Based on Generalized Linear Models, second ed. Springer, Berlin.
Fong, Y., Rue, H., Wakefield, J., 2010. Bayesian inference for generalized linear mixed models. Biostatistics 11, 397–412.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis. Chapman & Hall, London, Boca Raton.
Henderson, R., Shimakura, S., Gorst, D., 2002. Modeling spatial variation in leukemia survival data. Journal of the American Statistical Association 97, 965–972.
Illian, J.B., Sørbye, S.H., Rue, H., Hendrichsen, D.K., 2012. Fitting a log Gaussian Cox process with temporally varying effects—a case study. Journal of Environmental Statistics 3, 1–29.
Kneib, T., Fahrmeir, L., 2007. A mixed model approach for geoadditive hazard regression. Scandinavian Journal of Statistics 34, 207–228.
Lesaffre, E., Lawson, A.B., 2012. Bayesian Biostatistics. John Wiley & Sons, Ltd.

Lindgren, F., Rue, H., 2008. On the second-order random walk model for irregular locations. Scandinavian Journal of Statistics 35, 691–700.
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. Series B 73, 423–498.
Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: evolution, critique and future directions. Statistics in Medicine 28, 3049–3067.
Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10, 325–337.
Martino, S., Akerkar, R., Rue, H., 2011. Approximate Bayesian inference for survival models. Scandinavian Journal of Statistics 38, 514–528.
Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log Gaussian Cox processes. Scandinavian Journal of Statistics 25, 451–482.
Roos, M., Held, L., 2011. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. Bayesian Analysis 6, 259–278.
Rue, H., Held, L., 2005. Gaussian Markov Random Fields. Chapman & Hall/CRC, Boca Raton.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society. Series B 71, 319–392.
Schreeg, L.A., Kress, W.J., Erickson, D.L., Swenson, N.G., 2010. Phylogenetic analysis of local-scale tree soil associations in a lowland moist tropical forest. PLoS One 5, 1–10.
Sørbye, S.H., Rue, H., 2011. Simultaneous credible bands for latent Gaussian models. Scandinavian Journal of Statistics 38, 712–725.