

# Markov Chain Monte Carlo with the Integrated Nested Laplace Approximation

Virgilio Gómez-Rubio<sup>\*1</sup> and Håvard Rue<sup>†2</sup>

<sup>1</sup> Department of Mathematics, School of Industrial Engineering, Universidad de Castilla-La Mancha, Avda España s/n, 02071 Albacete, Spain

<sup>2</sup> CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

April 6, 2017

## Abstract

The Integrated Nested Laplace Approximation (INLA) has established itself as a widely used method for approximate inference on Bayesian hierarchical models which can be represented as a latent Gaussian model (LGM). INLA is based on producing an accurate approximation to the posterior marginal distributions of the parameters in the model and some other quantities of interest by using repeated approximations to intermediate distributions and integrals that appear in the computation of the posterior marginals.

INLA focuses on models whose latent effects are a Gaussian Markov random field (GMRF). For this reason, we have explored alternative ways of expanding the number of possible models that can be fitted using the INLA methodology. In this paper, we present a novel approach that combines INLA and Markov chain Monte Carlo (MCMC). The aim is to consider a wider range of models that cannot be fitted with INLA unless some of the parameters of the model have been fixed. Hence, conditioning on these parameters the model could be fitted with the **R-INLA** package. We show how new values of these

---

\*Virgilio.Gomez@uclm.es

†haavard.rue@kaust.edu.sa

parameters can be drawn from their posterior by using conditional models fitted with INLA and standard MCMC algorithms, such as Metropolis-Hastings. Hence, this will extend the use of INLA to fit models that can be expressed as a conditional LGM. Also, this new approach can be used to build simpler MCMC samplers for complex models as it allows sampling only on a limited number parameters in the model.

We will demonstrate how our approach can extend the class of models that could benefit from INLA, and how the **R-INLA** package will ease its implementation. We will go through simple examples of this new approach before we discuss more advanced problems with datasets taken from relevant literature.

**Keywords:** Bayesian Lasso, INLA, MCMC, Missing Values, Spatial Models

## 1 Introduction

Bayesian inference for complex hierarchical models has almost entirely relied upon computational methods, such as Markov chain Monte Carlo (MCMC, Gilks et al., 1996). Rue et al. (2009) propose a new paradigm for Bayesian inference on hierarchical models that can be represented as latent Gaussian models (LGMs), that focuses on approximating marginal distributions for the parameters in the model. This new approach, the Integrated Nested Laplace Approximation (INLA, henceforth), uses several approximations to the conditional distributions that appear in the integrals needed to obtain the marginal distributions. See Section 2 for details.

INLA is implemented as an R package, called **R-INLA**, that allows us to fit complex models often in a matter of seconds. Hence, this is often much faster than fitting the same model using MCMC methods. Fitting models using INLA is restricted, in practice, to the classes of models implemented in the **R-INLA** package. Several authors have provided ways of fitting other models with INLA by fixing some of the parameters in the model so that conditional models are fitted with **R-INLA**. We have included a brief summary below.

Li et al. (2012) provide an early application of the idea of fitting conditional models on some of the model parameters with **R-INLA**. They developed this idea for a very specific example on spatiotemporal models in which some of the models parameters are fixed at their maximum likelihood estimates, which are then plugged-in the overall model, thus ignoring the uncertainty about these parameters but greatly reducing the dimensionality of the model. However, they do not tackle the problem of fitting the complete

model to make inference on all the parameters in the model.

Bivand et al. (2014, 2015) propose an approach to extend the type of models that can be fitted with **R-INLA** and apply their ideas to fit some spatial models. They note how some models can be fitted after conditioning on one or several parameters in the model. For each of these conditional models **R-INLA** reports the marginal likelihood, which can be combined with a set of priors for the parameters to obtain their posterior distribution. For the remainder of the parameters, their posterior marginal distribution can be obtained by Bayesian model averaging (Hoeting et al., 1999) the family of models obtained with **R-INLA**.

Although Bivand et al. (2014, 2015) focus on some spatial models, their ideas can be applied in many other examples. They apply this to estimate the posterior marginal of the spatial autocorrelation parameter in some models, and this parameter is known to be bounded, so that computation of its marginal distribution is easy because the support of the distribution is a bounded interval.

For the case of unbounded parameters, the previous approach can be applied, but a previous search may be required. For example, the (conditional) maximum log-likelihood plus the log-prior could be maximised to obtain the mode of the posterior marginal. This will mark the centre of an interval where the values for the parameter are taken from and where the posterior marginal can be evaluated.

In this paper, we will propose a different approach based on Markov chain Monte Carlo techniques. Instead of trying to obtain the posterior marginal of the parameters we condition on, we show how to draw samples from their posterior distribution by combining MCMC techniques and conditioned models fitted with **R-INLA**. This provides several advantages, as described below.

This will increase the number of models that can be fitted using INLA and its associated R package **R-INLA**. In particular, models that can be expressed as a conditional LGM could be fitted. The implementation of MCMC algorithms will also be simplified as only the important parameters will be sampled, while the remaining parameters are integrated out with INLA and **R-INLA**. Hubin and Storvik (2016a) have also effectively combined MCMC and INLA for efficient variable selection and model choice.

The paper is structured as follows. The Integrated Nested Laplace Approximation is described in Section 2. Markov chain Monte Carlo methods are summarised in Section 3. Our proposed combination of MCMC and INLA is detailed in Section 4. Some simple examples are developed in Section 5 and some real applications are provided in Section 6. Finally, a discussion and some final remarks are provided in Section 7.

## 2 Integrated Nested Laplace Approximation

We will now describe the types of models that we will be considering and how the Integrated Nested Laplace Approximation method works. We will assume that our vector of  $n$  observed data  $\mathbf{y} = (y_1, \dots, y_n)$  are observations from a distribution in the exponential family, with mean  $\mu_i$ . We will also assume that a linear predictor on some covariates plus, possibly, other effects can be related to mean  $\mu_i$  by using an appropriate link function. Note that this linear predictor  $\eta_i$  may be made of linear terms on some covariates plus other types of terms, such as non-linear functions on the covariates, random effects, spatial random effects, etc. All these terms will define some latent effects  $\mathbf{x}$ .

The distribution of  $\mathbf{y}$  will depend on a vector of hyperparameters  $\boldsymbol{\theta}_1$ . Because of the approximation that INLA will use, we will also assume that the vector of latent effects  $\mathbf{x}$  will have a distribution that will depend on a vector of hyperparameters  $\boldsymbol{\theta}_2$ . Altogether, the hyperparameters can be represented using a single vector  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ .

From the previous formulation, it is clear that observations are independent given the values of the latent effects  $\mathbf{x}$  and the hyperparameters  $\boldsymbol{\theta}$ . That is, the likelihood of our model can be written down as

$$\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i \in \mathcal{I}} \pi(y_i|x_i, \boldsymbol{\theta}) \quad (1)$$

Here,  $i$  is indexed over a set of indices  $\mathcal{I} \subseteq \{1, \dots, n\}$  that indicates observed responses. Hence, if the value of  $y_i$  is missing then  $i \notin \mathcal{I}$  (but the predictive distribution  $y_i$  could be computed once the model is fitted).

Under a Bayesian framework, the aim is to compute the posterior distribution of the model parameters and hyperparameters using Bayes' rule. This can be stated as

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}, \boldsymbol{\theta}) \quad (2)$$

Here,  $\pi(\mathbf{x}, \boldsymbol{\theta})$  is the prior distribution of the latent effects and the vector of hyperparameters. As the latent effects  $\mathbf{x}$  have a distribution that depends on  $\boldsymbol{\theta}_2$ , it is convenient to write this prior distribution as  $\pi(\mathbf{x}, \boldsymbol{\theta}) = \pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ .

Altogether, the posterior distribution of the latent effects and hyperparameters can be expressed as

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

The joint posterior, as presented in Equation (3), is seldom available in a closed form. For this reason, several estimation methods and approximations have been developed over the years.

Recently, Rue et al. (2009) have provided approximations based on the Laplace approximation to estimate the marginals of all parameters and hyperparameters in the model. They develop this approximation for the family of latent Gaussian Markov random fields models. In this case, the vector of latent effects is a Gaussian Markov random field (GMRF). This GMRF will have zero mean (for simplicity) and precision matrix  $Q(\boldsymbol{\theta}_2)$ .

Assuming that the latent effects are a GMRF will let us develop Equation (3) further. In particular, the posterior distribution of the latent effects  $\mathbf{x}$  and the vector of hyperparameters  $\boldsymbol{\theta}$  can be written as

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

With INLA, the aim is not the joint posterior distribution  $\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})$  but the marginal distributions of latent effects and hyperparameters. That is,  $\pi(x_j | \mathbf{y})$  and  $\pi(\theta_k | \mathbf{y})$ , where indices  $j$  and  $k$  will take different ranges of values depending on the number of latent effects and hyperparameters.

Before computing these marginal distributions, INLA will obtain an approximation to  $\pi(\boldsymbol{\theta} | \mathbf{y})$ ,  $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$ . This approximation will later be used to compute an approximation to marginals  $\pi(x_j | \mathbf{y})$ . Given that the marginal can be written down as

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

the approximation is as follows:

$$\tilde{\pi}(x_j | \mathbf{y}) = \sum_g \tilde{\pi}(x_j | \boldsymbol{\theta}_g, \mathbf{y}) \times \tilde{\pi}(\boldsymbol{\theta}_g | \mathbf{y}) \times \Delta_g. \quad (6)$$

Here,  $\tilde{\pi}(x_j | \boldsymbol{\theta}_g, \mathbf{y})$  is an approximation to  $\pi(x_j | \boldsymbol{\theta}_g, \mathbf{y})$ , which can be obtained using different methods (see, Rue et al., 2009, for details).  $\boldsymbol{\theta}_g$  refers to an ensemble of hyperparameters, that take values on a grid (for example), with weights  $\Delta_g$ .

INLA is a general approximation that can be applied to a large number of models. An implementation for the R programming language is available in the **R-INLA** package at [www.r-inla.org](http://www.r-inla.org), which provides simple access to model fitting. This includes a simple interface to choose the likelihood, latent effects and priors. The implementation provided by **R-INLA** includes the computation of other quantities of interest. The marginal likelihood  $\pi(\mathbf{y})$

is approximated, and it can be used for model choice. As described in Rue et al. (2009), the approximation to the marginal likelihood provided by INLA is computed as

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

Here,  $\pi(\boldsymbol{\theta}, \mathbf{x}, \mathbf{y}) = \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ ,  $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$  is a Gaussian approximation to  $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$  and  $\mathbf{x}^*(\boldsymbol{\theta})$  is the posterior mode of  $\mathbf{x}$  for a given value of  $\boldsymbol{\theta}$ . This approximation is reliable when the posterior of  $\boldsymbol{\theta}$  is unimodal, as it is often the case for latent Gaussian models. Furthermore, Hubin and Storvik (2016b) demonstrate that this approximation is accurate for a wide range of models.

Other options for model choice and assessment include the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002) and the Conditional Predictive Ordinate (CPO, Pettit, 1990). Other features in the **R-INLA** package include the use of different likelihoods in the same model, the computation of the posterior marginal of a certain linear combination of the latent effects and others (see, Martins et al., 2013, for a summary of recent additions to the software).

### 3 Markov Chain Monte Carlo

In the previous Section we have reviewed how INLA computes an approximation of the marginal distributions of the model parameters and hyperparameters. Instead of focusing on an approximation to the marginals, Markov chain Monte Carlo methods could be used to obtain a sample from the joint posterior marginal  $\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$ . To simplify the notation, we will denote the vector of latent effects and hyperparameters by  $\mathbf{z} = (\mathbf{x}, \boldsymbol{\theta})$ . Hence, the aim now is to estimate  $\pi(\mathbf{z}|\mathbf{y})$  or, if we are only interested on the posterior marginals,  $\pi(z_i|\mathbf{y})$ .

Several methods to estimate or approximate the posterior distribution have been developed over the years (Gilks et al., 1996). In the case of MCMC, the interest is in obtaining a Markov chain whose limiting distribution is  $\pi(\mathbf{z}|\mathbf{y})$ . We will not provide a summary of MCMC methods here, and the reader is referred to Gilks et al. (1996) for a detailed description.

The values generated using MCMC are (correlated) draws from  $\pi(\mathbf{z}|\mathbf{y})$  and, hence, can be used to estimate quantities of interest. For example, if we are interested in marginal inference on  $z_i$ , the posterior mean from the  $N$  sampled values  $\left\{z_i^{(j)}\right\}_{j=1}^N$  can be estimated using the empirical mean of

$\{z_i^{(j)}\}_{j=1}^N$ . Similarly, estimates of the posterior expected value of any function on the parameters  $f(\mathbf{z})$  can be found using that

$$E[f(\mathbf{z})|\mathbf{y}] \simeq \frac{1}{N} \sum_{j=1}^N f(\mathbf{z}^{(j)}) \quad (7)$$

Multivariate estimates inference can be made by using the multivariate nature of vector  $\mathbf{z}^{(j)}$ . For example, the posterior covariance between parameters  $z_k$  and  $z_l$  could be computed by considering samples  $\{(z_k^{(j)}, z_l^{(j)})\}_{j=1}^N$ .

### 3.1 The Metropolis-Hastings algorithm

This algorithm was firstly proposed by Metropolis et al. (1953) and Hastings (1970). The Markov chain is generated by proposing new moves according to a proposal distribution  $q(\cdot|\cdot)$ . The new point is accepted with probability

$$\alpha = \min \left\{ 1, \frac{\pi(\mathbf{z}^{(j+1)}|\mathbf{y})q(\mathbf{z}^{(j)}|\mathbf{z}^{(j+1)})}{\pi(\mathbf{z}^{(j)}|\mathbf{y})q(\mathbf{z}^{(j+1)}|\mathbf{z}^{(j)})} \right\} \quad (8)$$

In the previous acceptance probability, the posterior probabilities of the current point and the proposed new point appear as  $\pi(\mathbf{z}^{(j)}|\mathbf{y})$  and  $\pi(\mathbf{z}^{(j+1)}|\mathbf{y})$ , respectively. These two probabilities are unknown, in principle, but using Bayes' rule they can be rewritten as

$$\pi(\mathbf{z}|\mathbf{y}) = \frac{\pi(\mathbf{y}|\mathbf{z})\pi(\mathbf{z})}{\pi(\mathbf{y})} \quad (9)$$

Hence, the acceptance probability  $\alpha$  can be rewritten as

$$\alpha = \min \left\{ 1, \frac{\pi(\mathbf{y}|\mathbf{z}^{(j+1)})\pi(\mathbf{z}^{(j+1)})q(\mathbf{z}^{(j)}|\mathbf{z}^{(j+1)})}{\pi(\mathbf{y}|\mathbf{z}^{(j)})\pi(\mathbf{z}^{(j)})q(\mathbf{z}^{(j+1)}|\mathbf{z}^{(j)})} \right\} \quad (10)$$

This is easier to compute as the acceptance probability depends on known quantities, such as the likelihood  $\pi(\mathbf{y}|\mathbf{z})$ , the prior on the parameters  $\pi(\mathbf{z})$  and the probabilities of the proposal distribution. Note that the term  $\pi(\mathbf{y})$  that appears in Equation (9) is unknown but that it cancels out as it appears both in the numerator and denominator.

In Equation (10) we have described the move to sample from the joint ensemble of model parameters. However, this can be applied to individual parameters one at a time, so that acceptance probabilities will be

$$\alpha = \min \left\{ 1, \frac{\pi(\mathbf{y}|z_i^{(j+1)})\pi(z_i^{(j+1)})q(z_i^{(j)}|z_i^{(j+1)})}{\pi(\mathbf{y}|z_i^{(j)})\pi(z_i^{(j)})q(z_i^{(j+1)}|z_i^{(j)})} \right\} \quad (11)$$

## 4 INLA within MCMC

In this Section, we will describe how INLA and MCMC can be combined to fit complex Bayesian hierarchical models. In principle, we will assume that the model cannot be fitted with **R-INLA** unless some of the parameters or hyperparameters in the model are fixed. This set of parameters is denoted by  $\mathbf{z}_c$  so that the full ensemble of parameters and hyperparameters is  $\mathbf{z} = (\mathbf{z}_c, \mathbf{z}_{-c})$ . Here  $\mathbf{z}_{-c}$  is used to denote all the parameters in  $\mathbf{z}$  that are not in  $\mathbf{z}_c$ . Our assumptions are that the posterior distribution of  $\mathbf{z}$  can be split as

$$\pi(\mathbf{z}|\mathbf{y}) \propto \pi(\mathbf{y}|\mathbf{z}_{-c})\pi(\mathbf{z}_{-c}|\mathbf{z}_c)\pi(\mathbf{z}_c) \quad (12)$$

and that  $\pi(\mathbf{y}|\mathbf{z}_{-c})\pi(\mathbf{z}_{-c}|\mathbf{z}_c)$  is a latent Gaussian model suitable for INLA. This means that conditional models (on  $\mathbf{z}_c$ ) can still be fitted with **R-INLA**, i.e., we can obtain marginals of the parameters in  $\mathbf{z}_{-c}$  given  $\mathbf{z}_c$ . The conditional posterior marginals for the  $k$ -th element in vector  $\mathbf{z}_{-c}$  will be denoted by  $\pi(z_{-c,k}|\mathbf{z}_c, \mathbf{y})$ . Also, the conditional marginal likelihood  $\pi(\mathbf{y}|\mathbf{z}_c)$  can be easily computed with **R-INLA**.

### 4.1 Metropolis-Hastings with INLA

We will now discuss how to implement the Metropolis-Hastings algorithm to estimate the posterior marginal of  $\mathbf{z}_c$ . Note that this is a multivariate distribution and that we will use block updating in the Metropolis-Hastings algorithm. Say that we start from an initial point  $\mathbf{z}_c^{(0)}$  then we can use the Metropolis-Hastings algorithm to obtain a sample from the posterior of  $\mathbf{z}_c$ .

We will draw a new proposal value for  $\mathbf{z}_c$ ,  $\mathbf{z}_c^{(1)}$ , using the proposal distribution  $q(\cdot|\cdot)$ . The acceptance probability, shown in Equation (10), becomes now:

$$\alpha = \min \left\{ 1, \frac{\pi(\mathbf{y}|\mathbf{z}_c^{(j+1)})\pi(\mathbf{z}_c^{(j+1)})q(\mathbf{z}_c^{(j)}|\mathbf{z}_c^{(j+1)})}{\pi(\mathbf{y}|\mathbf{z}_c^{(j)})\pi(\mathbf{z}_c^{(j)})q(\mathbf{z}_c^{(j+1)}|\mathbf{z}_c^{(j)})} \right\} \quad (13)$$

Note that  $\pi(\mathbf{y}|\mathbf{z}_c^{(j)})$  and  $\pi(\mathbf{y}|\mathbf{z}_c^{(j+1)})$  are the conditional marginal likelihoods on  $\mathbf{z}_c^{(j)}$  and  $\mathbf{z}_c^{(j+1)}$ , respectively. All these quantities can be obtained by fitting a model with **R-INLA** with the values of  $\mathbf{z}_c$  set to  $\mathbf{z}_c^{(j)}$  and  $\mathbf{z}_c^{(j+1)}$ . Hence, at each step of the Metropolis-Hastings algorithm only a model conditional on the proposal needs to be fitted.

$\pi(\mathbf{z}_c^{(j)})$  and  $\pi(\mathbf{z}_c^{(j+1)})$  are the priors of  $\mathbf{z}_c$  evaluated at  $\mathbf{z}_c^{(j)}$  and  $\mathbf{z}_c^{(j+1)}$ , respectively, and they can be easily computed as the priors are known in the model. Values  $q(\mathbf{z}_c^{(j)}|\mathbf{z}_c^{(j+1)})$  and  $q(\mathbf{z}_c^{(j+1)}|\mathbf{z}_c^{(j)})$  can also be computed as the proposal distribution is known. Hence, the Metropolis-Hastings algorithm can be implemented to obtain a sample from the posterior distribution of



$\mathbf{z}_c$ . The marginal distributions of the elements of  $\mathbf{z}_c$  can be easily obtained as well.

Regarding the marginals of  $z_{-c,k}$ , it is worth noting that at step  $j$  of the Metropolis-Hastings algorithm a conditional marginal distribution on  $\mathbf{z}_c^{(j)}$  (and the data  $\mathbf{y}$ ) is obtained:  $\pi(z_{-c,k}|\mathbf{z}_c^{(j)}, \mathbf{y})$ . The posterior marginal can be approximated by integrating over  $\mathbf{z}_c$  as follows:

$$\pi(z_{-c,k}|\mathbf{y}) = \int \pi(z_{-c,k}|\mathbf{z}_c, \mathbf{y})\pi(\mathbf{z}_c|\mathbf{y})d\mathbf{z}_c \simeq \frac{1}{N} \sum_{j=1}^N \pi(z_{-c,k}|\mathbf{z}_c^{(j)}, \mathbf{y}), \quad (14)$$

where  $N$  is the number of samples of the posterior distribution of  $\mathbf{z}_c$ . That is, the posterior marginal of  $z_{-c,k}$  can be obtained by averaging the conditional marginals obtained at each iteration of the Metropolis-Hastings algorithm.

## 4.2 Effect of approximating the marginal likelihood

So far, we have ignored the fact that the conditional marginal likelihood  $\pi(\mathbf{y}|\mathbf{z}_c)$  used in the acceptance probability  $\alpha$  is actually an approximation. In this section, we will discuss how this approximation will impact the validity of the inference.

The situations where a Metropolis-Hastings algorithm has inexact acceptance probabilities are often called pseudo-marginal MCMC algorithms and were first introduced in Beaumont (2003) in the context of statistical genetics where the likelihood in the acceptance probability is approximated using importance sampling. Andrieu and Roberts (2003) provided a more general justification of the pseudo-marginal MCMC algorithm, whose properties are further studied in Sherlock et al. (2015) and Medina-Aguayo et al. (2016). These results show that if the (random) acceptance probability is unbiased then the Markov chain will still have as stationary distribution the posterior distribution of the model parameters.

In our case, the error in the acceptance rate is coming from a deterministic estimate of the conditional marginal likelihood, hence the framework of pseudo-marginal MCMC does not apply. However, since it is deterministic, our MCMC chain will converge to a stationary distribution. This limiting distribution will be

$$\tilde{\pi}(\mathbf{z}_c|\mathbf{y}) \propto \pi(\mathbf{z}_c)\tilde{\pi}(\mathbf{y}|\mathbf{z}_c) \quad (15)$$

where the “ $\sim$ ” indicates an approximation. **R-INLA** returns an approximation to the conditional marginal likelihood term, which implies an approximation to  $\pi(\mathbf{z}_c|\mathbf{y})$ . This leaves the question, about how good this approximation is, for which we have to rely on asymptotic results, heuristics and numerical experience.

The conditional marginal likelihood estimate returned from **R-INLA** is based on numerical integration and uses a sequence of Laplace approximations (Rue et al., 2009, 2017). This estimate is more accurate than the classical estimate using one Laplace approximation. This approximation has, with classical assumptions, relative error  $\mathcal{O}(n^{-1})$  (Tierney and Kadane, 1986), where  $n$  is the number of replications in the observations. For our purpose, this error estimate is sufficient, as it demonstrates that

$$\frac{\tilde{\pi}(\mathbf{z}_c|\mathbf{y})}{\pi(\mathbf{z}_c|\mathbf{y})} \propto \frac{\tilde{\pi}(\mathbf{y}|\mathbf{z}_c)}{\pi(\mathbf{y}|\mathbf{z}_c)} = 1 + \mathcal{O}(n^{-1}) \quad (16)$$

for plausible values of  $\mathbf{z}_c$ . However, as discussed by Rue et al. (2009, 2017), the classical assumptions are rarely met in practice due to “random effects”, smoothing etc. Precise error estimates under realistic assumptions are difficult to obtain; see Rue et al. (2017) for a more detailed discussion of this issue.

About numerical experience with the conditional marginal likelihood estimate, Hubin and Storvik (2016b) have studied empirically its properties and accuracy for a wide range of latent Gaussian models. They have compared the estimates with those obtained using MCMC, and in all their cases the approximates of the marginal likelihood provided by INLA were very accurate. For this reason, we believe that the approximate stationary distribution  $\tilde{\pi}(\mathbf{z}_c|\mathbf{y})$  should be close to the true one, without being able to quantify this error in more details.

Although the error in Equation (16) is pointwise, we do expect the error would be smooth in  $\mathbf{z}_c$ . This is particularly important, as in most cases we are interested in the univariate marginals of  $\tilde{\pi}(\mathbf{z}_c|\mathbf{y})$ . These marginals will typically have less error as the influence of the approximation error will be averaged out integrating out all the other components. A final renormalization would also remove constant offset in the error.

Additionally, we will validate the approximation error in a simulation study in Section 5 where we fit various models using INLA, MCMC and INLA within MCMC and very similar posterior distributions are obtained. Furthermore, the real applications in Section 6 also support that the approximations to the marginal likelihood are accurate.

### 4.3 Some remarks

Common sense is still not out of fashion, hence there is an implicit assumption that our INLA within MCMC approach should be only for models for which it is reasonable to use the INLA-approach to do the inference for the conditional model. The procedure that we have just shown will allow INLA

to be used together with the Metropolis-Hastings algorithm (and, possibly, other MCMC methods) to obtain the posterior distribution (and marginals) of  $\mathbf{z}_c$  and the posterior marginals of the elements in  $\mathbf{z}_{-c}$ . Hence, this will allow INLA to be used to fit models not implemented in the **R-INLA** package as well as providing other options for model fitting, that we summarise here.

The Metropolis-Hastings algorithm will allow any choice of the priors on the set of parameters  $\mathbf{z}_c$ . This is an advantage (as shown in the example in Section 6.1) of combining MCMC and INLA because priors that are not implemented in **R-INLA** can be used in the model. In particular, improper flat priors, multivariate priors and objective priors can be used.

The framework of conditional LGMs that we now can fit using our new approach is quite rich. It includes models with missing covariates that are imputed at each step of the Metropolis-Hastings algorithm (see example in Section 6.2), models with complex non-linear effects in the linear predictor (see example in Section 6.3) or models that have a mixture of effects in the linear predictor (Bivand et al., 2015).

## 5 Simulation study

In this section we develop simple examples to illustrate the method proposed in the previous sections, and we investigate how this new approach works in practice.

### 5.1 Bivariate linear regression

The first example is based on a linear regression with two covariates. Our aim is to use our proposed method to obtain the posterior distribution of the coefficients of the two covariates and then compare the estimated marginals to the results obtained when the full model is fitted with MCMC and INLA.

The simulated dataset contains 100 observations of a response variable  $\mathbf{y}$  and covariates  $\mathbf{u}_1$  and  $\mathbf{u}_2$ . The model used to generate the data is a typical linear regression, i.e.,

$$y_i = \alpha + \beta_1 u_{1i} + \beta_2 u_{2i} + \varepsilon_i; \quad i = 1, \dots, 100 \quad (17)$$

Here,  $\varepsilon_i$  is a Gaussian error term with zero mean and precision  $\tau$ . The dataset has been simulated using  $\alpha = 3$ ,  $\beta_1 = 2$ ,  $\beta_2 = -2$  and  $\tau = 1$ . Covariates  $u_{1i}$  and  $u_{2i}$  have also been simulated using a uniform distribution between 0 and 1 in both cases.

This model can be easily fitted using **R-INLA**. Given that we are using a Gaussian model, inference is exact in this case (up to integration error). For

this reason, we can compare the marginal distribution provided of  $\beta_1$  and  $\beta_2$  by INLA to the ones obtained with our combined approach. Note that the Metropolis-Hastings algorithm will provide the joint posterior distribution of  $\boldsymbol{\beta} = (\beta_1, \beta_2)$  that can be used to obtain the posterior marginals of  $\beta_1$  and  $\beta_2$ . Furthermore, we can also compare the marginals of  $\alpha$  and  $\tau$ , that will be estimated by averaging the different conditional marginals obtained in the Metropolis-Hastings steps.

In order to implement the Metropolis-Hastings algorithm to obtain a sample from  $\pi(\boldsymbol{\beta}|\mathbf{y})$  we have chosen a starting point of  $\boldsymbol{\beta}^{(0)} = (0, 0)$ . The transition kernel to obtain a candidate  $\boldsymbol{\beta}^{(t+1)}$  at iteration  $t$  has been a bivariate Gaussian kernel centered at  $\boldsymbol{\beta}^{(t)}$  with diagonal variance-covariance matrix with values  $1/0.75^2$  in the diagonal as this provided a reasonable acceptance rate. The prior distribution of  $\boldsymbol{\beta}$  has been the product of two Gaussian distributions with zero mean and precision 0.001 because these are the default priors for linear effects in **R-INLA**.

Figure 1 shows a summary of the results. Given that both covariates are independent, their coefficients should show small correlation and this can clearly be seen in the plot of the joint posterior distribution of  $\boldsymbol{\beta}$ . Also, it can be seen how the marginals obtained with INLA within MCMC for  $\beta_1$  and  $\beta_2$  match those obtained with INLA and MCMC. In addition, we have included the estimates of the posterior marginals of the intercept  $\alpha$  and the precision  $\tau$ . When using INLA within MCMC these are obtained by Bayesian model averaging over the fitted models at every step of the Metropolis-Hastings algorithm, whilst when computed with **R-INLA** these are obtained by using INLA alone. The three estimation methods provide very similar posterior distributions of the posterior marginals of the intercept and the precision, which again confirms the accuracy of INLA within MCMC.

## 5.2 Missing covariates

In the next example, we will discuss the case of missing covariates. In this example we will consider a linear regression with covariate  $\mathbf{u}_1$  only and we will assume that a number of values of the covariates are missing. The aim is to include the imputation of these variables into the model, so that the output is a marginal distribution of the missing values. We will not discuss here the different frameworks under which the values have gone missing, but this is something that should take into account in the model. In particular, we have removed the values of 9 covariates, which is almost 10% of our data and summary plots can be included in a 3x3 plot. Hence, in this case the missingness mechanism is of the type missing completely at random (Little and Rubin, 2002).

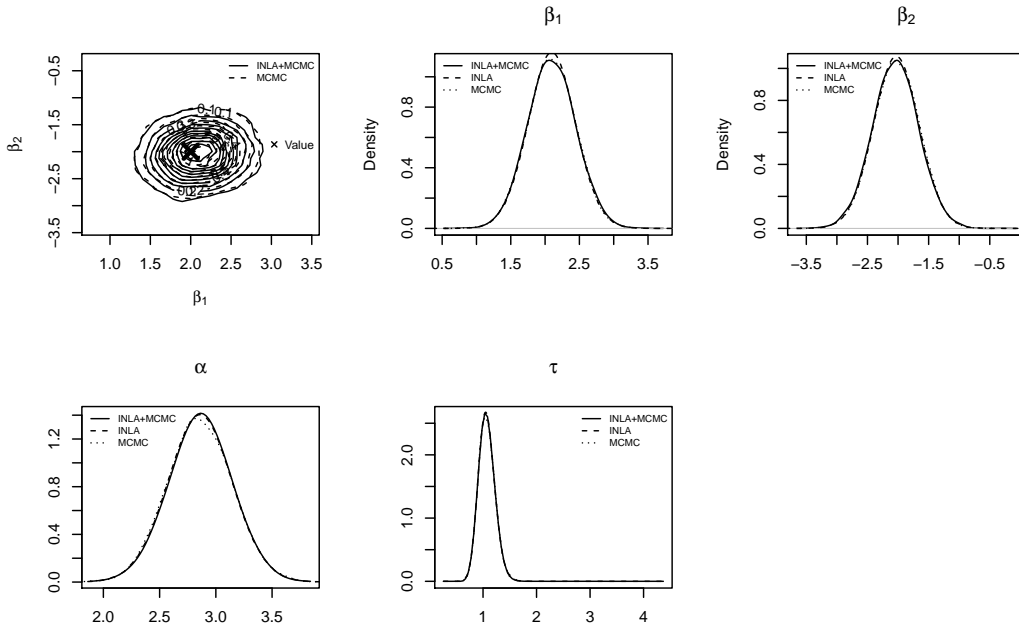


Figure 1: Summary of results of model fitting combining INLA and MCMC in the bivariate case. Joint posterior distribution of  $(\beta_1, \beta_2)$  (top-left plot) and posterior marginals of the model parameters.

Now, we will treat the missing values as if they were covariates. We will use a block updating scheme as we can have a large number of missing covariates. The transition kernel will be a multivariate Gaussian with diagonal variance-covariance. The mean and variance for all values are the mean and variance of the observed covariates, respectively. The prior distribution is also a multivariate Gaussian, but now with zero mean and diagonal variance-covariance matrix with entries four times the variance of a uniform random variable in the unit interval (the one used to simulate the covariates). This is done so that the prior information is small compared to the information provided by the covariates.

Figure 2 shows the posterior marginals obtained from the samples. As it can be seen, most of them are centered at the actual values removed from the model. Note that this time the model with missing covariates cannot be fitted with **R-INLA** so that we can only compare the marginals to those obtained with MCMC. In all cases the marginals obtained with INLA within MCMC and full MCMC are very similar.

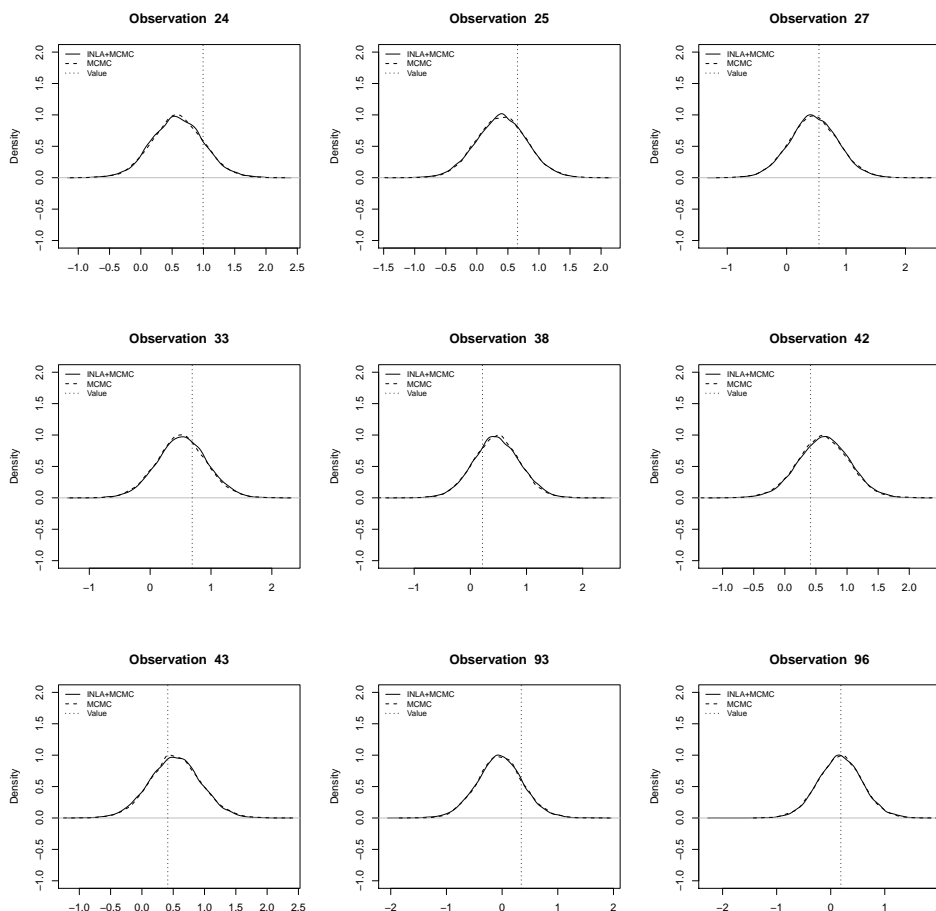


Figure 2: Posterior marginals of the missing values in the covariates obtained by fitting a model with INLA within MCMC, and MCMC.

### 5.3 Poisson regression

In this example we consider a Poisson regression with two covariates:

$$y_i \sim Po(\mu_i); \log(\mu_i) = \alpha + \beta_1 u_{1i} + \beta_2 u_{2i}; \quad i = 1, \dots, 100. \quad (18)$$

The values of the parameters used to simulate the dataset are  $\alpha = 0.5$ ,  $\beta_1 = 2$  and  $\beta_2 = -2$ .

As in Section 5.1, our purpose is to estimate the joint posterior distribution of  $(\beta_1, \beta_2)$ . The prior distribution used now is the same as in the previous example. Hence, the posterior marginal of  $\alpha$  is obtained by combining the different conditional marginals obtained at the different steps of

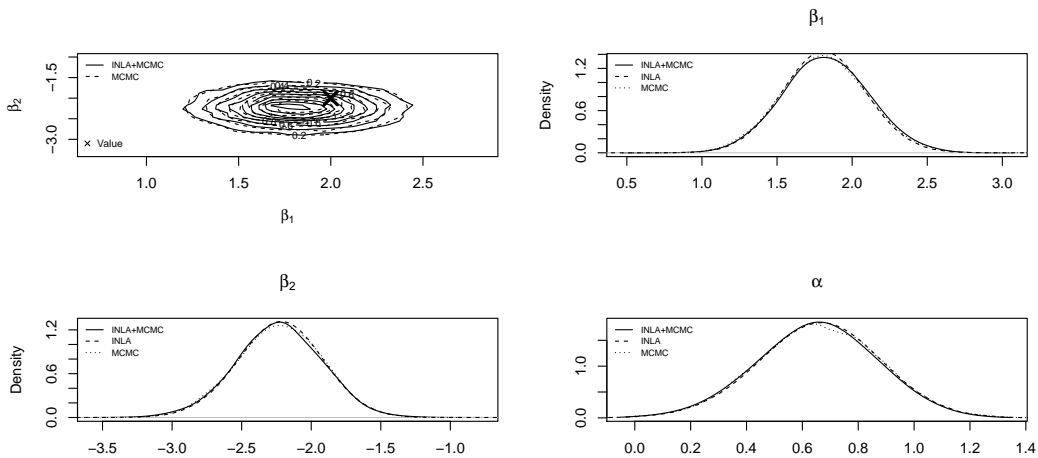


Figure 3: Summary of results of model fitting combining INLA and MCMC for the Poisson regression example. Joint posterior distribution of  $(\beta_1, \beta_2)$  (top-left plot) and posterior marginals of the model parameters.

the Metropolis-Hastings algorithm.

Figure 3 shows the estimates of the marginal distributions of the three parameters in the model, together with the joint posterior distribution of  $\beta_1$  and  $\beta_2$ . In all cases, there is a very good agreement between the estimates obtained with INLA, MCMC and INLA within MCMC of the marginals of the parameters in the model.

## 6 Applications

In this section, we will focus on some real life applications that provide a more realistic test of this methodology. In all the examples, we have run INLA within MC and MCMC for a total of 100500 simulations and discarded the first 500. Then we applied a thinning to keep one in ten iterations, to obtain a final chain of 10000 samples. This includes samples from the missing observations and fitted models. To fit the model using MCMC alone, we have used **rjags** (Plummer, 2016) with the same number of iterations and thinning.

### 6.1 Bayesian Lasso

The Lasso (Tibshirani, 1996) is a popular regression and variable selection method for variable selection. It has the nice property of providing coeffi-

cient estimates that are exactly zero and, hence, it performs model fitting and variable selection at the same time. For a linear model with a Gaussian likelihood, the Lasso is trying to estimate the regression coefficient by minimising

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Here,  $y_i$  is the response variable and  $x_{ij}$  are associated covariates.  $n$  is the number of observations and  $p$  the number of covariates.  $\lambda$  is a non-negative penalty term to control how the shrinkage of the coefficients is done. If  $\lambda = 0$  then the fitted coefficients are those obtained by maximum likelihood, whilst higher values of  $\lambda$  will shrink the estimates towards zero.

The Lasso is closely related to Bayesian inference as it can be regarded as a standard regression model with Laplace priors on the variable coefficients. The Laplace distribution is defined as

$$f(\beta) = \frac{1}{2\sigma} \exp\left(-\frac{|\beta - \mu|}{\sigma}\right), \quad x \in \mathbb{R}$$

where  $\mu$  and  $\sigma$ , a positive number, are parameters of location and scale, respectively. The Laplace prior distribution is not available for (parts of) the latent field in **R-INLA**. However, conditioning on the values of the  $\beta$ -coefficients the model can be easily fitted with **R-INLA**.

We will apply the methodology described in this paper to implement the Bayesian Lasso by combining INLA and MCMC. We will be using the **Hitters** dataset described in James et al. (2013). This dataset records several statistics about players in the Major League Baseball, including salary in 1987, number of times at bat in 1986 and other variables. Our aim is to build a model to predict the player's salary in 1987 on some of the other variables recorded in 1986 (the previous season).

We will focus on a smaller model than the one described in James et al. (2013) and will consider predicting salary in 1987 on only five variables measured from the 1986 season: number of times at bat (AtBat), number of hits (Hits), the number of home runs (HmRun), number of runs (Runs) and the number of runs batted (RBI).

For our implementation of the Bayesian Lasso, we will be fitting models conditioning on the covariate coefficients  $\beta = (\beta_1, \dots, \beta_p)$ . Also, we will assume that  $\beta$  and the error term precision  $\tau$  are independent a priori, i.e.,  $\pi(\beta, \tau) = \pi(\beta)\pi(\tau)$ . This will provide a simpler way to compare our results with the Lasso and it will also make computations a bit simpler. However, note that choosing it is also possible to choose a prior so that  $\pi(\beta, \tau) =$



Coefficient	Lasso	INLA+MCMC	MCMC
AtBat	0.00	-0.01 (0.08)	-0.02 (0.08)
Hits	0.18	0.17 (0.12)	0.17 (0.12)
HmRun	0.00	0.02 (0.07)	0.02 (0.07)
Runs	0.00	0.07 (0.09)	0.07 (0.08)
RBI	0.23	0.21 (0.11)	0.22 (0.11)

Table 1: Summary estimates of the Lasso and Bayesian Lasso (posterior mean and standard deviation, between parenthesis).

$\pi(\boldsymbol{\beta}|\tau)\pi(\tau)$  (see, for example, Lykou and Ntzoufras, 2011) The posterior distribution of these variables will be obtained using MCMC.

The summary of the Lasso estimates are available in Table 1 and the posterior distribution of the coefficients is in Figure 4. In all cases, there is agreement between the Lasso and Bayesian Lasso estimates. Also, the posterior distribution of the model coefficients is the same for MCMC and combining INLA with MCMC. For those coefficients with a zero estimate with the Lasso, the posterior distribution obtained with the Bayesian Lasso is centered at zero.

## 6.2 Imputation of Missing Covariates

van Buuren and Groothuis-Oudshoorn (2011) describe the R package **mice** that implements several multiple imputation methods. We will be using the **nhanes** dataset to illustrate how our approach can be used to provide imputation of missing covariates in a real dataset. This dataset contains data from Schafer (1997) on age, body mass index (**bmi**), hypertension status (**hyp**) and cholesterol level (**ch1**). Age is divided into three groups: 20-39, 40-59, 60+.

Our aim is to impute missing covariates in order to fit a model that explains the cholesterol level on age and body mass index. Although the values of age have been completely observed, there are missing values in body mass index and cholesterol level. INLA can handle missing values in the response (and will provide a predictive distribution) but, as already stated, is not able to handle models with missing values in the covariates.

We will consider a very simple imputation mechanism by assigning a Gaussian prior to the missing values of body mass index. This Gaussian is centred at the average value of the observed values (26.56) and variance four times the variance of the observed values (71.07, altogether). With this, we expect to provide some guidance on how the imputed values should be but allowing for a wide range of variation. More complex imputation mechanisms

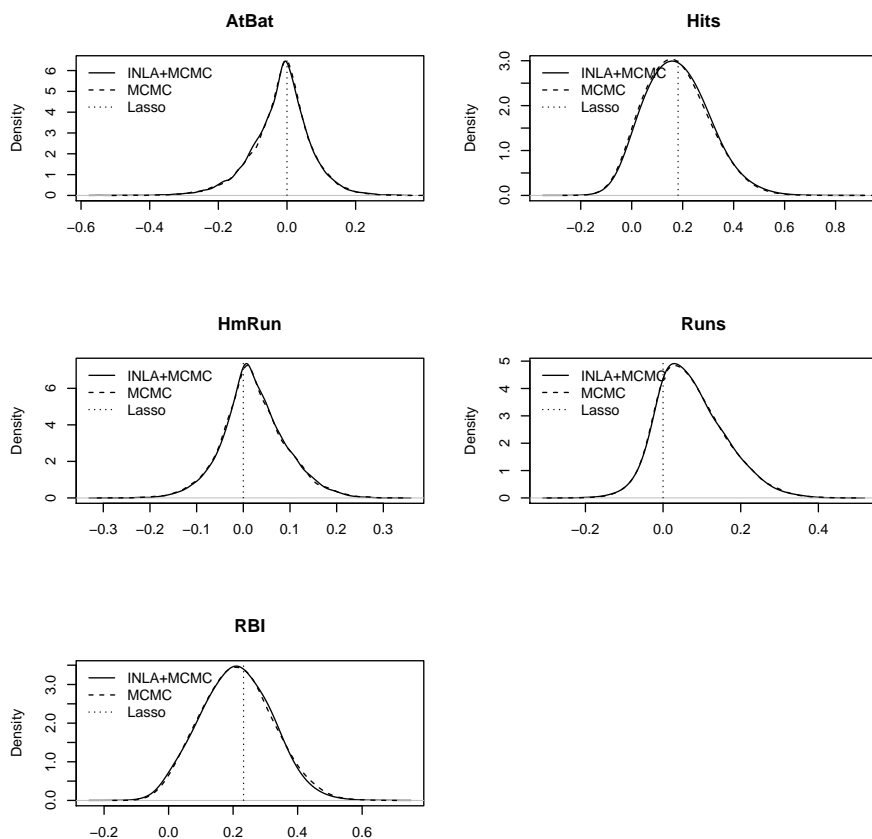


Figure 4: Summary of results for the Lasso and Bayesian Lasso.

could be considered (see, for example, Little and Rubin, 2002). As in previous examples, we will fit the same model using MCMC in order to compare both results. The model that we will fit is:

$$\begin{aligned}
 \text{chl}_i &= \beta_0 + \beta_1 \text{bmi}_i + \beta_2 \text{age}_{2_i} + \beta_3 \text{age}_{3_i} + \varepsilon_i \\
 \beta_0 &\propto 1 \\
 \beta_k &\propto N(0, 0.001); \quad k = 1, 2, 3 \\
 \varepsilon_i &\sim N(0, \tau) \\
 \tau &\sim Ga(1, 0.00005)
 \end{aligned} \tag{19}$$

Figure 5 shows the posterior marginal distributions of the imputed values of the body mass index. Both MCMC and our approach provide very similar point estimates. Table 2 summarises the model parameters obtained both with MCMC and our approach and Figure 6 displays the posterior marginals

of the model parameters obtained with our approach and MCMC. In all cases, the marginals agree, and the point estimates look very similar.

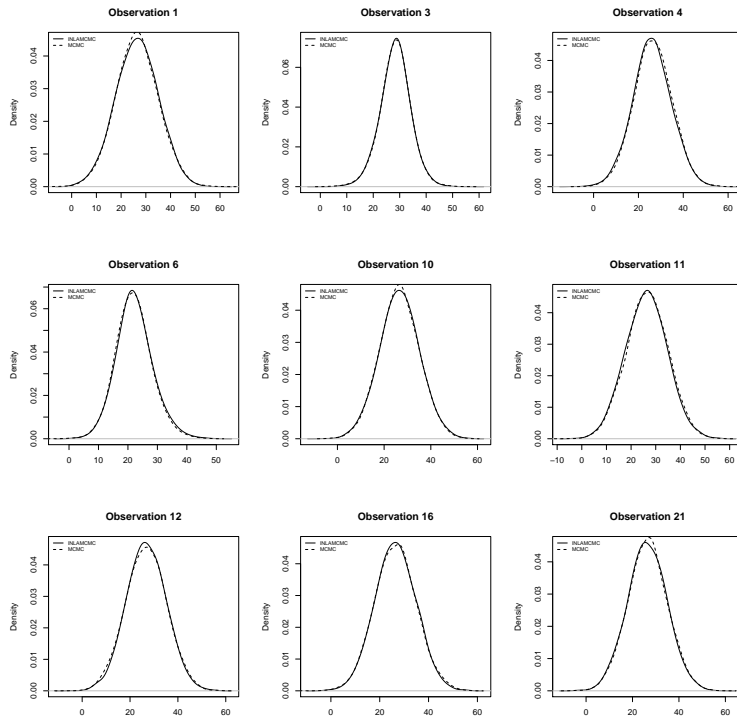


Figure 5: Marginal distributions of the imputed values of body mass index.

### 6.3 Spatial econometrics models

Bivand et al. (2014) describe a novel approach to extend the classes of models that can be fitted with **R-INLA** to fit some spatial econometrics models. In particular, they fit several conditional models by fixing the values of some of the parameters in the model, and then they combine these models using a Bayesian model averaging approach (Hoeting et al., 1999). Bivand et al. (2015) show a practical implementation with a spatial statistics model using R package **INLABMA**. Some of these models have already been included in **R-INLA** (Gómez-Rubio et al., 2017) but are still considered as experimental.

In this example we will focus on one of the spatial econometrics models described in Bivand et al. (2014) to illustrate how our new approach to combine MCMC and **R-INLA** can be used to fit unimplemented models. In

Parameter	MCMC	INLA+MCMC
$\beta_0$	39.760 (61.463)	43.469 (62.603)
$\beta_1$	4.994 (2.167)	4.864 (2.206)
$\beta_2$	29.989 (17.542)	29.501 (17.871)
$\beta_3$	50.049 (23.277)	49.449 (23.207)
$\tau$	0.001 (0.0005)	0.001 (0.0005)

Table 2: Summary of model parameter posterior estimates: posterior mean and standard deviation (in parentheses), model with missing covariates.

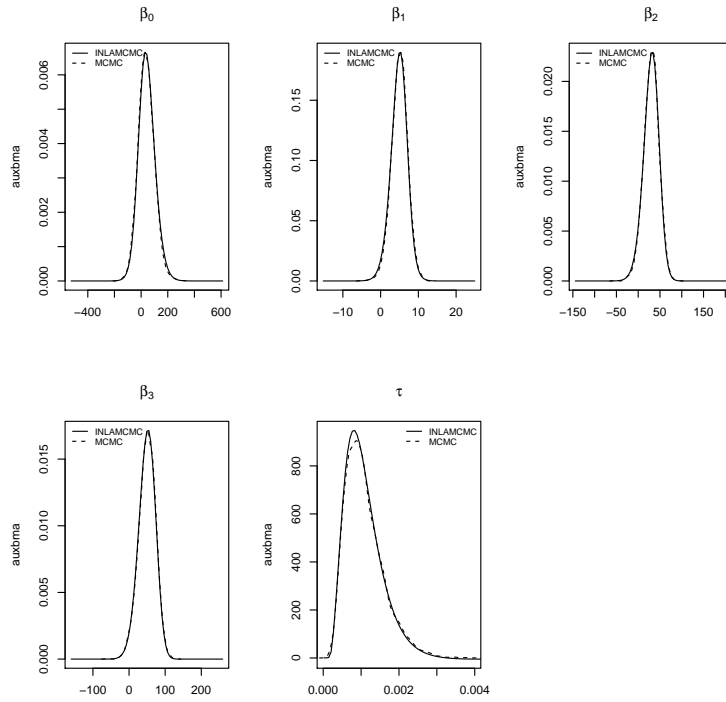


Figure 6: Marginal distributions of the model parameters, model with missing values in the covariates.

particular, we will consider the spatial lag model (LeSage and Pace, 2009):

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u}; \quad \mathbf{u} \sim N(0, \sigma_u^2 \mathbf{I})$$

Here,  $\mathbf{y}$  is a vector of observations at  $n$  areas,  $\mathbf{W}$  is an adjacency matrix,  $\rho$  a spatial autocorrelation parameter,  $\mathbf{X}$  a  $n \times p$  matrix of covariates with associated coefficients  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$  and  $\mathbf{u} = (u_1, \dots, u_n)$  an error term.  $u_i$ ,  $i = 1, \dots, n$ , is Normally distributed with zero mean and precision  $\tau_u$ .

Parameter	MaxLik	INLA+MCMC	MCMC	INLA+BMA
Intercept	61.05 (5.31)	60.62 (6.08)	58.53 (6.92)	60.81 (5.33)
$\beta_{\text{h. income}}$	-1.00 (0.34)	-0.97 (0.37)	-0.91 (0.39)	-0.98 (0.33)
$\beta_{\text{h. value}}$	-0.31 (0.09)	-0.31 (0.09)	-0.30 (0.10)	-0.31 (0.09)
$\rho$	0.52 (0.14)	0.55 (0.13)	0.55 (0.16)	0.54 (0.11)
$\tau_u$	0.01 (-)	0.01 (0.002)	0.01 (0.002)	0.01 (0.00004)

Table 3: Posterior means (and standard deviation) of the spatial lag model fitted to the Columbus data set using three different methods.

This model can be rewritten as follows:

$$\mathbf{y} = (\mathbf{I}_n - \rho \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}; \boldsymbol{\varepsilon} \sim N \left( 0, \frac{1}{\tau_u} ((\mathbf{I}_n - \rho \mathbf{W})(\mathbf{I}_n - \rho \mathbf{W}'))^{-1} \right)$$

This model is difficult to fit with any standard software for mixed-effects models because of parameter  $\rho$ . If the value of  $\rho$  is fixed, then it is easy to fit the model with **R-INLA** as it becomes a linear term on the covariates plus a random effects term with a known structure. Hence, by conditioning on the value of  $\rho$  we will be able to fit the model with **R-INLA**. In order to use our new approach, we will be drawing values of  $\rho$  using MCMC and conditioning on this parameter to fit the models with **R-INLA**.

Regarding prior distributions,  $\rho$  is assigned a uniform between  $-1.5$  and  $1$ ,  $\beta_i$ ,  $i = 1, \dots, p$  a Gaussian prior with zero mean and precision  $0.001$  (the default), and  $\tau_u$  is assigned a Gamma distribution with parameters  $1$  and  $0.00005$  (the default for the precision of a 'generic0' latent class in **R-INLA**).

We have fitted this model to the Columbus dataset available in **R** package **spdep**. This dataset contains information about 49 neighbourhoods in Columbus (Ohio) and we have considered a model with crime rates as the response and household income and housing value as covariates. We have also fitted the spatial lag model using a maximum likelihood approach, the method proposed by Bivand et al. (2014) and MCMC using an implementation of the model for the Jags software included in package **SEjags**, which can be downloaded from Github. The results are shown in Table 3. All Bayesian approaches have very similar estimates, and these are also very similar to the maximum likelihood estimates.

## 7 Discussion

In this paper, we have developed a novel approach to extend the models that can be fitted with INLA. For this, we have used INLA within the Metropolis-Hastings algorithm, so that only a few number of parameters are sampled.

We have shown three important applications. In the first one, we have implemented a Bayesian Lasso for variable selection using Laplace priors on the coefficient of the covariates. By following this example, other priors could be easily used with INLA. This includes not only univariate priors but multivariate priors, that are seldom available in **R-INLA**.

In our second example we have tackled the problem of imputation of missing covariates in model fitting. Here, we have included a very simple imputation method for the missing values in the covariates, so that model fitting and imputation were done at the same time. Compared to fitting the same model with MCMC, we obtained the same posterior estimates. In an ongoing work, Cameletti et al. (2017) explore how this can be extended to larger problems and how different imputation models and missingness mechanisms can be properly addressed with INLA and MCMC.

Finally, we have also shown how other models not included in the **R-INLA** software can be fitted with INLA and MCMC. In particular, we have fitted a spatial econometrics model by fitting conditional models on the spatial autocorrelation parameter. This method can be easily modified to suit any other models. In particular, Gibbs sampling could be used if the full conditionals are available for a subset of model parameters.

To sum up, we believe that this approach can be employed together with INLA to fit more complex models and that it can also be combined with other MCMC algorithms to develop simple samplers to fit complex Bayesian hierarchical models. This method can work well when the conditional models are hard to explore with current approaches for which INLA provides a fast approximation, such as geostatistical models. Furthermore, INLA could be embedded into a Reversible Jump MCMC algorithm so that once the model dimension has been set, the resulting model is approximated with INLA. See, for example, Chen et al. (2000) for a comprehensive list of MCMC algorithms that could benefit from embedding INLA.

## 8 Acknowledgements

Virgilio Gómez-Rubio has been supported by grant PPIC-2014-001, funded by Consejería de Educación, Cultura y Deportes (JCCM) and FEDER, and grant MTM2016-77501-P, funded by Ministerio de Economía y Competitividad.

## References

- Andrieu, C. and G. O. Roberts (2003). The pseudo-marginal approach to efficient monte carlo computations. *Genetics* 37(2), 697–725.
- Beaumont, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* 164, 1139–1160.
- Bivand, R. S., V. Gómez-Rubio, and H. Rue (2014). Approximate Bayesian inference for spatial econometrics models. *Spatial Statistics* 9, 146–165.
- Bivand, R. S., V. Gómez-Rubio, and H. Rue (2015). Spatial data analysis with **R-INLA** with some extensions. *Journal of Statistical Software* 63(20), 1–31.
- Cameletti, M., V. Gómez-Rubio, and M. Blangiardo (2017). Missing data analysis with the Integrated Nested Laplace Approximation. *In preparation*.
- Chen, M.-H., Q.-M. Shao, and J. G. Igrahim (2000). *Monte Carlo Methods in Bayesian Computation*. Springer, New York.
- Gilks, W., W. Gilks, S. Richardson, and D. Spiegelhalter (1996). *Markov Chain Monte Carlo in practice*. Boca Raton, Florida: Chapman & Hall.
- Gómez-Rubio, V., R. S. Bivand, and H. Rue (2017). Estimating spatial econometrics models with integrated nested laplace approximation. *Arxiv preprint at <https://arxiv.org/abs/1703.01273>*.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97 – 109.
- Hoeting, J., A. R. David Madigan and, and C. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science* 14, 382–401.
- Hubin, A. and G. Storvik (2016a, April). Efficient mode jumping MCMC for Bayesian variable selection in GLMM. *ArXiv e-prints*.
- Hubin, A. and G. Storvik (2016b, November). Estimating the marginal likelihood with Integrated nested Laplace approximation (INLA). *ArXiv e-prints*.
- James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An Introduction to Statistical Learning with Applications in R*. Springer.

- LeSage, J. and R. K. Pace (2009). *Introduction to Spatial Econometrics*. Chapman and Hall/CRC.
- Li, Y., P. Brown, H. Rue, M. al-Maini, and P. Fortin (2012). Spatial modelling of Lupus incidence over 40 years with changes in census areas. *Journal of the Royal Statistical Society, Series C* 61, 99–115.
- Little, R. J. A. and D. B. Rubin (2002). *Statistical analysis with missing data*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Lykou, A. and I. Ntzoufras (2011). WinBUGS: a tutorial. *Wiley Interdisciplinary Reviews: Computational Statistics*, 385 – 396.
- Martins, T. G., D. Simpson, F. Lindgren, and H. Rue (2013). Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis* 67, 68–83.
- Medina-Aguayo, F. J., A. Lee, and G. O. Roberts (2016). Stability of noisy metropolis-hastings. *Statistical Computing* 26, 1187–1211.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics* 21, 1087 – 1091.
- Pettit, L. I. (1990). The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society. Series B (Methodological)* 52(1), pp. 175–184.
- Plummer, M. (2016). *rjags: Bayesian Graphical Models using MCMC*. R package version 4-6.
- Rue, H., S. Martino, and N. Chopin (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.
- Rue, H., A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren (2017). Bayesian computing with INLA: A review. *Annual Reviews of Statistics and Its Applications*. To appear.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall, London.



- Sherlock, C., A. H. Thiery, G. O. Roberts, and J. S. Rosenthal (2015). On the efficiency of pseudo-marginal random walk metropolis algorithms. *The Annals of Statistics* 43(1), 238–275.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 64(4), 583–616.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58(1), 267 – 288.
- Tierney, L. and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *81*(393), 82–86.
- van Buuren, S. and K. Groothuis-Oudshoorn (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software* 45(1), 1–67.