

Correspondence to: Virgilio Gómez-Rubio

Title: Spatial models using Laplace approximation methods

Name: Virgilio Gómez-Rubio^{1,*}, Roger S. Bivand², Håvard Rue³

Affil./Addr. 1: Department of Mathematics
School of Industrial Engineering-Albacete
University of Castilla-La Mancha
Avda. España s/n, 02071 Albacete, Spain
E-mail: {Virgilio.Gomez, Franciso.Palmi}@uclm.es

*: Corresponding author

Affil./Addr. 2: Department of Economics
Norwegian School of Economics
Helleveien 30, N-5045 Bergen, Norway
E-mail: Roger.Bivand.nhh.no

Affil./Addr. 3: CEMSE Division
King Abdullah University of Science and Technology
Thuwal 23955-6900, Saudi Arabia
E-mail: Haavard.Rue@kaust.edu.sa

Spatial models using Laplace approximation methods

Abstract

Bayesian inference has been at the center of the development of spatial statistics in recent years. In particular, Bayesian hierarchical models including several fixed and random effects have become very popular in many different fields. Given that inference on these models is seldom available in closed form, model fitting is usually based on simulation methods such as Markov chain Monte Carlo.

However, these methods are often very computationally expensive and a number of approximations have been developed. The integrated nested Laplace approximation (INLA) provides a general approach to computing the posterior marginals of the parameters in the model. INLA focuses on latent Gaussian models, but this is a class of methods wide enough to tackle a large number of problems in spatial statistics.

In this chapter, we describe the main advantages of the integrated nested Laplace approximation. Applications to many different problems in spatial statistics will be discussed as well.

Keywords

areal data, geostatistics, integrated nested Laplace approximation, multivariate statistics, point patterns, spatial statistics

Introduction

Spatial models provide a suitable way of analyzing data when observations are thought to be correlated because of their locations in space. Bayesian inference has proven useful when dealing with spatial models and modeling local dependence. In Bayesian analysis (see, for example, Gelman et al 2003), inference about the vector of model parameters \boldsymbol{x} is based on computing their joint posterior distribution given the vector of observed data \boldsymbol{y} . This is done by means of Bayes' rule:

$$\pi(\boldsymbol{x}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}) \quad (1)$$

Here $\pi(\boldsymbol{y}|\boldsymbol{x})$ represents the likelihood of the model given its parameters and $\pi(\boldsymbol{x})$ is the prior distribution of the parameters of the model. Hence, the posterior distribution depends on the mechanism which generates the data (i.e., the likelihood) and the

Bayesian method
spatial models
integrated nested
Laplace
approximation
Laplace
approximation
INLA—see integrat
nested Laplace
approximation

previous information about the model parameters (i.e., the prior distribution). Note that $\pi(\mathbf{x})$ is often supposed to depend on some hyperparameters which in turn have their own prior distributions.

$\pi(\mathbf{x}|\mathbf{y})$ is a multivariate distribution of the ensemble of model parameters which is often hard to obtain. In many applications it is sufficient with obtaining a separate posterior distribution for some of the parameters in the model because no joint inference is needed (for example, the estimates of the relative risk in different areas). These distributions are called posterior marginals and can be denoted $\pi(x_l|\mathbf{y})$. As these are univariate distributions, they are often easier to compute or approximate than the joint posterior distribution.

Given that in most cases there is no closed form for the posterior distributions of most parameters in the model, Markov chain Monte Carlo (MCMC, see Gelman et al 2003) techniques have been employed to estimate the joint posterior. Furthermore, a number of sound techniques for model criticism, comparison and selection make Bayesian inference appealing.

For models with complex spatial dependence or large datasets, MCMC may not be a convenient solution due to computational time. For this reason, Rue et al (2009) propose the use of approximate inference based on what they have called the integrated nested Laplace approximation (INLA). This approximation will focus on the posterior marginals which are easier to compute than obtaining an approximation to the joint posterior distribution. Also, INLA will only consider approximations for hierarchical models whose latent effects can be expressed as a Gaussian Markov random field (GMRF).

Successful applications of INLA include disease mapping geostatistics point patterns spatial econometrics and others. Furthermore, see Rue et al (2017) and Bakka et al (2018) for recent reviews of INLA.

Integrated Nested Laplace Approximation

The integrated nested Laplace approximation (INLA) focuses on providing a good approximation to the posterior marginal distributions of the parameters in the model. In particular, this approximation has been developed for latent Gaussian models. These cover a general class of models which appear in many areas of interest. Spatial statistics is one of them, as spatial correlation can be introduced by means of correlated random effects.

First of all, let us assume that we have n observed variables y_i , $i = 1, \dots, n$ with a distribution (usually from the exponential family) with a mean μ_i which is related to a linear predictor η_i through a convenient link function. In turn, η_i is modeled additively on different effects:

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \varepsilon_i \quad (2)$$

Here, $f^{(j)}$ represents some non-linear function or random effects (of which there are n_f) on a set of covariates \mathbf{u} , n_β is the number of covariates, β_k are coefficients for linear effects on a vector of covariates \mathbf{z} and ε_i are unstructured terms. The latent effects $\mathbf{x} = \{\{\eta_i\}, \alpha, \{\beta_k\}, \dots\}$ are assumed to be Gaussian with zero mean and precision matrix $\mathbf{Q}(\boldsymbol{\theta}_1)$, where $\boldsymbol{\theta}_1$ is a vector of hyperparameters. Hence, the observations will have a likelihood which will depend on the latent effects \mathbf{x} and a set of parameters $\boldsymbol{\theta}_2$. Furthermore, the observations y_i are supposed to be independent given \mathbf{x} and $\boldsymbol{\theta}_2$.

In the particular case of spatial statistics, terms $f^{(j)}(u_{ji})$ can be taken as $f_i^{(j)}$ (or u_i oversimplifying the notation) to represent a random effect at a spatial location i . Hence, covariate u_{ji} acts as the spatial index i of area i for the set of random effects j . For example, taking $n_f = 2$ we can define $u_i = f^1(u_{1i})$ and $v_i = f^2(u_{2i})$, where $\mathbf{u} = \{u_1, \dots, u_n\}$ is a vector of independent random effects and $\mathbf{v} = \{v_1, \dots, v_n\}$ a vector of spatially correlated random effects.

integrated nested
Laplace
approximation
Laplace
approximation

Rue et al (2009) focus on the posterior distribution of \mathbf{x} and the vector of hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$:

$$\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta}) \pi(\mathbf{x} | \boldsymbol{\theta}) \prod_{i \in \mathcal{I}} \pi(y_i | x_i, \boldsymbol{\theta}) \propto \quad (3)$$

$$\pi(\boldsymbol{\theta}) |\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\left\{-\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} + \sum_{i \in \mathcal{I}} \log(\pi(y_i | x_i, \boldsymbol{\theta}))\right\} \quad (4)$$

Here \mathcal{I} is the subset of indices (from 1 to n) that are observed with observations \mathbf{y} and their respective linear predictors $\{\eta_i\}$. Furthermore, the n first elements of \mathbf{x} are the linear predictors, i.e., $x_i = \eta_i$, $i = 1, \dots, n$. Note that η_i is the only observed latent effect (through y_i) and that all the other latent effects are not observed directly and need to be estimated. In addition, the latent effects may be subject to some linear constraints of the form $\mathbf{A}\mathbf{x} = \mathbf{e}$. Here, \mathbf{A} is a matrix and \mathbf{e} a vector, both of the appropriate dimensions, that represent the linear constraints of \mathbf{x} . Finally, the latent field is supposed to have conditional independence properties, so that \mathbf{x} becomes a Gaussian Markov random field (GMRF). As we will show later, these Markov properties play an important role when modeling spatial data.

The likelihood of the data $\pi(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta})$ is not constrained to be Gaussian. At the moment, INLA can deal with several likelihoods from the exponential family as well as with mixtures, such as zero-inflated distributions. Furthermore, INLA is flexible enough to allow different observations to have different likelihoods. Hence, INLA can deal with a myriad of models.

Instead of aiming at the full posterior distribution of the model parameters \mathbf{x} and $\boldsymbol{\theta}$, Rue et al (2009) focus on obtaining an approximation to the posterior marginal distributions $\pi(x_l | \mathbf{y})$ and $\pi(\theta_p | \mathbf{y})$. These marginals can be written down as

$$\pi(x_l | \mathbf{y}) = \int \pi(x_l | \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (5)$$

and

$$\pi(\theta_p|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-p} \quad (6)$$

Here $\boldsymbol{\theta}_{-p}$ denotes $\boldsymbol{\theta}$ minus component θ_p .

The approximations will be for the conditional distributions in the right hand sides of the previous expressions. Note that an approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$ is also required and that numerical integrations will be feasible only if the dimension of $\boldsymbol{\theta}$ is small (as it often happens in practice).

A first approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$ using Gaussian distributions can be constructed using the Laplace approximation:

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} \quad (7)$$

$\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation to the full conditional of \mathbf{x} and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional for a given value of $\boldsymbol{\theta}$.

Hence, the marginals of interest can be computed using numerical integration over a multidimensional grid of values of $\boldsymbol{\theta}$. For example,

$$\tilde{\pi}(x_l|\mathbf{y}) = \sum_{k=1}^K \tilde{\pi}(x_l|\boldsymbol{\theta}_k, \mathbf{y}) \times \tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y}) \times \Delta_k \quad (8)$$

where Δ_k represents the weights for each of the K vector of values $\boldsymbol{\theta}_k$ in the grid.

Rue et al (2009) stress the importance of having a good approximation $\tilde{\pi}(x_l|\boldsymbol{\theta}, \mathbf{y})$ to $\pi(x_l|\boldsymbol{\theta}, \mathbf{y})$. A Gaussian approximation $\tilde{\pi}_G(x_l|\boldsymbol{\theta}, \mathbf{y})$ is based on using a normal distribution with mean $\mu_l(\boldsymbol{\theta})$ and marginal variance $\sigma_l^2(\boldsymbol{\theta})$. The approximation provided by INLA (and in particular the Gaussian approximation for $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$) is exact for Gaussian data and the approximation is only due to integration (with respect to $\boldsymbol{\theta}$) error. This may be a good starting point, but it may not suffice because of possible inaccuracy if it is not centered at the correct point and because of its lack of skewness.

For this reason, they also propose other alternatives such as the Laplace approximation and the *integrated nested Laplace approximation (INLA)*. Firstly, an improved approximation may be obtained by using a Laplace approximation:

$$\tilde{\pi}_{LA}(x_l|\boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-l}|x_l, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-l}=\mathbf{x}_{-l}^*(x_l, \boldsymbol{\theta})} \quad (9) \quad \text{Gaussian Markov random field}$$

Here $\tilde{\pi}_{GG}(\mathbf{x}_{-l}|x_l, \boldsymbol{\theta}, \mathbf{y})$ is a Gaussian approximation to $\mathbf{x}_{-l}|x_l, \boldsymbol{\theta}, \mathbf{y}$ which is centered around the mode $\mathbf{x}_{-l}^*(x_l, \boldsymbol{\theta})$. As this approximation must be computed for every x_l , some numerical techniques are required to speed up computation. GMRF—see Gaussian Markov random field

Finally, Rue et al (2009) derive a simplified Laplace approximation to improve the approximation given by $\tilde{\pi}_{LA}(x_l|\boldsymbol{\theta}, \mathbf{y})$ by means of a series expansion of the Laplace approximation around $x_l = \mu_l(\boldsymbol{\theta})$. This provides a better approximation and it corrects for location and skewness. As $\tilde{\pi}_{LA}(x_l|\boldsymbol{\theta}, \mathbf{y})$ is very expensive to compute, the simplified Laplace approximation seems the best tradeoff between speed and accuracy.

It should be noted that while these approximations will center on the posterior marginal of a single latent effect x_l or hyperparameter θ_p the methodology behind them could be applied to obtain an approximation of the joint posterior of any subset S of latent effects \mathbf{x}_S (see Section 6.1, Rue et al 2009). However, in that case the approximations become more complex and the numerical integration needed is more demanding.

Gaussian Markov random field

Approximate inference using INLA is based on the assumption that the latent field \mathbf{x} is Gaussian, and fulfills some conditional independence properties. In particular, any two latent effects x_l and x_m in \mathbf{x} should be independent given the remaining latent effects \mathbf{x}_{-lm} . Furthermore, the number of hyperparameters appearing in the distribution of \mathbf{x} is assumed to be small.

Rue and Held (2005) provide a description of methods for efficient computation of Gaussian Markov random fields (GMRF) which can be used to speed up computations and provide fast approximations. GMRF are the key to providing good Gaussian

approximations for the posterior marginals. INLA is based on providing Gaussian approximations to densities like

$$\pi(\mathbf{x}|\boldsymbol{\theta}, y) \propto \exp\left\{-\frac{1}{2}\mathbf{x}^T \mathbf{Q} \mathbf{x} + \sum_{i \in \mathcal{I}} \log(y_i|x_i, \boldsymbol{\theta})\right\} \quad (10)$$

where \mathbf{Q} is the precision matrix of the GMRF. Note that if \mathbf{Q} is a known matrix its determinant (sometimes termed Jacobian) can be ignored at this stage as the posterior distribution can be rescaled later. This distribution may be subject to a set of linear constraints $\mathbf{A}\mathbf{x} = \mathbf{e}$. In any case, the approximation will result in a Gaussian distribution with mean \mathbf{x}^* and precision matrix $\mathbf{Q}^* = \mathbf{Q} + \text{diag}(\mathbf{c}^*)$, with $\text{diag}(\mathbf{c}^*)$ a diagonal matrix with entries \mathbf{c}^* (see Rue et al 2009, Section 2 for details). If linear constraints are present, the mean and precision matrix of the Gaussian approximation are conveniently corrected.

These constrained models are useful for fitting geostatistical models and adjacency-based spatial correlation effects for areal data (for example, using an intrinsic conditional autoregressive model). Other spatial and temporal random effects can be modeled by using intrinsic GMRFs with linear constraints (see Rue and Held 2005, Chapter 3). Linear constraints are often employed to impose a sum-to-zero constraint on intrinsic GMRFs in order to make these effects identifiable. This is particularly important when dealing with complex spatio-temporal effects.

Priors

So far, we have dealt with how the likelihood and the latent Gaussian Markov random fields are defined. As in all Bayesian approaches, a set of priors needs to be assigned to the θ parameters.

First of all, covariate coefficients in the linear predictor will be assigned a normal distribution with zero mean and precision τ . A similar distribution will be used for the random errors ε_i .

In principle, the latent random effects will be all Gaussian with zero mean. Hence, only the parameters in the precision matrix will need a prior. For the case in which the precision matrix is of the form $\tau\mathbf{Q}$, where \mathbf{Q} is a known matrix, τ can be assigned either a Gamma, truncated normal or improper flat (in the log-scale) prior. If the whole precision matrix is to be assigned a prior, then a Wishart distribution is available for correlated random effects of small dimension (up to five). Finally, the INLA software provides other prior distributions. For example, correlation parameters, such as the ones used to model spatial autocorrelation, can be assigned a beta prior.

Note that, for simple models, these choices are equivalent to setting a conjugate prior distribution and that in all cases the prior parameters are supposed to be known (i.e., these cannot be assigned a prior in turn). It should be mentioned that these priors are the ones implemented in the R-INLA package (available from <http://www.r-inla.org>), but user-defined priors can be used as well by providing the mathematical expression for them.

Other priors can be built on upon simpler prior specifications. For example, spatially varying coefficients on a covariate can be implemented by using a prior which is the sum of independent and spatially correlated random effects. More information about how priors can be specified are available in the documentation of the R-INLA package.

Model criticism and selection

INLA provides a number of ways of comparing and assessing models. First of all, an approximation to the marginal likelihood $\pi(\mathbf{y})$ is provided. This approximation is based on

$$\tilde{\pi}(\mathbf{y}) = \int \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} d\boldsymbol{\theta} \quad (11)$$

where $\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y}) = \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$. Models with a larger value of the marginal likelihood will be preferred. Also, marginal likelihood can be used to compute Bayes factors in order to compare models.

Predictive measures can also be computed very easily. In particular, INLA can compute the predictive distribution of y_i given all the other observations, i.e., $\pi(y_i|\mathbf{y}_{-i})$. Following Pettit (1990), INLA reports the probability integral transform (PIT):

$$\text{PIT}_i = \text{Prob}(y_i^{new} \leq y_i|\mathbf{y}_{-i}) \quad (12)$$

This criterion has been used to assess the validity of spatial models in disease mapping and it avoids the use of other sampling-based methods which may be less accurate.

Roos and Held (2011) discuss sensitivity to priors for binary data using the conditional predictive ordinate (CPO, Geisser 1993), which is defined as $\pi(y_i|\mathbf{y}_{-i})$. They use the mean logarithmic CPO to build the following statistic

$$\overline{\text{CPO}} = -\frac{1}{n} \sum_{i=1}^n \log(\pi(y_i|\mathbf{y}_{-i})) \quad (13)$$

as a measure of the predictive quality of the model. Lower values of $\overline{\text{CPO}}$ indicate a better model. As the authors state, this criterion can easily be extended to other hierarchical models. The CPO and PIT have been compared using “exact” Bayesian inference (using MCMC) and approximate inference (with INLA), showing that the approximated values are very close in general to the exact ones.

Finally, INLA can also compute the deviance information criterion (DIC) and the Watanabe–Akaike information criterion (WAIC) which are a popular way of comparing Bayesian hierarchical models. Both criteria also compute a measure of the effective number of parameters which is a measure of the complexity of the model.

PIT—see predictive
integral transfo
predictive integra
transform
CPO—see conditi
predictive
ordinate
conditional
predictive
ordinate
DIC—see devianc
information
criterion
deviance
information
criterion
WAIC—see Wata
Akaike
information
criterion
Watanabe-Akaike
information
criterion

Implementation

R programming
language
R-INLA

Besides the original paper, the authors have released open-source software (called INLA) which implements all the techniques mentioned here. In addition, an interface for the R programming language (R Development Core Team 2019) can be downloaded (from <http://www.r-inla.org>) which makes the use of the software easier and is able to produce summary statistics and plots of the results.

Other features

In addition to an easy to use interface, the INLA software provides some other features. The joint posterior distribution of the hyperparameters can be computed. In addition, it is possible to define several linear combinations of the latent effects so that their posterior marginals are computed. Furthermore, if several of these linear combinations are computed, the joint correlation matrix can be computed as well, and this can be used to approximate the joint posterior distribution.

Spatial Models

Spatial dependence can be modeled in different ways in Bayesian hierarchical models (Banerjee et al 2004). Given that INLA focuses on latent Gaussian models and given that the latent effects are Gaussian, spatial correlation can be embedded in the precision matrix. Furthermore, because of the Markov properties of the latent field these variance-covariance matrices are often very sparse. How these methods can be applied to the different areas of spatial statistics is discussed below.

Geoadditive mixed effects models

Geoadditive models appear when regression models on a set of covariates are combined with other types of random effects (Kamman and Wand 2003). A geoadditive model

will be based on modeling the mean μ_i at each location i on the sum of a set of fixed and random effects:

$$\mu_i = \mu + \mathbf{z}_i\boldsymbol{\beta} + u_i + v_i \quad (14)$$

where \mathbf{z}_i is a vector of covariates and $\boldsymbol{\beta}$ the vector of associated coefficients. \mathbf{u} is a vector of spatially correlated random effects, while \mathbf{v} is a vector of independent random effects.

Note that this modeling can be done regardless of the likelihood employed for the data. In the case of a generalized linear model (GLM) a convenient link function will be used to transform the linear predictor accordingly.

Other non-parametric approaches can be implemented taking advantage of this approach. Kammann and Wand (2003) show how Penalized Splines (P-Splines) can be expressed as a mixed-effects model. Lee and Durbán (2009) describe how P-Splines and a conditionally autoregressive (CAR) model can be used to model spatial data. They develop an expression of these models as mixed-effects models. Although this is not a fully Bayesian approach, these models could be fitted with INLA using the following representation:

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad (15)$$

Here \mathbf{X} and \mathbf{Z} represent design matrices for the fixed and random effects which have a particular structure derived from the fact that this mixed model represents a P-Spline. A fully Bayesian approach to P-Splines can be found in Lang and Brezger (2004) and it is based on imposing a prior on the coefficients $\boldsymbol{\gamma}$ of a design matrix \mathbf{B} (based on the basis functions):

$$\boldsymbol{\mu} = \mathbf{B}\boldsymbol{\gamma} \quad (16)$$

Different priors on $\boldsymbol{\gamma}$ lead to different types of splines (Fahrmeir and Kneib 2011). For producing smoothed values of an observed covariate using P-Splines, the prior should be a random walk. To achieve spatial smoothing, the prior on $\boldsymbol{\gamma}$ should be a GMRF

splines
P-splines
penalized splines

with spatial structure. See Lang and Brezger (2004) for details on how to define \mathbf{B} and the prior of γ for spatial smoothing.

Disease mapping

The analysis of public health data has played an important role in the development of spatial statistics in the last decades. Besag et al (1991) provided a suitable model in which spatial correlation and unstructured variation are combined in a geadditive way which is also computationally appealing. Other authors have extended this model later, some of them for spatio-temporal disease mapping.

Disease mapping is a particular example of the analysis of lattice data. In this case, observations are aggregated over some region (counties, states, health districts, etc.) and spatial models assume that neighboring areas will have similar behavior. Here, dependence is between neighbors and a popular criterion is that two areas are neighbors if they share a common boundary.

Besag et al (1991) proposed the use of two latent random effects: a spatially correlated one \mathbf{u} and an independent one \mathbf{v} . The first will account for any spatial correlation and the second will account for any other unstructured difference between the regions. While the non-structured random effects are Gaussian with zero mean and precision τI_n (where I_n is the identity matrix of size $n \times n$), the spatially correlated random effects are defined using conditional distributions given the values at the neighbors. This is equivalent to using an intrinsic GMRF (Rue and Held 2005, Chapter 3), which is known as intrinsic conditionally autoregressive (CAR) model.

In order to encode this spatial information into a GMRF with zero mean and precision \mathbf{Q} we will make use of the Markov property to note that if areas i and j are independent given the remaining areas then $Q_{ij} = Q_{ji} = 0$. Hence, the precision

disease mapping
CAR—secondi
autoregressive
model
conditionally
autoregressive
model

matrix \mathbf{Q} will be very sparse and the algorithms described in Rue and Held (2005) can be used for fast sampling from this GMRF.

In particular, the intrinsic CAR precision matrix is defined as

$$Q_{ij} = \kappa \begin{cases} n_i & i = j, \\ -1 & i \sim j, \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

Here $i \sim j$ means that areas i and j are neighbors, κ is a conditional precision and n_i is the number of neighbors of area i . This makes the conditional distribution of $u_i | \mathbf{u}_{-i}, \kappa$ Gaussian with mean $\frac{1}{n_i} \sum_{j \sim i} u_j$ and variance $\frac{1}{\kappa n_i}$.

Note that the intrinsic CAR is an improper GMRF of rank $n - 1$. For this reason the sum-to-zero constraint on the random effect is added so that these effects can be identified. This is a common assumption for random effects based on intrinsic GMRF (Martino and Rue 2010).

A proper version of the intrinsic CAR model is available and it has a precision matrix similar to the previous one but adding a term $d > 0$ to the diagonal elements, so that they become $Q_{ii} = n_i + d$. Here, $\log(d)$ is assigned a log-gamma prior distribution by default. Note that the main point of this model is to make the precision matrix strictly diagonally dominant so that it becomes invertible and the prior distribution is a proper one.

A more general approach is obtained when the precision matrix is defined as

$$\mathbf{Q} = \left(I - \frac{\rho}{\lambda_{max}} \mathbf{C} \right) \quad (18)$$

This can be used to define a general CAR spatial effect by taking \mathbf{C} as a matrix of spatial weights. A vignette in the `spdep` package shows how different spatial weights can be defined. ρ represents the spatial correlation (and it can be assigned a prior) and takes values between zero and one because the weight matrix is \mathbf{C} divided by λ_{max} , its maximum eigenvalue, and by default a Gaussian prior is on $\text{logit}(\rho)$. Note that this

will produce a proper distribution for the spatially correlated random effects. Negative spatial autocorrelation is often ignored in disease mapping.

In this general case, the conditional distribution of u_i is

$$u_i | \mathbf{u}_{-i}, \kappa \sim \mathcal{N} \left(\rho \frac{\sum_{j \neq i} w_{ij} u_j}{w_{i+}}, \frac{1}{\kappa w_{i+}} \right) \quad (19)$$

where $w_{ij} = c_{ij} / \lambda_{max}$ and $w_{i+} = \sum_{j=1}^n w_{ij}$. Note that if \mathbf{C} is row-standardized then $\lambda_{max} = 1$ and $w_{i+} = 1$ and the marginal distribution has a simpler form.

Geostatistical models

In addition to fitting a model to the data, geostatistics focuses on predicting a continuous surface (often approximated by a discrete grid of points) so these models are often computationally very expensive. Spatially correlated random effects are built for the set of sampling locations, which may lead to trouble if the number of locations is large. Geostatistical models are not restricted to Gaussian likelihoods, as described in Banerjee et al (2004), and they can be used to model other types of data using a geostatistical latent effect.

Spatial correlation in geostatistical models is built upon the distances between the sampling points, usually using a decay function on the distance. For example, a simple covariance function is defined such as $\Sigma_{ij} = \sigma^2 \exp(-d_{ij}/\varphi)$. Here d_{ij} is the distance between points i and j , and φ is a parameter to control for the spatial scale. Once the model is fitted, prediction relies on the posterior distributions of the parameters and the covariances for the points in the grid.

A more general class of spatial covariance is provided by the Matérn correlation function, of which the exponential decay function is a particular example. The Matérn covariance is defined as

$$\Sigma_{ij} = \sigma^2 \frac{\tau^\kappa K(\tau, \kappa)}{2^{\kappa-1} \Gamma(\kappa)}; \quad \tau = \alpha_\kappa d_{ij} / \varphi \quad (20)$$

$K(\cdot, \kappa)$ is the modified Bessel function of order κ and $\Gamma(\cdot)$ the Gamma function. α_κ and φ can be used to control the scale of the spatial variation. Setting κ to 0.5 leads to an exponential covariance. Other values of κ will lead to other known spatial covariance functions.

When it comes to provide a prediction on the grid, INLA treats the observation at each point on the grid as a missing value. This makes INLA compute the marginal posterior distribution at that point so that summary statistics can be obtained later. In this approach, modeling and prediction occur on a regular grid, and observations need to match to some location in the grid.

Lindgren et al (2011) aim at modeling the geostatistical model by using a mesh based on a triangulation of the sampling points (instead of a regular grid) and a stochastic partial differential equation (SPDE). This is equivalent to modeling a Gaussian process with a Matérn covariance, as in Equation (20). In this approach, the spatial effect \mathbf{u} is

$$u(s) = \sum_{k=1}^N \psi_k(s) w_k, \quad s \in \mathbb{R}^2 \quad (21)$$

where $\{\psi_k\}$ are some basis functions, $\{w_k\}$ are Gaussian distributed weights and N is the number of points in the triangulation used to split the study area. As this is a more complex approach, the reader is referred to the original paper (Lindgren et al 2011).

Finally, INLA can be used for geostatistical design. Methods and results discussed in Diggle et al (2010) for preferential sampling can be reproduced with INLA (see the Case Studies section in <http://www.r-inla.org>). Anisotropic models could also be employed and use of these models is being integrated into the software package.

Point process models

Rue et al (2009) show an example of the analysis of a point pattern with INLA using a Poisson process. Rather than modeling the continuous intensity of the point process,

they divide the study area in N disjoint cells (not necessarily of equal size) and model the data as coming from a counting process. Hence, the response variable y_i represents the number of occurrences of the process in square w_i ; $i = 1, \dots, N$. For simplicity a square lattice may be employed. In a square lattice all the squares have the same area and spatially correlated random effects can be defined similarly as in lattice data (i.e., two squares are neighbors if they have a common boundary).

In their example, Rue et al (2009) use a hierarchical Poisson process to model the number of trees in each square using a log-Gaussian Cox process (LGP). In this case, the intensity function is $\lambda(s) = \exp\{Z(s)\}$; $s \in \mathbb{R}^2$, where $Z(s)$ is a Gaussian field at $s \in \mathbb{R}^2$.

Hence, y_i is the observed number of occurrences in cell w_i . If η_i is the realization of $Z(s_i)$ then $\pi(\mathbf{y}|\boldsymbol{\eta}) = \prod_i \pi(y_i|\eta_i)$, where $\pi(y_i|\eta_i)$ represents a Poisson distribution with mean $|w_i| \exp(\eta_i)$. $|w_i|$ is the area of cell w_i .

In turn, η_i is modeled according to a number of covariates plus some random effects:

$$\eta_i = X_i\beta + u_i + v_i \tag{22}$$

\mathbf{u} and \mathbf{v} are modeled in a similar way as with the lattice data case. v_i are independent Gaussian with zero mean and variance σ_v^2 so that they represent independent variation between the squares. On the other hand, u_i are modeled using a second-order polynomial intrinsic GMRF. In this way, first, second and third order neighbors are taken into account, each one with a different weight, to mimic thin plate splines. See Rue and Held (2005) for details.

Simpson et al (2016) extend the ideas in Lindgren et al (2011) to model the latent LGP in a continuous way using a mesh on the study area. They show that this is a better approach that reduces the computational burden as a mesh is used instead of a regular grid and there is no need to aggregate cases into small cells.

More complex models cannot be fully addressed using INLA. In particular, those for which a closed likelihood does not exist as, for example, Gibbs processes. In a Gibbs process future observations depend on present observations and, hence, producing a likelihood in closed form is not feasible.

Examples

As it happens, INLA is one of many alternatives for fitting Bayesian hierarchical models. In this section we provide a comparison to other software available for the R programming language, including computing times. Our aim here is not to provide a full comparison of computation times but to indicate how different approaches compare in terms of time and accuracy of results when used to fit a similar model to the same data set.

Geostatistics

For geostatistical models, we will use the Rongelap data set analyzed in several works on model-based geostatistics. This data set records radionuclide concentration at 157 different locations, and the interest is on providing an estimate of the concentration over the whole Rongelap island.

In this example computations are made on a regular grid, and we have considered a 5x5 regular grid on one of the clusters in the north-east part of the island to make a fair comparison between computing times. We have used the INLA software (using the Laplace approximation) and the R package `geoRglm`, which provides model fitting using MCMC. The different computation times are shown in Table 1, while a map comparing the different estimates is shown in Fig. 1.

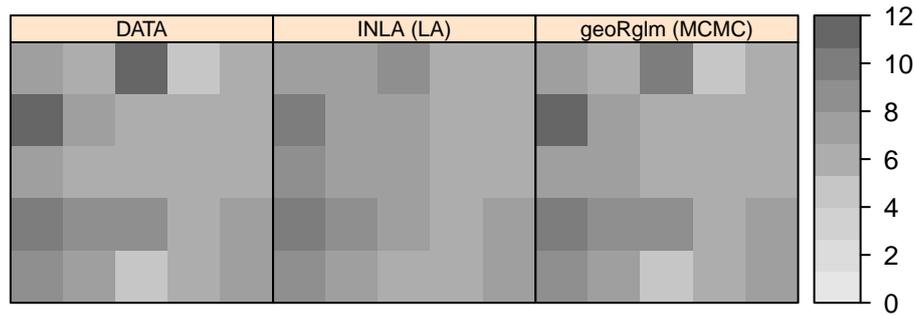


Fig. 1. Estimates of the radionuclide concentration using different methods: integrated Laplace approximation (INLA) and MCMC (using geoRglm).

Lattice data

For the case of lattice data, we have used the number of total malignant neoplasms deaths in South Carolina in 1999. We have fitted the model proposed in Besag et al (1991) with population density as a covariate. In this case, we have used the INLA software as well as WinBUGS. Times are available in Table 1 and a graphical comparison of the estimates is available in Fig. 2.

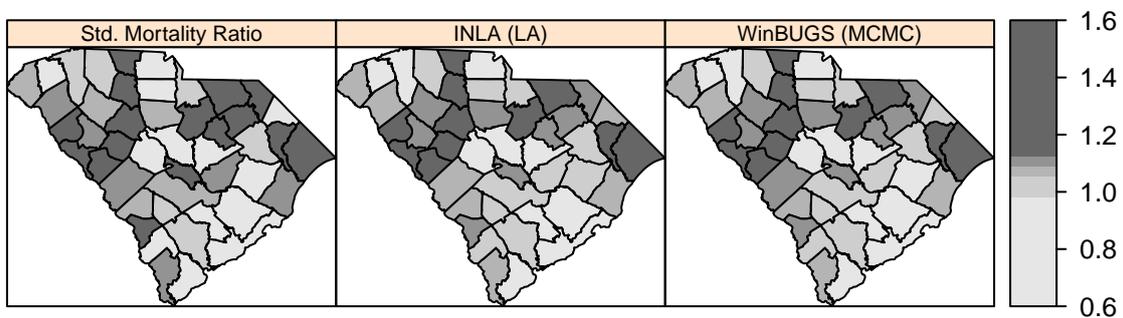


Fig. 2. Estimates of the relative risk using different methods: Standardized Mortality Ratio (SMR), integrated Laplace approximation (INLA) and MCMC (using WinBUGS).

Point patterns

As an example on point patterns, we have performed an analysis of the Japanese pines data set available in R package `spatstat` (Baddeley et al 2015). This data set provides the location of Japanese pine saplings in a square region in a natural forest. As in the geostatistics example, we have used a regular square grid so that the data are the number of saplings in each grid square. A 10x10 square grid has been used in this case and the model to account for spatial dependence is the same as in the previous example (Besag et al 1991). This will also give us an idea of how INLA behaves as the grid size increases.

Figure 3 summarizes the fitted number of saplings and computing times are available in Table 1. It is worth noting how the differences between INLA and WinBUGS have increased now.

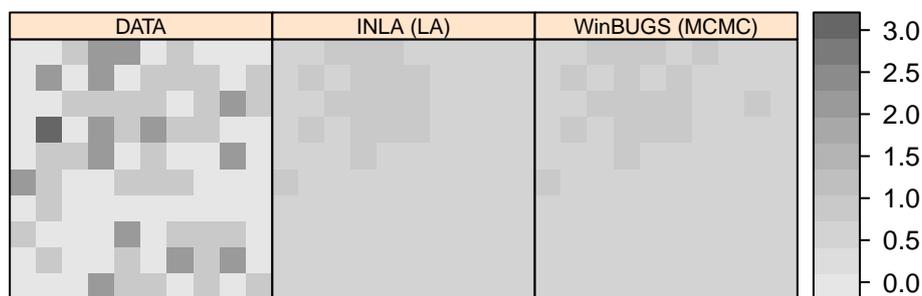


Fig. 3. Estimates of the number of saplings per square using two different methods: integrated Laplace approximation (INLA) and MCMC (using WinBUGS).

Conclusions

The integrated nested Laplace approximation developed in Rue et al (2009) provides a series of approximations for the posterior marginals of the parameters of a Bayesian hierarchical model in which the latent effects are a Gaussian Markov random field.

		Geostatistics		Lattice data		Point patterns	
Software	Method	# Iter.	Time (sec.)	# Iter.	Time (sec.)	# Iter.	Time (sec.)
R-INLA	INLA	—	0.251	—	0.422	—	0.758
geoRglm	MCMC	22,000	0.409	—	—	—	—
WinBUGS	MCMC	—	—	22,000	11.420	22,000	35.336

Table 1. Summary of computation times for different problems, software and fitting method.

This family of models covers a good number of Bayesian hierarchical models, including several of those most used in spatial statistics. In addition, Markov properties are very convenient in dealing with spatial data and they can be used to model local dependence. Besides an approximation to the posterior marginals of the parameters in the model, INLA can compute several criteria for model criticism and selection, such as PIT and the DIC.

Regarding spatial models, INLA has been used to tackle problems in the analysis of lattice data, geostatistics and point processes. Note that several authors have proposed strategies to use INLA to fit spatial models not implemented in the R-INLA package. In all cases, spatial dependence is modeled via the precision matrix of Gaussian random effects. The recent developments by Lindgren et al (2011) allow for continuous modeling of latent spatial effects, which avoids the use of a grid and provides a good computational approach as well. Bakka et al (2018) provides a summary of recent developments of INLA for spatial analysis.

The availability of associated software that implements all these methods provides a suitable framework for their wider use. Other external software may be required to display the results in maps or create adjacency matrices for the analysis of lattice data. For this reason, the authors of the INLA software have provided an interface to the R programming language. The R-INLA web site (<http://www.r-inla.org>) pro-

vides the latest version of the software and its documentation as well as an updated list of published and working papers.

Acknowledgements

V. Gómez-Rubio has been supported by grants PPIC-2014-001-P and SBPLY/17/180501/000491, funded by Consejería de Educación, Cultura y Deportes (JCCM, Spain) and Fondo Europeo de Desarrollo Regional, and grants MTM2008-03085 and MTM2016-77501-P, funded by the Ministerio de Economía y Competitividad (Spain).

References

- Baddeley A, Rubak E, Turner R (2015) *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC Press, London, URL <http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/>
- Bakka H, Rue H, Fuglstad GA, Riebler A, Bolin D, Krainski E, Simpson D, Lindgren F (2018) Spatial modelling with R-INLA: A review. *WIREs Comput Stat* 10(6):1–24, URL <https://doi.org/10.1002/wics.1443>
- Banerjee S, Gelfand AE, Carlin BP (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Boca Raton, Florida
- Besag J, York J, Mollie A (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43(1):1–59
- Diggle PJ, Menezes R, Su Tl (2010) Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59(2):191–232
- Fahrmeir L, Kneib T (2011) *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxford University Press, New York
- Geisser S (1993) *Predictive Inference: An Introduction*. Chapman & Hall, New York
- Gelman A, Carlin JB, Stern HS, Rubin DB (2003) *Bayesian Data Analysis*, 2nd edn. Chapman & Hall/CRC, Boca Raton, Florida

- Kammann EE, Wand MP (2003) Geoaddivitive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52(1):1–18
- Lang S, Brezger A (2004) Bayesian P-splines. *Journal of Computational and Graphical Statistics* 13(1):183–212
- Lee DJ, Durbán M (2009) Smooth-car mixed models for spatial count data. *Computational Statistics and Data Analysis* 53(8):2968 – 2979
- Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion). *Journal of the Royal Statistical Society, Series B* 73(4):423–498
- Martino S, Rue H (2010) Case studies in Bayesian computation using INLA. In: Mantovan, P and Secchi, P (ed) *Complex Data Modeling and Computationally Intensive Statistical Methods, Contributions to Statistics*, Springer, pp 99–114
- Pettit LI (1990) The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society Series B (Methodological)* 52(1):pp. 175–184
- R Development Core Team (2019) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>
- Roos M, Held L (2011) Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis* 6(2):259–278
- Rue H, Held L (2005) *Gaussian Markov Random Fields. Theory and Applications*. Chapman & Hall, New York
- Rue H, Riebler AI, Sørbye SH, Illian JB, Simpson DP, Lindgren FK (2017) Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application* 4:395–421
- Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B* 71(Part 2):319–392
- Simpson D, Illian J, Lindgren F, Sørbye SH, Rue H (2016) Going off grid: Computationally efficient inference for log-gaussian cox processes. *Biometrika* 103(1):49–70

Index

- Bayesian methods, 3
- CAR, *see* conditionally autoregressive model
- conditional predictive ordinate, 11
- conditionally autoregressive model, 14
- Cox process, 18
- CPO, *see* conditional predictive ordinate
- deviance information criterion, 11
- DIC, *see* deviance information criterion
- disease mapping, 14
- Gaussian Markov random field, 8
- geostatistics, 16
- GMRF, *see* Gaussian Markov random field
- INLA, *see* integrated nested Laplace approximation
- integrated nested Laplace approximation, 3, 5
- Laplace approximation, 3, 5
- Matérn correlation, 16
- P-splines, 13
- penalized splines, 13
- PIT, *see* predictive integral transform
- point pattern, 17
- point process, 17
- predictive integral transform, 11
- prior distribution, 9
- R programming language, 12
- R-INLA, 12
- spatial models, 3
- SPDE, *see* stochastic partial differential equations
- splines, 13
- stochastic partial differential equations, 17
- WAIC, *see* Watanabe-Akaike information criterion
- Watanabe-Akaike information criterion, 11