PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
ESCUELA DE POSGRADO

APPROXIMATE BAYESIAN INFERENCE FOR
DIRECTED ACYCLIC GRAPH AUTOREGRESSIVE MODELS

TESIS PARA OPTAR POR EL GRADO ACADÉMICO DE MAGÍSTER EN ESTADÍSTICA

AUTOR

Julio Cesar Buendia Narvaez

ASESORA

Dra. Zaida Jesús Quiroz Cornejo

Diciembre, 2021

Dedication

To my mother Sophia, who taught me that with hard work and focus we can achieve everything and that our dreams can come true. Without your love and sacrifices, I would not be who I am today.

To my father Cipriano, who has stuck with me through everything and always supported me, no matter what. To my sister Ruth, who has given me invaluable advices and encouragement. I will be forever grateful.

To Marilyn, she has also been my biggest supporter.

Acknowledgement

First, I thank God for giving me the opportunity to complete this thesis despite all the challenges.

This thesis would not have been possible without Prof. Zaida Quiroz, who has been my mentor and is one of the most inspiring scholars I know. Her guidance, motivation, insights, support and patience have led me to complete my thesis. I would like to express my gratitude for introducing me to the field of spatial statistics.

Moreover, I would like to thank all professors at PUCP for the time they spent reading the thesis, and their critical examination to improve my research.

And last but not least, my thanks to Declan, my godson, for your contagious smile.

Resumen

La prevalencia de enfermedades epidemiológicas recolectadas en áreas geográficamente limitadas, como distritos o provincias, son cruciales para la toma de decisiones en salud pública. Usualmente esta variable respuesta presenta dependencia espacial, es decir, es similar en áreas vecinas, debido a la naturaleza de la enfermedad, clima, nivel económico y cultural, entre otras razones. En este sentido, se proponen modelos espaciales de datos áreas para identificar tendencias y factores asociados a enfermedades epidemiológicas, tomando en cuenta la dependencia espacial entre áreas geográficas. Por lo general, estos modelos ajustan a la dependencia espacial a través de efectos aleatorios derivados a través de grafos. En particular, el modelo autorregresivo de gráfico acíclico dirigido (DAGAR) se basa en un grafo acíclico dirigido y algunos efectos aleatorios "del pasado". Como consecuencia, la matriz de precisión (inversa de la covarianza) del modelo es dispersa. Este modelo tiene una interpretación intuitiva de los parámetros asociados con la dependencia espacial y se puede representar como un modelo gaussiano latente. En este contexto, en esta tesis se propone implementar el modelo DAGAR a través del método de inferencia bayesiano aproximado INLA que es determinista, bastante preciso y eficiente. Dentro de este enfoque, la estimación de datos grandes se puede realizar en segundos o minutos, y permite ajustar los datos con distribución gaussiana o no gaussiana. Finalmente, para mostrar el aporte de esta propuesta, el modelo DAGAR se ajusta a datos reales.

Palavras-clave: CAR, DAGAR, INLA, modelos gaussianos latentes.

Abstract

The prevalence of epidemiological diseases collected in geographically limited areas, such as districts or provinces, are crucial for making public health decisions. It is common that this response variable presents spatial dependence, that is, it is similar in neighboring areas, due to the nature of the disease, weather, economy and cultural level, among other reasons. In this sense, spatial models for areal data are proposed to identify trends and factors associated with epidemiological diseases, taking into account the spatial dependence between geographic areas. Usually, these models fit the spatial dependence through spatial random effects built from graphs and conditional distributions. In particular, the directed acyclic graph autoregressive (DAGAR) model is based on a directed acyclic graph and some "past" random effects. As a consequence, the precision matrix (inverse of the covariance) of the model is sparse. This model has an intuitive interpretation of the parameters associated with spatial dependence and can be represented as a latent Gaussian model. In this context, we propose in this project to implement the DAGAR model throughout the approximate Bayesian inference method INLA which is deterministic, quite accurate and efficient. Within this approach, estimation of large data can be carried out in seconds or minutes, and it allows to fit data following a Gaussian distribution or non-Gaussian distributions. Finally, in order to show the contribution of this proposal, the DAGAR model will be fitted to real data.
Keywords: CAR, DAGAR, INLA, latent Gaussian models.

Contents

Abbreviations and acronyms

INLA	Integrated nested Laplace approximations.
MCMC	Markov chain Monte Carlo.
AR	Autoregressive.
CAR	Conditional autoregressive model.
SAR	Simultaneous autoregressive model.
DAGAR	Directed acyclic graph autoregressive model.
GMRF	Gaussian Markov Random Field.
pdf	Probability distribution (density) function.
RAM	Random Access Memory.
WAIC	Widely Applicable Information Criterion.
CPO	Conditional Predictive Ordinate.
LPML	Logarithm of the pseudo marginal likelihood.
RMSEE	Root of mean squared estimation error.
SIDS	Sudden Infant Death Syndrome.
NC	North Carolina.
PCR	Polymerase chain reaction.

Symbols

Q	Precision matrix.
$E(X)$	Expected value of a random variable X.
$f_X(\cdot)$	Probability density function of X.

Chapter 1

Introduction

Over the last decade a lot of variables were collected on geographical areas due to the advancement of technology. It is quite often that the value of a variable on nearest neighbor areas is similar, this can be evidence of spatial dependence. For instance, in the jungle it is known that the number of cases of dengue must be higher in nearest areas, essentialy due to the precipitations and high temperatures. In practice many disciplines deal with spatial dependence, for instance, public health, environmental sciences, social sciences, among others. Statistical models for areal data are important to detect covariates that influence the response variable of interest as well as to describe the spatial distribution of the response variable.

Usually the conditional autoregressive model (CAR) (Besag, 1974) and the simultaneous autoregressive model (SAR) (Whittle, 1954; Cressie, 1993) are used to fit areal data. Both models need graphs to define the neighbors of each area. The random spatial effects of the CAR model are built assuming conditional dependence between neighbor areas and jointly follow a multivariate Gaussian distribution. The random spatial effects of the SAR model are linearly dependent of the neighbor spatial effects and independent errors and jointly also follow a multivariate Gaussian distribution. Both models depend on a spatial parameter, whose interpretation is quite difficult because a high value of these parameter implies a modest spatial autocorrelation between the areas.

Considering the limitation of spatial models for areal data in terms of interpretation of the spatial parameters and availability of valid priors Datta et al. (2019) developed a new model for areal data called autoregressive model of acyclic directed graphs (DAGAR) model, which joins the SAR model and a directed acyclic graph. They extend the idea of nearest neighbor Gaussian processes (NNGP, Datta et al. (2016)) for geostatistical data to areal data. As a result, the multivariate Gaussian distribution of the DAGAR model enjoys an sparse precision matrix, thus DAGAR model is scalable for large spatial data. Moreover, the interpretation of a parameter related to the spatial autocorrelation is better, that is, a high value of these parameter implies a high spatial autocorrelation between the areas.

The model is an spatial generalized linear model and can be represented hierarchically, thus Bayesian inference is suitable for this model. For instance, Datta et al. (2019) estimate the parameters of the DAGAR model, through Markov Monte Carlo chains (MCMC) methods. Nevertheless, this method is computationally expensive due to the large number of iterations required to update the parameters. In this thesis, the estimation of parameters of the DAGAR model will be achieved through the Integrated Nested Laplace Approach

(INLA) Rue et al. (2009). This method consists in applying the Laplace approximation to integrate out high-dimensional latent components. It is more efficient than the MCMC and quite accurate. In fact, the implementation will exploit the sparsity of the precision matrix of the DAGAR model, therefore, it allows to fit large spatial datasets. Furthermore, the DAGAR model will be implemented for Gaussian and non-Gaussian data through the binomial, Poisson and Gamma distributions, and zero-inflated distributions, among others. Extensions to spatio-temporal DAGAR models are also quite feasible. In summary, this implementation will accommodate a wide range of DAGAR models for areal data.

In addition, this model will be fitted to a real diseases and epidemiological dataset using INLA. Thus this thesis is of great relevance in the public health aspect of the country, since it will allow to identify which factors can influence the prevalence of diseases, as well as to study the spatial distribution of the disease along the districts or counties of a country. These new epidemiological models are expected to make a significant contribution to our understanding of the behavior and immunology of the viruses and its control in the near future.

The main goal of this thesis is to implement approximate Bayesian inference method through INLA for the DAGAR spatial model for large datasets. In particular, the DAGAR model will be implemented using INLA, some simulation studies will be performed in order to assess the performance of the DAGAR model, these results will be contrasted with competing models like the CAR and SAR models. Finally, the DAGAR model will be applied to fit real large areal data using INLA.

The outline of the thesis is detailed as follows: Chapter 2 focuses on a review of areal data models such as the conditional autoregressive model and the directed acyclic graph autoregressive model. Furthermore, we describe Bayesian inference methods and the integrated nested Laplace approximation method. Chapter 3 presents the properties of the DAGAR model. The details of bayesian inference for this model using INLA are also presented in this chapter. Chapter 4 shows the performance of bayesian inference for DAGAR models using INLA through simulations. Chapter 5 shows applications of the proposed model. Finally, some conclusions are remarked in 6.

Chapter 2

Preliminary concepts

This chapter presents some important concepts for areal data models. Areal data can be generated when a fixed domain is divided into a finite number of areas regular (lattice) or irregular (counties, provinces, among others). Formally, let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ be random variables for each area $1, 2, \ldots, n$. An important feature of areal units is the evidence of spatial autocorrelation. It suggests that measurements for areal units which are near to each other will tend to take more similar values. In order to incorporate the spatial autocorrelation in the model, the neighborhood structure among areas is represented through a graph that is used to built the adjacency matrix associated to this graph.

Let $\mathcal{G} = \{V, E\}$ denote an undirected graph with n nodes in V, corresponding to each area, and edges E between neighbor areas. For instance, Figure 2.1 shows an undirected graph with $n = 5$ nodes $\boldsymbol{V} = \{v_1, v_2, v_3, v_4, v_5\}$ and edges $\boldsymbol{E} = \{(v_1, v_2), (v_2, v_3), (v_2, v_5), (v_4, v_5)\}$ representing that the node one is neighbor of node two, node two is neighbor of nodes one, three and five, node three is neighbor of nodes two and five, node four is neighbor of area five, and node five is neighbor of nodes two, three and four. In practice, for areal data, the nodes represent the areas.

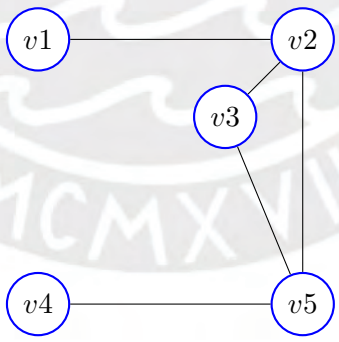


Figure 2.1: Undirected graph of five nodes.

Another key concept is the adjacency matrix $\boldsymbol{W}$, that is a matrix of weights w_{ij} representing the neighborhood structure. The weights w_{ij} are related to areas i and j. The

relationship among these neighbors areas can be expressed in the adjacency matrix:

$$\boldsymbol{W} = \begin{pmatrix} w_{1,1} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,n} \end{pmatrix}.$$

So $w_{ij} \neq 0$ if and only if i and j areas are neighbors and $w_{ij} = 0$ in any other case (Blangiardo y Cameletti, 2015). Therefore, w_{ij} can be seen as weights where the weight will be greater as there is greater proximity between the areas. For instance, Figure 2.2 shows the adjacency matrix representation of the graph $\mathcal{G}$, where $w_{ij} = 1$ if i and j share some common boundary, and $w_{ij} = 1$ otherwise.

$$\boldsymbol{W} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Figure 2.2: Adjacency matrix of the graph in Figure 2.1.

In order to define spatial models for areal data, the spatial correlation can be incorporated through a spatial random effects u_i for each area $i = 1, \ldots, n$. In the next section it is presented a brief review of the most common models for areal data.

2.1 Models for areal data

2.1.1 Conditional autoregressive (CAR) model

CAR models were introduced by Besag (1974) and have been extensively used in spatial statistics to fit areal data. Let assume that a geographic region is partitioned into areas indexed by integers $i = 1, ..., n$. Then the vector of spatial random effects for n areas is defined as $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^T$. In the CAR model, the structured random spatial effect u_i of an area i, conditioned on the rest random spatial effects $\boldsymbol{u_{-i}} = (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n)^T$, follows a Gaussian distribution, that is:

$$u_i | \boldsymbol{u_{-i}} \sim N(\sum_{j=1}^{n} b_{ij} u_j, \frac{1}{\tau_i}),$$

where $b_{ij} \geq 0$ are covariance parameters, with $b_{ii} = 0$ for all i and τ is a precision parameter. This full conditional distributions determine a well defined joint distribution for $\boldsymbol{u}$ if $b_{ij}\tau_i = b_{ji}\tau_j$ for all $i, j = 1, \ldots, n$, and $\boldsymbol{F}(\boldsymbol{I} - \boldsymbol{B})$ is positive definite, for $\boldsymbol{B} = (b_{ij})$ and $\boldsymbol{F} = \text{diag}(1/\tau_1, \ldots, 1/\tau_n)$. Therefore

$$\boldsymbol{u} \sim N(0, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{F}).$$

In particular, in this thesis it is assumed that $\boldsymbol{B} = \rho\boldsymbol{W}$ where ρ is an unknown spatial parameter, also known as spatial autocorrelation parameter, $\boldsymbol{W}$ is an adjacency matrix, and $\boldsymbol{F} = \boldsymbol{I_n}/\tau$ then

$$\boldsymbol{u} \sim N(0, (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}/\tau).$$

It is important to remark that to guarantee that $(\boldsymbol{I}-\rho\boldsymbol{W})$ is positive definite, ρ is constrained into the interval $(\lambda_{(1)}^{-1}, \lambda_{(n)}^{-1})$ where $\lambda_{(1)} \leq \lambda_{(2)} \leq \cdots \leq \lambda_{(n)})$ are the ordered eigenvalues of $\boldsymbol{W}$, with $\lambda_{(1)} < 0 < \lambda_{(n)}$. In order to constrain $\rho < 1$, the adjacency matrix can be re-scaled by dividing it by $\lambda_{(n)}$, that is, $\widetilde{\boldsymbol{W}} = \boldsymbol{W}/\lambda_{(n)}$ (Haining, 2003). Usually, $\rho \in (0, 1)$, thus it can be interpreted similarly as a correlation parameter. Finally, the CAR distribution of $\boldsymbol{u}$ is given by:

$$\boldsymbol{u} \sim N(0, (\boldsymbol{I} - \rho\widetilde{\boldsymbol{W}})^{-1}/\tau),$$

where the precision matrix is $\boldsymbol{Q}_{CAR} = \tau(\boldsymbol{I} - \rho\widetilde{\boldsymbol{W}})$.

Banerjee et al. (2014) shows for a similar specification of the CAR model that within a Bayesian framework, a prior on ρ that encourages a consequential amount of spatial association would place most of its mass near one. This means that even the spatial association is not too strong, the ρ parameter would tend to take high values.

The spatial random effect $\boldsymbol{u}$ can be incorporated into a spatial model. Specifically, if $E(Y_i) = \mu_i$ then,

$$g(\mu_i) = z_i^\top \boldsymbol{\beta} + u_i,$$

where $g(\cdot)$ is some suitable link function, z_i is a vector of covariates and $\boldsymbol{\beta}$ is a vector of regression coefficients. And from a Bayesian approach, a CAR prior can be assigned for $\boldsymbol{u}$.

2.1.2 Simultaneous autoregressive (SAR) model

The Simultaneous autoregressive (SAR) model was introduced by Whittle (1954) and proceeds by simultaneously modeling the spatial random effects u_i. It is considered that the random spatial effect of an area i, depends on its neighbor random effects, that is

$$u_i = \rho \sum_j b_{ij} u_j + \epsilon_i, \text{ for } i = 1, 2, \ldots, n,$$

where $\epsilon_i \overset{\text{ind}}{\sim} N(0, 1/\tau_i)$ are non-structured errors and b_{ij} are known constants. And it is assumed that the errors ϵ_i are independent of u_i. Then from the next set of equations:

$$u_1 = \rho \sum_{j \neq 1} b_{1j} u_j + \epsilon_1,$$

$$u_2 = \rho \sum_{j \neq 2} b_{2j} u_j + \epsilon_2,$$

$$\vdots$$

$$u_n = \rho \sum_{j \neq n} b_{nj} u_j + \epsilon_n,$$

$\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^T$ follows a normal distribution, that is,

$$\boldsymbol{u} \sim N\left(0, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{F}((\boldsymbol{I} - \boldsymbol{B})^{-1})^T\right),$$

where $\boldsymbol{F} = diag(1/\tau_1, \ldots, 1/\tau_n) = \boldsymbol{I_n}/\tau$ and the matrix $\boldsymbol{B} = \rho\boldsymbol{W}$, where ρ is referred as a spatial parameter, and $(\boldsymbol{I} - \boldsymbol{B})$ is positive definite if ρ is constrained into the interval $(\lambda_{(1)}^{-1}, \lambda_{(n)}^{-1})$ being $\lambda_{(1)} \leq \lambda_{(2)} \leq \cdots \leq \lambda_{(n)})$ the ordered eigenvalues of $\boldsymbol{W}$, with $\lambda_{(1)} < 0 < \lambda_{(n)}$. In order to constrain $\rho < 1$, the adjacency matrix can be re-scaled by dividing it by $\lambda_{(n)}$, that is, $\widetilde{\boldsymbol{W}} = \boldsymbol{W}/\lambda_{(n)}$ (Haining, 2003). Usually, $\rho \in (0, 1)$, thus it can be interpreted similarly as an autocorrelation spatial parameter. Nevertheless, with respect to interpretation of the parameter ρ, it leads to similar problems as the CAR models. Finally, the SAR distribution of $\boldsymbol{u}$ is given by:

$$\boldsymbol{u} \sim N\left(0, \frac{1}{\tau}(\boldsymbol{I} - \rho\widetilde{\boldsymbol{W}})^{-1}((\boldsymbol{I} - \rho\widetilde{\boldsymbol{W}})^{-1})^T\right),$$

where the precision matrix is $\boldsymbol{Q}_{SAR} = \tau(\boldsymbol{I} - \rho\widetilde{\boldsymbol{W}})(\boldsymbol{I} - \rho\widetilde{\boldsymbol{W}})^\top$.

The spatial random effect $\boldsymbol{u}$ can be incorporated into a spatial model. Specifically, if $E(Y_i) = \mu_i$ then,

$$g(\mu_i) = z_i^\top \boldsymbol{\beta} + u_i,$$

where $g(\cdot)$ is some suitable link function, z_i is a vector of covariates and $\boldsymbol{\beta}$ is a vector of regression coefficients. And from a Bayesian approach, a SAR prior can be assigned for $\boldsymbol{u}$.

2.2 Bayesian Inference

2.2.1 Markov Chain Monte Carlo (MCMC) methods

Given the availability of more powerful computational resources, MCMC algorithms were established as the core of Bayesian inference over the last decades. Let $\pi(\cdot)$ and $\pi(\cdot|\cdot)$ represent a probability density function or a conditional probability density function (pdf). The main goal of Bayesian inference relies on posterior estimation of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$, based on the the observed data $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ (likelihood function) and the prior information. Let $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ represent the likelihood function and let $\pi(\boldsymbol{\theta})$ be the prior distribution of $\boldsymbol{\theta}$. Then through the Bayes theorem, the posterior distribution of $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}|y) \propto \pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

For simple models, the posterior distribution of $\boldsymbol{\theta}$ is known, however for complex models like the spatial models for areal data, $\boldsymbol{\theta}$ has a high dimension and its posterior distribution does not enjoy a known form. In this context, MCMC algorithms like Gibbs sampling and the Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970) can be used to sample from the posterior distribution.

The Gibbs sampling breaks the problem of sampling from the posterior distribution in a serie of samples of full conditional (FC) distributions of lower dimension. The Gibbs sampling algorithm is summarized in the next steps:

1. Define initial values $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, ..., \theta_p^{(0)})$
2. For each iteration t,

   FC1 Sample from $\theta_1^{(t)}|\theta_2^{(t-1)}, ..., \theta_p^{(t-1)}, \mathbf{Y}$
   FC2 Sample from $\theta_2^{(t)}|\theta_1^{(t)}, \theta_3^{(t-1)}, ..., \theta_p^{(t-1)}, \mathbf{Y}$
   $\vdots$
   FCp Sample from $\theta_p^{(t)}|\theta_1^{(t)}, ..., \theta_{p-1}^{(t)}, \mathbf{Y}$

3. Repeat step 2, S times, to obtain the posterior samples

$$\boldsymbol{\theta}^{(1)}, ..., \boldsymbol{\theta}^{(S)}.$$

The Gibbs sampling algorithm ensures that for any initial values the Markov chain will converge to the posterior distribution. This algorithm works if the full conditional distributions are known. When we are not able to sample from the full conditional distributions, another nice MCMC algorithm is the Metropolis and Metropolis-Hasting.

Let assume that $\boldsymbol{\theta}_{(j)} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p)$. Without loss of generality the Metropolis algorithm to update θ_j is summarized in the next steps:

1. Choose an initial value $\theta_j^{(0)}$, such that $p(\theta_j^{(0)}|\theta_{(j)}^{(t-1)}, \mathbf{Y}) > 0$.
2. In iteration t, sample a candidate value θ_j^* from the proposal distirbution $q(\theta|\theta_j^{t-1})$.
3. Compute the acceptance probability:

$$R = \min\left\{1, \frac{\pi(\theta_j^*|\theta_{(j)}^{t-1}, \mathbf{Y})}{\pi(\theta_j^{t-1}|\theta_{(j)}^{t-1}, \mathbf{Y})}\right\}$$

4. If $R = 1$ then accept θ_j^* , and update $\theta_j^{(t)} = \theta_j^*$. Otherwise, if $R < 1$, sample from $r \sim Uniforme(0, 1)$, such that if:
   - $r < R$ then accept θ_j^*, and update $\theta_j^{(t)} = \theta_j^*$.
   - Other case, θ_j^* is rejected and $\theta_j^{(t)} = \theta_j^{(t-1)}$.
5. Repeat 2 to 4 until achieve the convergence of the Markov chain.

In the Metropolis-Hasting algorithm the proposal distribution can be asymmetric. The Metropolis and Metropolis-Hasing sampling algorithms ensure that for any initial values the Markov chain will converge to the posterior distribution. Further, note that the Gibbs sampling is a particular case of the Metropolis algorithm.

2.2.2 Integrated Nested Laplace approximation

The integrated nested Laplace approximation (INLA) (Rue et al., 2009) approach is a method to perform Bayesian inference for latent Gaussian models (LGMs). INLA allows more fast Bayesian inference than the MCMC method. In order to define the INLA algorithm, it is important to define Gaussian Approximation and latent Gaussian models.

Let $\pi(x|y)$ be a posterior density distribution of x given by

$$\pi(x|y) \propto \pi(x)\pi(y|x) = \exp(f(x)),$$

where $f_X(x)$ is the log-posterior of x that can be approximated using as quadratic Taylor expansion around $x = x_0$ by:

$$f(x) \approx f(x_0) + f^{'}(x_0)(x - x_0) + \frac{1}{2}f^{''}(x_0)(x - x_0)^2$$

$$= a + bx - \frac{1}{2}cx^2,$$

where $b = f^{'}(x_0) - f^{''}(x_0)x_0$ and $c = -f^{''}(x_0)$. The value of a is not relevant in the following. Therefore, the Gaussian approximation of $\pi(x|y)$ is defined as follows:

$$\tilde{\pi}_G(x|y) \propto \exp(-\frac{1}{2}cx^2 + bx),$$

which has the form of a normal distribution with mean b/c and precision c. As a result, the Gaussian approximation for the density $\pi(x|y)$ is better when x_0 is closer to the mode of $\pi(x|y)$. For more details, see Rue y Held (2005).

Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ be a vector of response variables, and assume that the distribution of Y_i is characterized by a parameter μ_i, which is the mean or some quantile. Then, the linear predictor can be defined as follows:

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{m=1}^{M} \beta_m z_{mi} + \sum_{l=1}^{L} f^{(l)}(u_{li}),$$

where $g(\cdot)$ is a link function, β_0 represents the intercept, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)^T$ represents the vector of regression (or fixed) coefficients, z_{mi} is the m-th covariate for the i-th observation, and $\boldsymbol{f} = (f^{(1)}(\cdot), \ldots, f^{(L)}(\cdot))^T$ is a collection of functions defined in terms of covariates u_{li}, the l-th covariate for the i-th observation. The terms $f^l(\cdot)$ can assume different forms such as spatial or temporal random effects. The vector of parameters $\boldsymbol{x} = \{\beta_0, \beta_1, \ldots, \beta_M, \boldsymbol{f}\}$ is a Gaussian Markov random field. Then a latent Gaussian model (LGM) is a hierarchical model that is composed by the observed data $\boldsymbol{y}$, latent Gaussian field $\boldsymbol{x}$ and hyperparameters $\boldsymbol{\theta}$ as follows:

- observed data $\boldsymbol{y}$: Assuming conditional independence of Y_i's given the Gaussian Random Markov Field and its hyperparameters, the conditional distribution of $\boldsymbol{Y}$ is defined

as follows:

$$\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi(y_i|x_i,\theta).$$

- latent Gaussian field $\boldsymbol{x}$: Assuming a multivariate Gaussian prior on $\boldsymbol{x}$ with mean 0 and precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$, then

$$\boldsymbol{x}|\boldsymbol{\theta} \sim N(0, \boldsymbol{Q}^{-1}(\boldsymbol{\theta})).$$

- hyperparameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$,

$$\pi(\boldsymbol{\theta}).$$

So, the joint posterior distribution of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}).$$

INLA computes the marginal posterior distributions for each element of the latent Gaussian field $\boldsymbol{x}$, that is

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

Also, the marginal posterior of the hyperparameter vector is defined as follows

$$\pi(\theta_m|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{-m},$$

where $\boldsymbol{\theta}_{-m} = (\theta_1, \ldots, \theta_{m-1}, \theta_{m+1}, \ldots, \theta_k)$. Specifically, the INLA approach exploits the assumptions of the model to produce a numerical approximation of the posteriors of interest based on the Gaussian approximation (Tierney y Kadane, 1986) method, that is

$$\begin{aligned}
\pi(\boldsymbol{\theta}|\boldsymbol{y}) &= \frac{\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}, \\
&= \frac{\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x},\boldsymbol{\theta})}{\pi(\boldsymbol{y})}\frac{1}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}, \\
&= \frac{\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{y})}\frac{1}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}, \\
&\propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|,\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})},
\end{aligned}$$

which is approximated by:

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x},\theta)\pi(\boldsymbol{\theta})}{\tilde{\pi_G}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}|_{x=x^*(\boldsymbol{\theta})} = \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}), \tag{2.1}$$

where $\tilde{\pi_G}$ denotes a Gaussian approximation to the full conditional of x, and $x^*(\theta)$ is the mode of the full conditional of $\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}$. Note that the approximation presented in Equation (2.1) is called Laplace approximation.

In a similar way,

$$\begin{aligned}\pi(x_j|\boldsymbol{\theta},\boldsymbol{y}) &= \frac{\pi((x_j,x_{-j}|\boldsymbol{\theta},\boldsymbol{y}))}{\pi(x_j|x_j,\boldsymbol{\theta},\boldsymbol{y})},\\ &= \frac{\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}{\pi(\boldsymbol{\theta}|\boldsymbol{y})}\frac{1}{\pi(x_{-j}|x_j,\boldsymbol{\theta},\boldsymbol{y})},\\ &\propto \frac{\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}{\pi(x_{-j}|x_j,\boldsymbol{\theta},\boldsymbol{y})},\end{aligned}$$

where $\boldsymbol{x} = (x_j, x_{-j})$, x_{-j} indicates the vector x without the i^{th} element. Thus $\pi(x_j|\boldsymbol{\theta},\boldsymbol{y})$ can be approximated using the Gaussian approximation by

$$\widetilde{\pi}(x_j|\boldsymbol{\theta},\boldsymbol{y}) \approx \frac{\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}{\widetilde{\pi}(x_{-j}|x_j,\boldsymbol{\theta},\boldsymbol{y})}|_{x_{-j}=x^*_{(x_j,\boldsymbol{\theta})}} = \tilde{\pi}(x_j|\boldsymbol{\theta},\boldsymbol{y}), \tag{2.2}$$

where $x^*_{(x_j,\theta)}$ is the mode of the full conditional distribution of x_{-j} that is obtained by using some optimization method. Also note that $\tilde{\pi}(x_j|\boldsymbol{\theta},\boldsymbol{y})$ is the Laplace approximation of $\pi(x_j|\boldsymbol{\theta},\boldsymbol{y})$.

Hence, the marginals of each parameter x_j and θ_m can be obtained using approximations in Equation (2.1) and (2.2) as follows,

$$\tilde{\pi}(x_j|\boldsymbol{y}) = \int \tilde{\pi}(x_j|\boldsymbol{\theta},\boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\theta. \tag{2.3}$$

$$\tilde{\pi}(\theta_m|\boldsymbol{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\theta_{-m}. \tag{2.4}$$

2.3 Model assesment

In order to study the goodness of fit of the studied models, the logarithm of the pseudo marginal likelihood (LPML), the Watanabe-Akaike (or "widely applicable") information criterion (WAIC), and the root of mean squared estimation error (RMSEE) are considered to measure the performance of each model.

The WAIC was introduced by Watanabe (2010). From a Bayesian perspective,the WAIC is based on the posterior predictive density, this is its main advantage over other measures. It is stated that the WAIC is particularly helpful for hierarchical models. Gelman et al. (2013) states that it can be computed as follows:

$$WAIC = -2\sum_{i=1}^{N}\left[\log\left\{\frac{1}{K}\sum_{k=1}^{K}\pi\left(y_i|\boldsymbol{x}^{(k)},\boldsymbol{\theta}^{(k)}\right)\right\} - V_{k=1}^{K}\log\left\{\pi\left(y_i|\boldsymbol{x}^{(k)},\boldsymbol{\theta}^{(k)}\right)\right\}\right],$$

where K is the number of posterior samples, $\left(\boldsymbol{x}^{(k)},\boldsymbol{\theta}^{(k)}\right)$ are samples from $\pi\left(x_i,\boldsymbol{\theta}|\boldsymbol{y}\right)$ and $V_{k=1}^{K}\left(\cdot\right)$ is the sample variance. The lower the value of WAIC, the better the model.

Another alternative Bayesian model criteria is the conditional predictive ordinate (Geisser

y Eddy, 1979), defined as

$$CPO_i = \pi\left(y_i|y_{-i}\right) = \frac{1}{\int \frac{\pi(x_i|y)}{\pi(y_{-i}|x_i)}} dx_i,$$

where y_{-i} is given by $\boldsymbol{y}$ without the i^{th} component. The Monte Carlo estimation for the CPO_i (Held et al., 2009) is defined as the harmonic mean of the conditional density $\pi\left(y_i|x_m, \theta_m\right)$,

$$\widehat{CPO_i} = \left[\frac{1}{K}\sum_{k=1}^{K} \frac{1}{\pi\left(y_i|\boldsymbol{x}^{(k)}, \boldsymbol{\theta}^{(k)}\right)}\right]^{-1},$$

evaluated at samples $\left(\boldsymbol{x}^{(k)}, \boldsymbol{\theta}^{(k)}\right)$ from $\pi\left(x_i, \boldsymbol{\theta}|\boldsymbol{y}\right)$. Moreover, since the CPO_i is a goodness of fit measure for each observation, it can be summarized for all data via a single value called the logarithm of the pseudo marginal likelihood (LPML), so comparison between models can be done using,

$$LPML = \sum_{i=1}^{n} \log \pi\left(y_i|y_{-i}\right),$$

that is, the higher the value of LPML, the better the model.

Finally, to assess the closeness between the posterior mean estimation of $\boldsymbol{y}$ and the observed $\boldsymbol{y}$ value, it is computed the root of mean squared estimation error (RMSE). The root of mean squared estimation error (RMSE) is computed as follows:

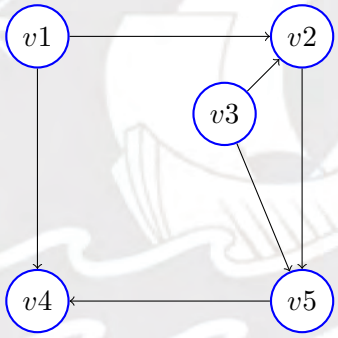
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2}; \text{ where } d_i = y_i - E\left(\widehat{Y_i|\boldsymbol{x}}, \boldsymbol{\theta}\right).$$

Chapter 3

Approximate Bayesian inference for DAGAR models

3.1 Directed acyclic graph

The key of the DAGAR model is a directed acyclic graph (DAG). A directed graph is a pair $(N, R) \subseteq N \times N$ consisting of a set of N nodes and a binary relation R, that specifies a directed edge from a node n to another node m whenever $(n, m) \in R$. An edge $(i, j) \in R$ is called directed if $(i, j) \in R$ but $(j, i) \notin R$. This graph does not contain any cycle. For instance, Figure 3.1 provides an example of a DAG structure $\mathcal{G}$ and the adjacency matrix of this graph.



$$\boldsymbol{W} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Figure 3.1: Directed graph and its adjacency matrix.

Another important feature of a DAG that we can also observe in this example is the sparsity of the adjacency matrix with respect to the graph, given that $w_{ij} \neq 0$ when there is a directed edge from i to j and $w_{ij} = 0$ otherwise. In practice many graphs give rise to sparse matrices when there are only a few edges between pairs of nodes.

3.2 Embedded spanning tree

A spanning tree $\mathcal{T}$ is a subgraph of an undirected graph $\mathcal{G}$, that is a tree which includes all of the nodes of $\mathcal{G}$ which are connected with the minimum possible number of edges. It cannot be disconnected as it does not have cycles. For instance, Figure 3.2 shows the spanning trees

of a graph $\mathcal{G}$ with three nodes $\boldsymbol{V} = \{v_1, v_2, v_3\}$ and edges $\boldsymbol{E} = \{(v_1, v_2), (v_1, v_3), (v_2, v_3)\}$. We can observe that a graph may have several spanning trees.

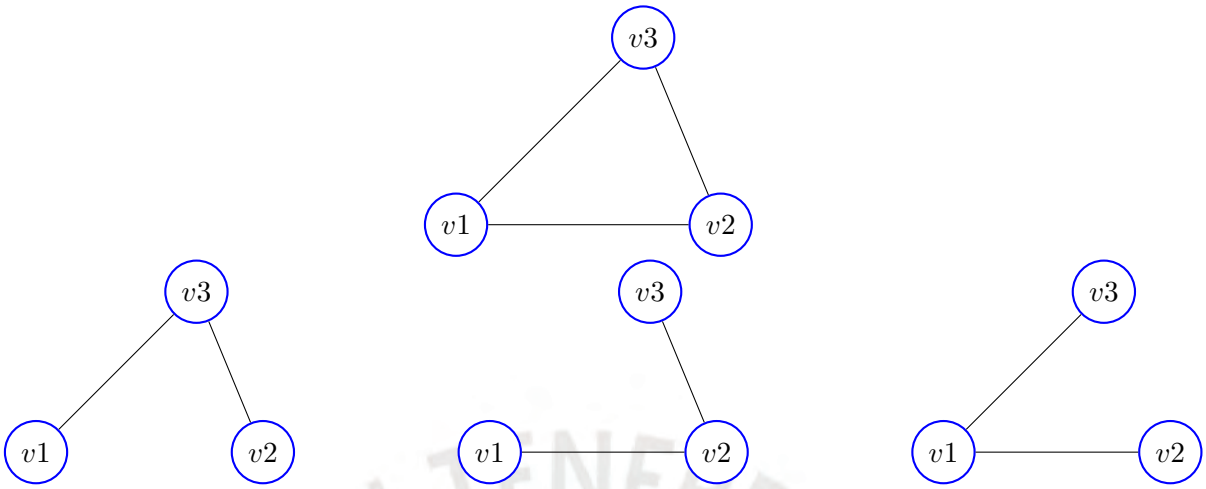


Figure 3.2: Spanning trees of $\mathcal{G}$.

A graph $\mathcal{G}$ is descomposed into a sequence of embedded spanning trees $\mathcal{T}_i$ when the joined embedded spanning trees give rise to the original graph. For instance Figure 3.3 shows the embedded spanning trees of a graph $\mathcal{G}$ with five nodes $\boldsymbol{V} = \{v_1, v_2, v_3, v_4, v_5\}$ and edges $\boldsymbol{E} = \{(v_1, v_2), (v_1, v_3), (v_2, v_3), (v_2, v_4), (v_3, v_4), (v_4, v_5)\}$.

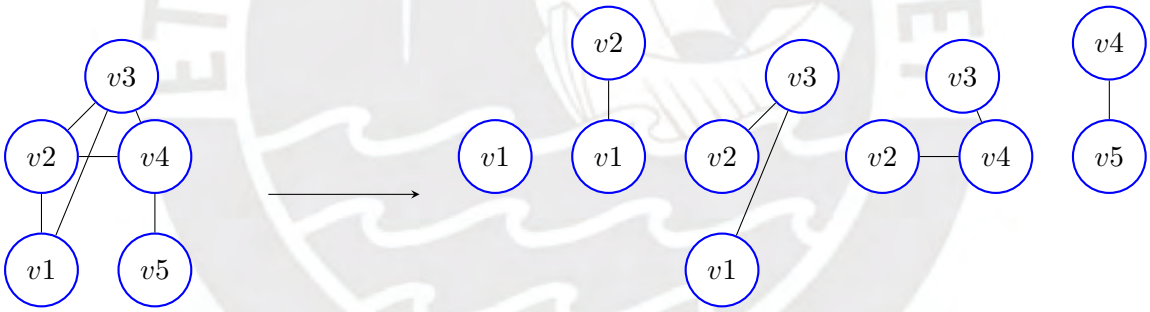


Figure 3.3: Decomposing a graph $\mathcal{G}$ into a sequence of embedded spanning trees $\mathcal{T}_i$.

3.3 Directed acyclic autoregressive (DAGAR) model

Let $\boldsymbol{u} = (u_1, u_2, \cdots, u_n)^T$ be a $n \times 1$ vector consisting of spatial random effects corresponding to each area $1, 2, \ldots, n$. From basic probability the joint density of $\boldsymbol{u}$ can be determined by the next conditional distributions of u_i's:

$$\pi(\boldsymbol{u}) = \pi(u_1)\pi(u_2|u_1)\pi(u_3|u_1, u_2)\cdots\pi(u_k|u_1, u_2, \cdots, u_{k-1}).$$

Under this definition, it is assume that each spatial random effect u_i depends on its "past" nearest neighbor areas. This approach was already used in some way to define the SAR model, where each random spatial effect u_i's can be defined by

$$u_1 = \epsilon_1; u_2 = b_{21}u_1 + \epsilon_2; \ldots, u_n = b_{n1}u_1 + \cdots + b_{n,n-1}u_{n-1} + \epsilon_n, \tag{3.1}$$

where $\epsilon_i \overset{\text{ind}}{\sim} N(0, 1/\tau_i)$ are errors independent of $\boldsymbol{u}$.

A natural extension of the SAR approach, is to assume that each random spatial effect u_i's depends on "some past" nearest neighbor areas. This extension was the assumption of the DAGAR random effects. Hence the DAGAR random effect u is defined by:

$$u_i = \sum_{\substack{j<i \\ j\sim i}} b_{ij} u_j + \epsilon_i; \text{ for } i = 1, 2, \ldots, n,$$

where $j \sim i$ denotes neighbors areas of i, $\epsilon_i \overset{\text{ind}}{\sim} N(0, 1/\tau_i)$ and b_{ij} are coefficients related to the spatial parameter of the model.

In order to construct the DAGAR model, it is neccesary to define the neighbor set of area i as $N(i) = \{j < i, j \sim i\}$, where $j \sim i$ points out that the areas i and j are neighbors. Then $b_{ij} = 0$ for all $j \notin N(i)$. And, the constraint $j < i$ implies that $\boldsymbol{B}$ is a lower triangular matrix, which ensures the positive definite property of the precision matrix of the DAGAR random effects. Then, the DAGAR random effects can be rewritten by:

$$u_1 = \epsilon_1, u_i = \sum_{j\in N(i)} u_j b_{ij} + \epsilon_i, \text{ for } i = 1, 2, \cdots, n, \tag{3.2}$$

where $\epsilon_i \overset{\text{ind}}{\sim} N(0, 1/\tau_i)$. To define $b_{ij} \neq 0$, that is which areas are neighbors, let $\mathcal{G} = (\boldsymbol{V}, \boldsymbol{E})$ be an acyclic graph (DAG) with the areas as nodes V and edges E between neighbors (past and nearest), and $\mathcal{G}_i$ be a subgraph of $\mathcal{G}$ comprising vertices $i \cup N(i)$ and the edges among them. Datta et al. (2019) proposed to use a local spanning trees of small subgraphs of $\mathcal{G}$ to construct the lower dimensional conditional densities specified in Equation (3.2) and therefore to define $b_{ij} \neq 0$. Specifically, an emmbedded spanning tree T_i of $\mathcal{G}_i$ was proposed to establish the final set $N(i)$ and the conditional density $u_i|\boldsymbol{u}_{N(i)}$.

Then, the spatial autocorrelation (ρ) between the random effects can be taken into account through the spanning tree T_i and covariance matrix of an a autoregressive process of order one (AR(1)) that depends on a parameter ρ defined as follows:

$$cov(u_i, \boldsymbol{u}_{N(i)}) = \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho^2 & \cdots & \rho^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho^2 & \rho^2 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 1 & v_i^T \\ v_i & \sum_i \end{pmatrix}, \tag{3.3}$$

where $0 \leq \rho \leq 1$, the matrix with elements $\rho^{d_{ij}}$ is positive definite, for d_{ij} denoting the length of the shortest path on $\mathcal{G}$ between nodes i and j; and v_i is the vector of covariances between u_i and $\boldsymbol{u}_{N(i)}$, and $\sum_i$ is the covariance matrix of $\boldsymbol{u}_{N(i)}$ assuming an AR(1) model on the spanning tree of $\mathcal{G}_i$. Then, $E(u_i|\boldsymbol{u}_{N(i)}) = \sum_{j\in N(i)} b_{ij} u_j, \tau_i = 1/Var(u_i|\boldsymbol{u}_{N(i)})$, and from Equations (3.3) and (3.2),

$$b_{ij} = \frac{\rho}{1 + (n_{<i})\rho^2}; i = 2, \cdots, n; j \in N(i);$$

$$\tau_i = \frac{1 + (n_{<i} - 1)\rho^2}{1 - \rho^2}, i = 1, \cdots, n.$$

where $n_{<i}$ denotes the cardinality of $N(i)$, that is, the number of neighbors of an area i. Thus, from Equation (3.2) it follows that $\boldsymbol{u} = \boldsymbol{B}\boldsymbol{u}+\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{F})$, $\boldsymbol{F} = \text{diag}\,(\tau_1, \tau_2, \ldots, \tau_n)$ and $\boldsymbol{B} = \{b_{ij}\}$ is a strictly lower triangular matrix. Hence, $\boldsymbol{u} \sim N\left(0, (\boldsymbol{I} - \boldsymbol{B})^\top \boldsymbol{F}(\boldsymbol{I} - \boldsymbol{B})\right)$. It is important to remark that the DAGAR precision matrix $\boldsymbol{Q}(\rho) = [(\boldsymbol{I} - \boldsymbol{B})^T \boldsymbol{F}(\boldsymbol{I} - \boldsymbol{B})]^{-1}$ (inverse of covariance function) of the DAGAR model is sparse. Thereby the model is useful for massive areal datasets. Furthermore, the interpretation of the parameter related to the spatial autocorrelation is more clear when we adopt the autoregressive model of an acyclic directed graph (DAGAR).

Finally, the DAGAR specification, further assumes that the resulting model

$$\boldsymbol{u} \sim N(0, \boldsymbol{Q}^{-1}(\rho)/\tau_u),$$

where $\boldsymbol{Q}(\rho)$ is the DAGAR precision matrix, is homoskedastic, and hence $1/\tau_u$ is the marginal variance. As a consequence, the DAGAR model ensures interpretability for both ρ and τ_u. For more details on DAGAR models see Datta et al. (2019).

The spatial random effects $\boldsymbol{u}$ following a DAGAR specification can be incorporated into an hierarchical model framework as priors and fitted through Bayesian inference.

3.4 Bayesian inference for DAGAR models

Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$ be the vector of response random variables Y_i for areas $i = 1, 2, \ldots, n$ and let assume that Y_i's are conditionally independent, given $\boldsymbol{\beta}, \boldsymbol{u}, \rho, \tau_u, \tau_e$. Then the conditional mean $E(Y_i|.)$ can be linked to the linear predictor η_i through a suitable link function $h(\cdot)$, i.e.,

$$h(E(Y_i|.)) = \eta_i = z_i^T \boldsymbol{\beta} + u_i + \epsilon_i, \tag{3.4}$$

where z_i represents a vector of covariates, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^T$ are the regression coefficients, u_i is a random spatial effect corresponding to each area i and ϵ_i represents unstructured random errors, which can be included or not in the linear predictor depending on the distribucion of Y_i. In case they are included, they follow independent normal distributions $N(0, 1/\tau_e)$. In particular, to fully specify the DAGAR model, it is assumed that $\boldsymbol{u} = (u_1, u_2, \ldots, u_k)^T|\rho, \tau_u \sim N(0, \boldsymbol{Q}^{-1}(\rho)/\tau_u)$, where $\boldsymbol{Q}(\rho) = [(\boldsymbol{I}-\boldsymbol{B})^T \boldsymbol{F}(\boldsymbol{I}-\boldsymbol{B})]^{-1}$ is the DAGAR precision matrix, where $\boldsymbol{F} = \text{diag}\,(\tau_1, \tau_2, \ldots, \tau_n)$ and $\boldsymbol{B} = \{b_{ij}\}$, for

$$b_{ij} = \frac{\rho}{1 + (n_{<i})\rho^2}; i = 2, \cdots, n; j \in N(i);$$

$$\tau_i = \frac{1 + (n_{<i} - 1)\rho^2}{1 - \rho^2}, i = 1, \cdots, n;$$

where ρ is the spatial autocorrelation parameter, and $n_{<i}$ denotes the number of neighbors of an area i. This DAGAR model fits into the class of hierarchical areal models, therefore, Bayesian inference is suitable for this class of models.

Since hierarchical models can also be represented as latent Gaussian models (LGM), the general DAGAR model can be represented as a LGM as follows:

(i) In the third level, let define the set of hyperparameters $\boldsymbol{\theta} = (\theta_1, \tau_u, \rho, \tau_e)$:

$$\pi(\boldsymbol{\theta}) = \pi(\theta_1, \rho, \tau_u, \tau_e),$$

where θ_1 representes scale parameters of Y_i; τ_u, ρ are spatial parameters; and τ_e is the precision parameter of random errors. Furthermore, θ might have some of these parameters depending on the second and first level of the LGM.

(ii) In the second level, let define the latent Gaussian field $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{\beta})$:

$$\boldsymbol{x} \mid \boldsymbol{\theta} \sim N(\boldsymbol{u}|0, \boldsymbol{Q}(\rho)^{-1}/\tau_u)N(\boldsymbol{\beta}).$$

(iii) In the first level, let assume that Y_i's are conditionally independent, given $\boldsymbol{x}$ and $\boldsymbol{\theta}$. Then

$$\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \pi(y_i|\boldsymbol{x}, \boldsymbol{\theta}).$$

Some DAGAR models depending on the distribution of Y_i are:

- Gaussian distribution: $Y_i|\boldsymbol{x}, \boldsymbol{\theta} \overset{ind}{\sim} N(\mu_i, \theta_1)$, for $i = 1, \cdots, n$, where

$$\mu_i = z_i^T \boldsymbol{\beta} + u_i + \epsilon_i.$$

The latent field is $x = (u, \beta)$ and the vector of hyperparameters is $\boldsymbol{\theta} = (\tau_u, \rho, \tau_e)$.

- Poisson distribution: $Y_i|x, \theta \sim Poisson(\lambda_i)$, for $i = 1, \cdots, n$, and from a spatial generalized linear mixed model framework $h(\lambda_i) = z_i^T \boldsymbol{\beta} + u_i$, for instance,

$$\log(\lambda_i) = z_i^T \boldsymbol{\beta} + u_i.$$

The latent field is $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{\beta})$ and the vector of hyperparameters is $\boldsymbol{\theta} = (\tau_u, \rho)$.

- Let $Y_i|x, \theta \sim$ Binomial(p_i), for $i = 1, \cdots, n$, where from a spatial generalized linear model framework $h(p_i) = z_i^T \boldsymbol{\beta} + u_i$, for instance,

$$\text{logit}(p_i) = z_i^T \boldsymbol{\beta} + u_i.$$

The latent field is $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{\beta})$ and the vector of hyperparameters is $\boldsymbol{\theta} = (\tau_u, \rho)$.

- Let $Y_i|\boldsymbol{x}, \boldsymbol{\theta} \sim Gamma(\mu_i, \kappa)$, for $i = 1, \cdots, n$, where from a spatial generalized linear model framework $h(\mu_i) = z_i^T \boldsymbol{\beta} + u_i$, for instance, $\log(\mu_i) = z_i^T \boldsymbol{\beta} + u_i$. The latent field is $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{\beta})$ and $\theta_1 = \kappa$ then the vector of hyperparameters is $\boldsymbol{\theta} = (\tau_u, \rho, \kappa)$.

With regard to the prior distributions for $\boldsymbol{\theta}$, in the absence of information, the typical choice are non-informative prior distributions. Thus for $\boldsymbol{\beta}$ is assumed a normal distribution with mean zero and a big variance, that is $\beta_p \sim N(0, 10^6), p = 1, \ldots, P$. The prior distribution for τ_u, τ_e and θ_1 is a gamma(1, 0.0005). And finally, a uniform distribution $U(0, 1)$ for ρ.

Then the joint posterior distribution for the general DAGAR model is:

$$\begin{aligned}\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}),\\ &\propto \pi(\theta_1)\pi(\tau_u)\pi(\rho)\pi(\tau_e)\pi(\theta_1)\pi(u|0,Q(\rho)^{-1}/\tau_u)\pi(\beta)\prod_{i=1}^{n}\pi(y_i|x,\theta).\end{aligned}$$

The posterior estimations are not easy to compute, and that is the main aim of INLA, providing approximations to the posterior marginals of the latent variables and hyperparameters. In addition, since $\boldsymbol{u}$ is a GMRF and the precision matrix $\boldsymbol{Q}(\rho)$ is sparse, INLA can exploit these features to achieve fast Bayesian inference.

As indicated in section 2.2.2 the INLA approach was adopted, to obtain the marginal distributions of $\boldsymbol{x},\boldsymbol{\theta}$. The marginal distribution of θ given x is approximated as follows:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}|_{x=x^*(\boldsymbol{\theta})} = \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}). \tag{3.5}$$

where $\tilde{\pi}$ is the Laplace approximation, $x^*(\theta)$ is the mode of $\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}$. Similarly, the Laplace approximation is computed as follows:

$$\pi(x_j|\boldsymbol{\theta},\boldsymbol{y}) \approx \frac{\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}{\tilde{\pi}(x_{-j}|x_j,\boldsymbol{\theta},\boldsymbol{y})}|_{x_{-j}=x^*_{(x_j,\boldsymbol{\theta})}} = \tilde{\pi}(x_j|\boldsymbol{\theta},\boldsymbol{y}). \tag{3.6}$$

After computing $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ in Equation (3.5) and $\tilde{\pi}(x_j|\boldsymbol{\theta},\boldsymbol{y})$ in Equation (3.6), we use the equation (2.3), then

$$\tilde{\pi}(x_j|y) = \int \tilde{\pi}(x_j|\boldsymbol{\theta},\boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}|y)d\theta.$$

$$\tilde{\pi}(\theta_m|y) = \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\theta_{-m}.$$

To conclude this computing, we have to add that the DAGAR model with Laplace approximation can be solved by via the R-INLA-package. We will see the simulation in the next chapter.

Chapter 4

Simulation study

In this chapter, we implement Bayesian inference for DAGAR models using the INLA approach through the R-INLA package, which is available in www.r-inla.org. Simulations of areal data under different scenarios were performed, for small and big data, and for Gaussian distributions and non- Gaussian distributions. Then DAGAR, CAR and SAR models were fitted in order to show the main advantages of DAGAR models. The results in this chapter were produced on a computer with some requirements as well as Intel(R) Core(TM) i7 and 12GB RAM.

4.1 Bayesian Parameter estimation with the Gaussian distribution

In order to assess the performance of the DAGAR model for large data, two scenarios were compared throughout two samples sizes: $n = 100$ and $n = 900$. Since the DAGAR, CAR and SAR models are going to be compared, and in particular the interpretation of the parameter ρ is going to be assesed, instead of simulating the spatial random effect vector u from these models, it is simulated from a Gaussian process

$$\boldsymbol{u} \sim N(0, \boldsymbol{M}/\tau_u),$$

where $1/\tau_u$ is the spatial marginal variance and $M(\cdot)$ is a correlation function. Specifically, it is used the exponential correlation function

$$M(d_{ij}) = \exp(-\phi d(i,j)),$$

where $d(i,j)$ represents the Euclidean distance between the centroids of areas i and j and ϕ is a parameter related to the range. The values for simulating $\boldsymbol{u}$ were set by $\tau_u = 0.25$ and $\phi = -\log(\rho)$ for $\rho \in (0,1)$. The covariates z_1 and z_2 are generated from standard normal distributions. The value of the regression coefficients and precision parameter of the random error were set to $\beta_0 = 4, \beta_1 = 2, \beta_2 = -1$, and $\tau_e = 5$. Finally the response variable Y_i for $i = 1, \ldots, n$ is simulated for each scenario from a normal distribution with mean

$$\mu_i = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + u_i$$

and variance $1/\tau_e$.

In order to fit the areal data models, INLA requires the form of the precision matrix, the key to fit the DAGAR structure. In particular, Fig. 4.1 shows the sparse structure of

the precision matrix of a DAGAR model, for a small $n = 100$ (left panel) and large dataset $n = 900$ (right panel). All neighbor areas (i, j) have values different from zero (grey squares) if $j \sim N(i)$.

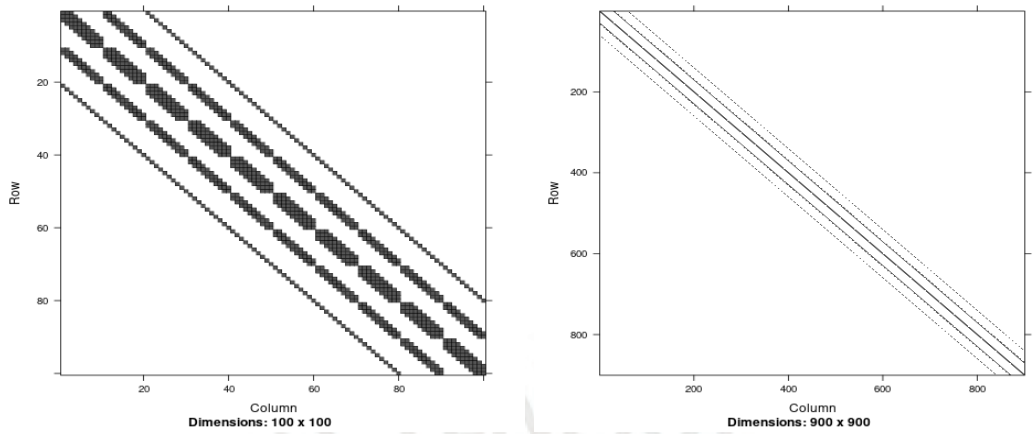


Figure 4.1: Precision matrices for $n = 100$ (left panel) and $n = 900$ (right panel).

4.1.1 Comparison between INLA and MCMC

The main advantage of performing Bayesian inference for DAGAR models through INLA is in terms of time requirements. In order to show this feature, small areal data (n=100) and large areal data (n = 900) were simulated as explained in the previous section, with $\rho = 0.5$.

The DAGAR model was fitted through INLA and MCMC for small and large areal data. The code to fit INLA for DAGAR model was written as part of the contribution of this thesis. MCMC was fitted using the `NIMBLE package` in R. Posterior inference of MCMC was based upon one chain of 10000 iterations (with a burn-in of 5001 iterations). The results of this experiment are shown in Table 4.1. In fact, the execution time for MCMC is very expensive compared to INLA. From these results, it is evident that fitting replications using MCMC would take a long time, for this reason the next simulation studies were performed through INLA.

Method	n	Time (sec)
MCMC	100	350.4
INLA	100	20.5
MCMC	900	9840.4
INLA	900	99.7

Table 4.1: Running times for DAGAR models using MCMC and INLA for small areal data (n=100) and large areal data (n=900).

4.1.2 Bayesian inference using INLA

The values of ρ for simulating u were set as $\rho = j/10$, for $j = 1, \ldots, 9$, that is considering nine subscenarios. Assuming $\phi = -\log(\rho)$ in this way implies that the average neighbor pair correlation ρ varies between 0.1 and 0.9. It is conducted 100 replications for the eighteen

scenarios (according to $n = 100$ and $n = 900$), thus covering a wide spectrum of scenarios for areal data.

Then, the DAGAR, CAR and SAR models were fitted throughout INLA. Bayesian inference for CAR and SAR models is essentially the same as the method described for the DAGAR model (see Section 3.4) but for specific precision matrices. The precision matrices of these models are summarized in Table 4.2, where b_{ij} are elements of $\boldsymbol{B}$ and $\boldsymbol{F}$ is the diagonal matrix with elements $\tau_1, \ldots, \tau_n$, where b_{ij} and τ_i are defined in terms of ρ.

Table 4.2: Areal data Models

Models	Precision matrix
DAGAR	$\tau_u[(\boldsymbol{I}-\boldsymbol{B})^T\boldsymbol{F}(\boldsymbol{I}-\boldsymbol{B})]^{-1}$
CAR	$\tau_u(\boldsymbol{I}-\rho\widetilde{\boldsymbol{W}})$
SAR	$\tau_u(\boldsymbol{I}-\rho\widetilde{\boldsymbol{W}})(\boldsymbol{I}-\rho\widetilde{\boldsymbol{W}})^\top$

Figure 4.2 shows the mean of the 100 median posterior estimates and the 95% credible bands intervals for ρ under the three models(CAR, SAR and DAGAR) and two scenarios, for small areal data, $n = 100$ areas (left panel) and large areal data, $n = 900$ areas (right panel). For both scenarios, estimates for ρ from the CAR model (red lines) are considerably higher than the true value, specially worst for large areal data. For small areal data the bias is higher when the true ρ is small, but at least the 95% confidence bands (red) cover the true ρ values. For large areal data the bias is also higher when the true ρ is small, but in this case the 95% confidence bands (red) do not cover the true ρ values.

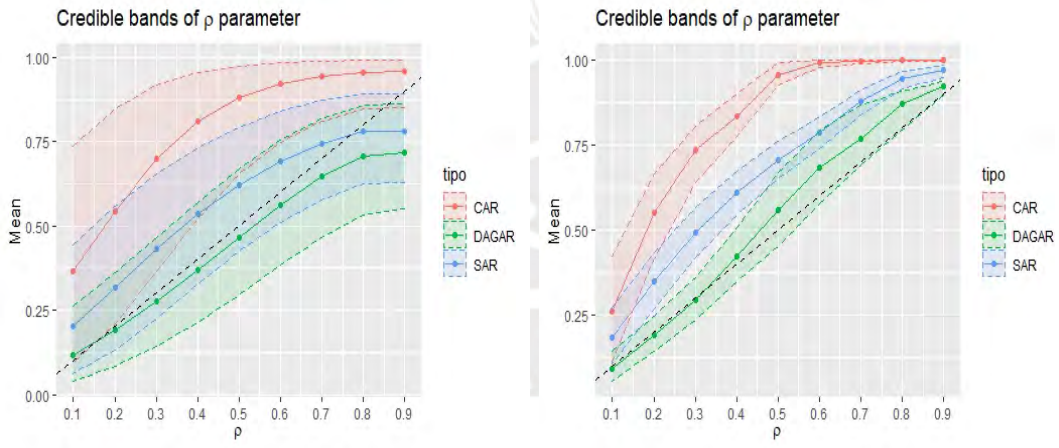


Figure 4.2: Posterior estimate and confidence bands of ρ as a function of the true ρ (x-axis) for 100 replicated simulations of small (left column) and large (right column) datasets with Gaussian distribution.

The SAR models generally perform better in this respect with less estimation bias (blue line), particularly for higher ρ and large areal data. The 95% confidence bands for the SAR models (blue bands) cover the true value of ρ for small areal data, while the bands clearly miss all the true ρ values for large areal data. Finally, the DAGAR models generally perform much better with much less estimation bias (green line), particularly for small ρ. The 95% confidence bands of the DAGAR models (green bands) always cover the true value of ρ when

the data is small. And for large areal data the 95% confidence bands of the DAGAR models (green bands) always cover the true value of ρ.

Figure 4.3 shows the mean of median posterior estimates and credible bands for regression coefficients $\beta_0, \beta_1, \beta_2$ and the error variance $1/\tau_\epsilon$ over the 100 replications of the three models for small areal data. The 95% confidence bands for all models cover the true value of the parameters. Since the spatial marginal precision τ_u has different interpretation for CAR, SAR and DAGAR models, the estimations for this parameter should not be compared.

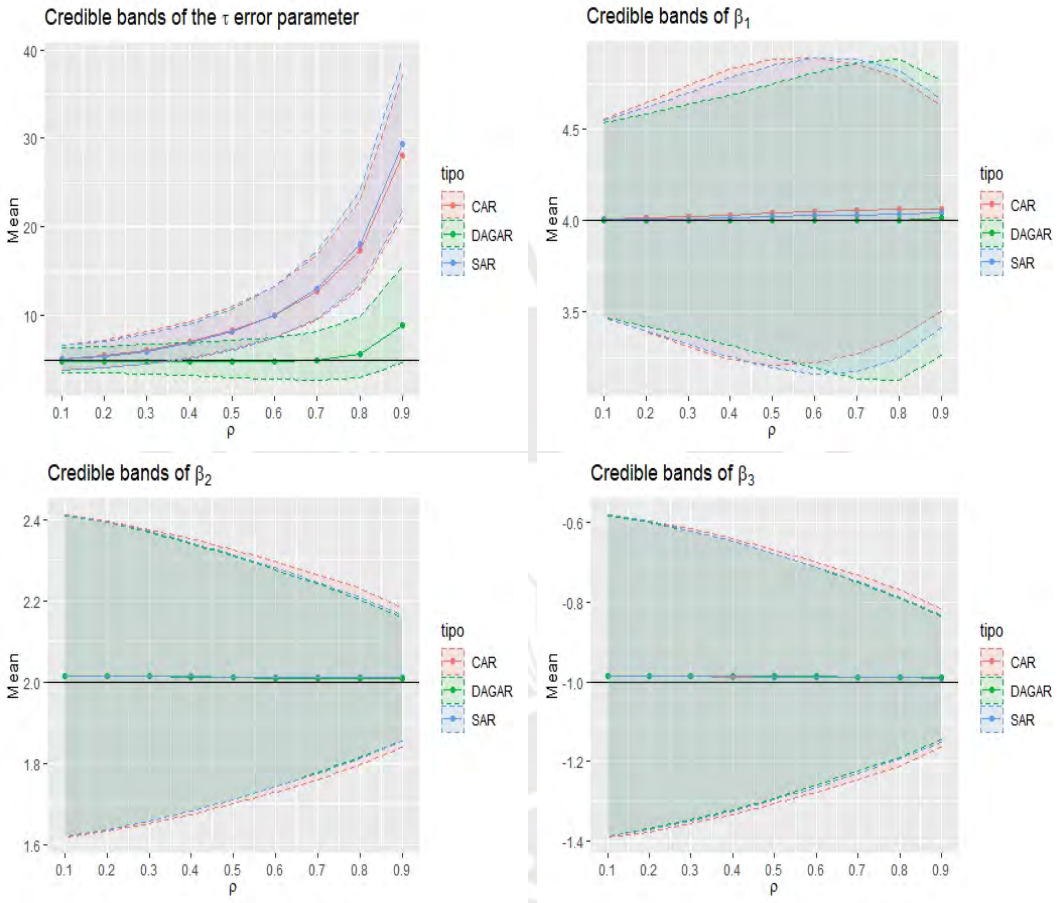


Figure 4.3: Credible bands of the parameters as a function of the true ρ (x-axis) for 100 replicated simulations of small areal data with Gaussian distribution.

Similar plots for large areal data are shown in Figure 4.4. All the models shown that the bias is lower for large areal data than for small areal data . It can also be seen that the credible bands for the 95% confidence bands for all models cover the true value of the parameters. And for the regression coefficients, the higher the ρ parameter, the smaller the credible bands of all models.

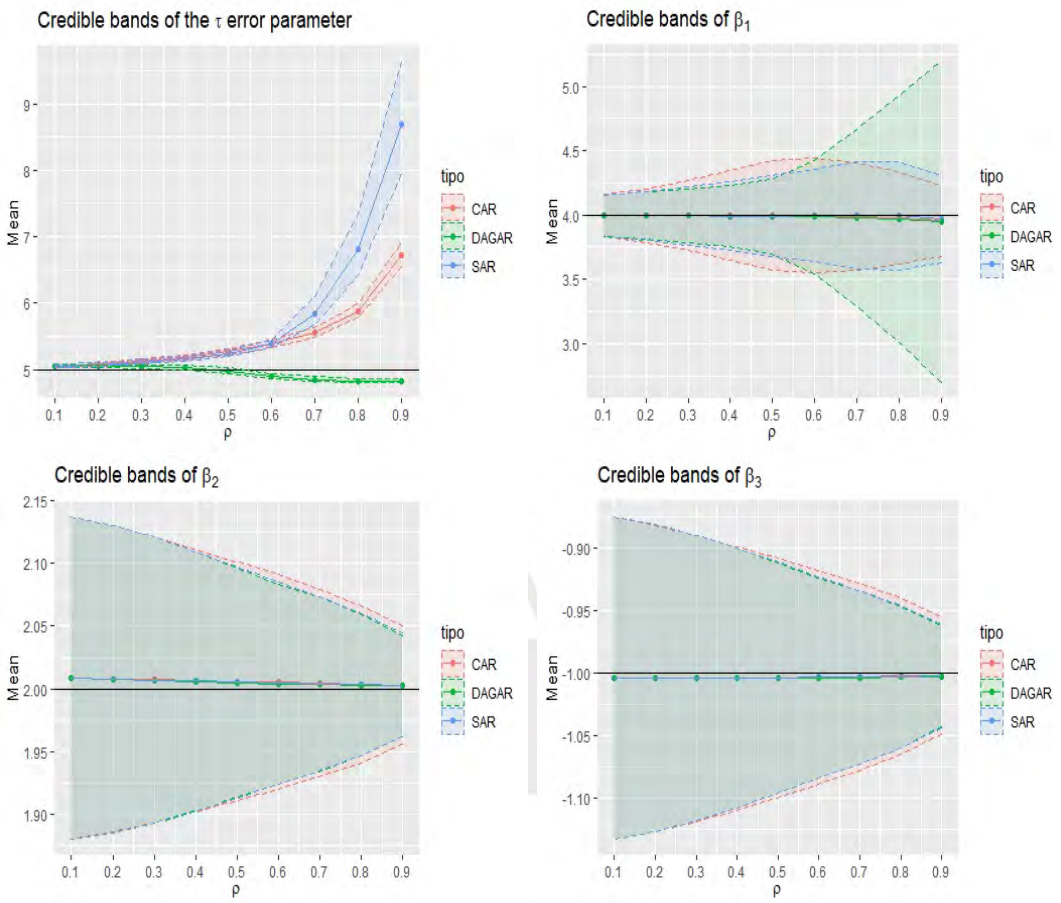


Figure 4.4: Credible bands of the parameters as a function of the true ρ (x-axis) for 100 replicated simulations of large areal data with Gaussian distribution.

4.2 Bayesian parameter estimation with non-Gaussian likelihood

Similarly to the previous section, two scenarios were compared throughout two samples sizes: $n = 100$ and $n = 900$.The spatial random effect vector u from these models, it is simulated from a Gaussian process $\boldsymbol{u} \sim N(0, M/\tau_u)$, where $1/\tau_u$ is the spatial marginal variance and $M(.)$ is a correlation function. Specifically, it is used the exponential correlation function $M(d_{ij}) = \exp(-\phi d(i,j))$, where $d(i,j)$ represents the Euclidean distance between the centroids of areas i and j and ϕ is a parameter related to the range. The values for simulating $\boldsymbol{u}$ were set as follows, $\tau_u = 0.25$ and $\phi = -\log(\rho)$ for $\rho = j/10$, for $j = 1, \ldots, 9$. Specifically, assuming ϕ in this way implies that the average neighbor pair correlation ρ varies between 0.1 and 0.9. The covariates z_1 and z_2 are generated from standard normal distributions. The value of the regression coefficients are set as $\beta_0 = 4, \beta_1 = 2$, and $\beta_2 = 2$. Simulation of the response variable is detailed in the next subsections.

4.2.1 DAGAR models with Poisson distribution

The response variable Y_i for $i = 1, \ldots, n$ is simulated for each scenario from a Poisson distribution with mean

$$\lambda_i = \exp(\beta_0 + \beta_1 z_1 + \beta_2 z_2 + u_i).$$

It is conducted 100 replications for the eighteen scenarios (nine scenarios for different ρ values and two scenarios for $n = 100$ and $n = 900$). Then, the DAGAR, CAR and SAR models were fitted throughout INLA.

Figure 4.5 shows the mean of the 100 median posterior estimates and 95% credible bands intervals of ρ under the three models(CAR, SAR and DAGAR) and two scenarios, for small areal data, $n = 100$ areas (left panel) and large areal data, $n = 900$ areas (right panel). For both scenarios, estimates for ρ from the CAR model (red lines) are higher than the true value. For small and large areal data the 95% confidence bands of CAR models (red bands) do not cover the true ρ values. The SAR models generally perform better in this respect with less estimation bias (blue line), particularly for higher ρ. The 95% confidence bands for the SAR models (blue bands) cover the true value of ρ for small and large areal data. Finally, the DAGAR models generally perform much better than CAR and SAR models, with much less estimation bias (green line) for all ρ values. The 95% confidence bands of the DAGAR models (green bands) always cover the true value of ρ when the data is small or large.

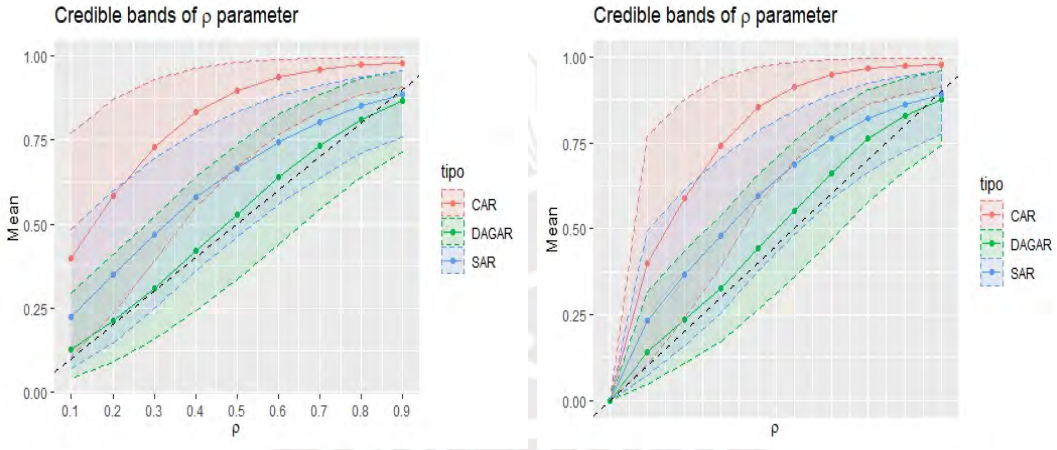


Figure 4.5: Estimate and credible intervals of ρ as a function of the true ρ (x-axis) for the simulation small and large datasets with Poisson distribution.

Figure 4.6 shows the mean of median posterior estimates and credible bands for regression coefficients $\beta_0, \beta_1, \beta_2$ over the 100 replications of the three models for small areal data (left column) and large areal data (right column). For small and large areal data, the 95% confidence bands for all models cover the true value of the parameters. In general, the confidence bands of all models are smaller as the value of ρ increases. And in some cases, the 95% confidence bands are a little bit smaller for SAR and DAGAR models. The spatial marginal precision τ_u has different interpretation for CAR, SAR and DAGAR models, therefore the estimations for this parameter are not compared.

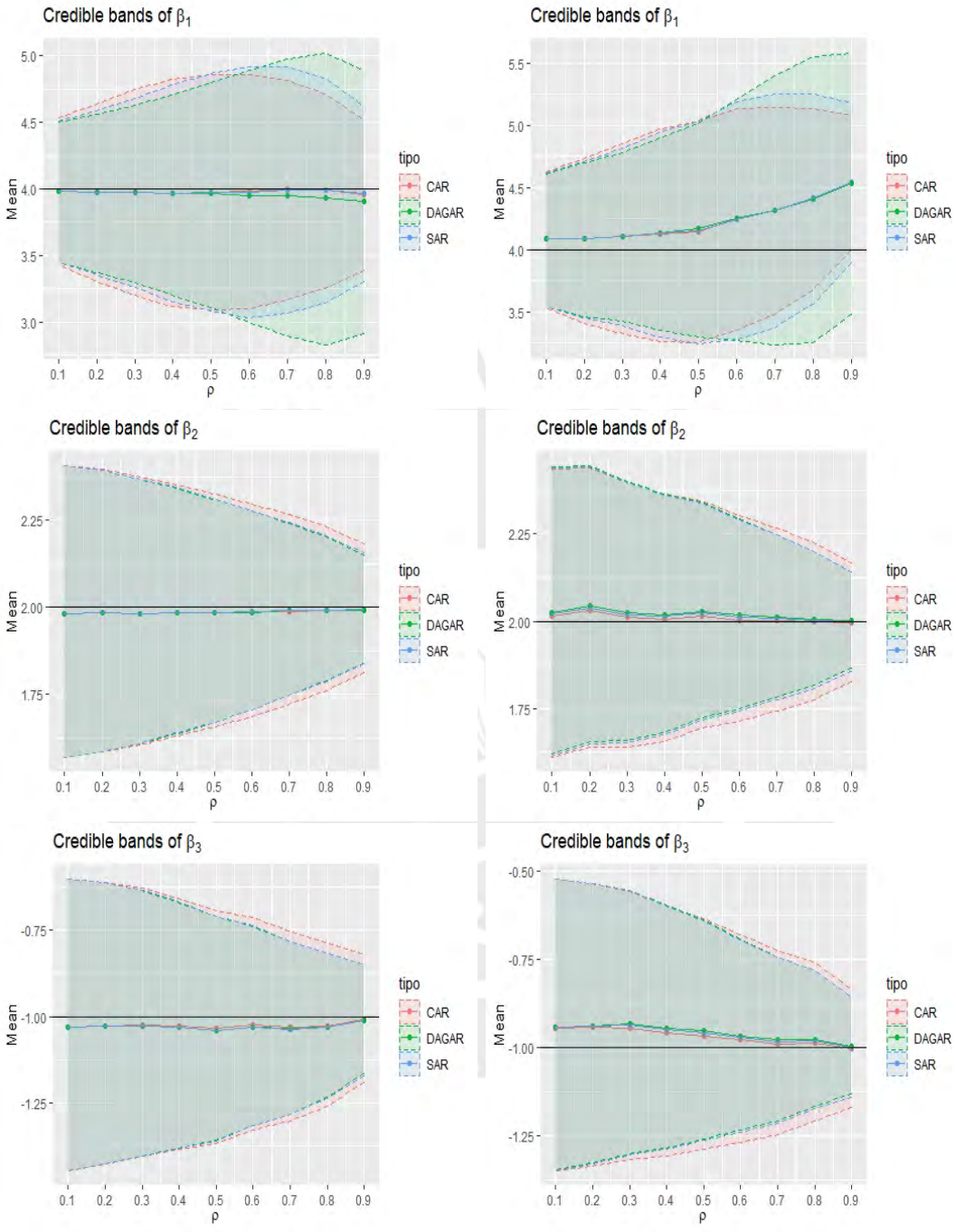


Figure 4.6: Credible bands of the parameters as a function of the true ρ (x-axis) for 100 replicated simulations for small areal data (left column) and large areal data (right column) with Poisson distribution.

4.2.2 Parameter estimation with the Binomial distribution

The response variable Y_i for $i = 1, \ldots, n$ is simulated for each scenario from a Binomial distribution $Bin(p_i, m_i)$, where p_i is sampled from:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + u_i,$$

and m_i is sampled from $\{1, 2, 3, \ldots, 100\}$. It is conducted 100 replications for the eighteen scenarios (nine scenarios for different ρ values and two scenarios for $n = 100$ and $n = 900$). Then, the DAGAR, CAR and SAR models were fitted throughout INLA.

Figure 4.7 shows the mean of the 100 median posterior estimates and 95% credible bands intervals of ρ under the three models(CAR, SAR and DAGAR) and two scenarios, for small areal data, $n = 100$ areas (left panel) and large areal data, $n = 900$ areas (right panel). For both scenarios, estimates for ρ from the CAR model (red lines) are higher than the true value. For small and large areal data the 95% confidence bands of CAR models (red bands) slightly cover the true ρ values. The SAR models generally perform better in this respect with less estimation bias (blue line), particularly for higher ρ. The 95% confidence bands for the SAR models (blue bands) cover the true value of ρ for small and large areal data. Finally, the DAGAR models generally perform much better than CAR and SAR models, with much less estimation bias (green line) for all ρ values. The 95% confidence bands of the DAGAR models (green bands) always cover the true value of ρ when the data is small or large.

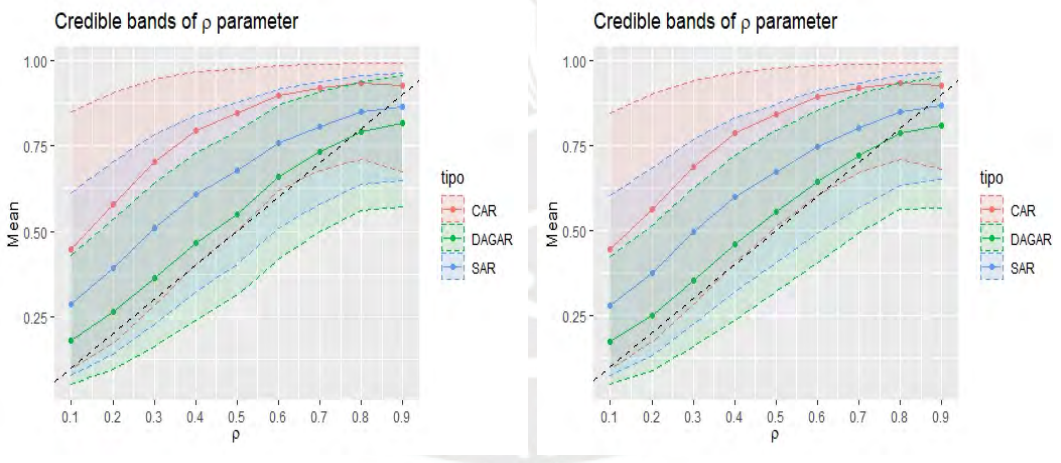


Figure 4.7: Estimate and credible bands of ρ as a function of the true ρ (x-axis) for 100 replicated simulations of small (left) and large (right) datasets with Binomial distribution.

Figure 4.10 shows the mean of median posterior estimates and credible bands for regression coefficients $\beta_0, \beta_1, \beta_2$ over the 100 replications of the three models for small areal data (left column) and large areal data (right column). For small and large areal data, the 95% confidence bands for all models cover the true value of the parameters. In general, the confidence bands of all models are smaller as the value of ρ increases. And in some cases, the 95% confidence bands are a little bit smaller for SAR and DAGAR models. The spatial marginal precision τ_u has different interpretation for CAR, SAR and DAGAR models, therefore the estimations for this parameter are not compared.

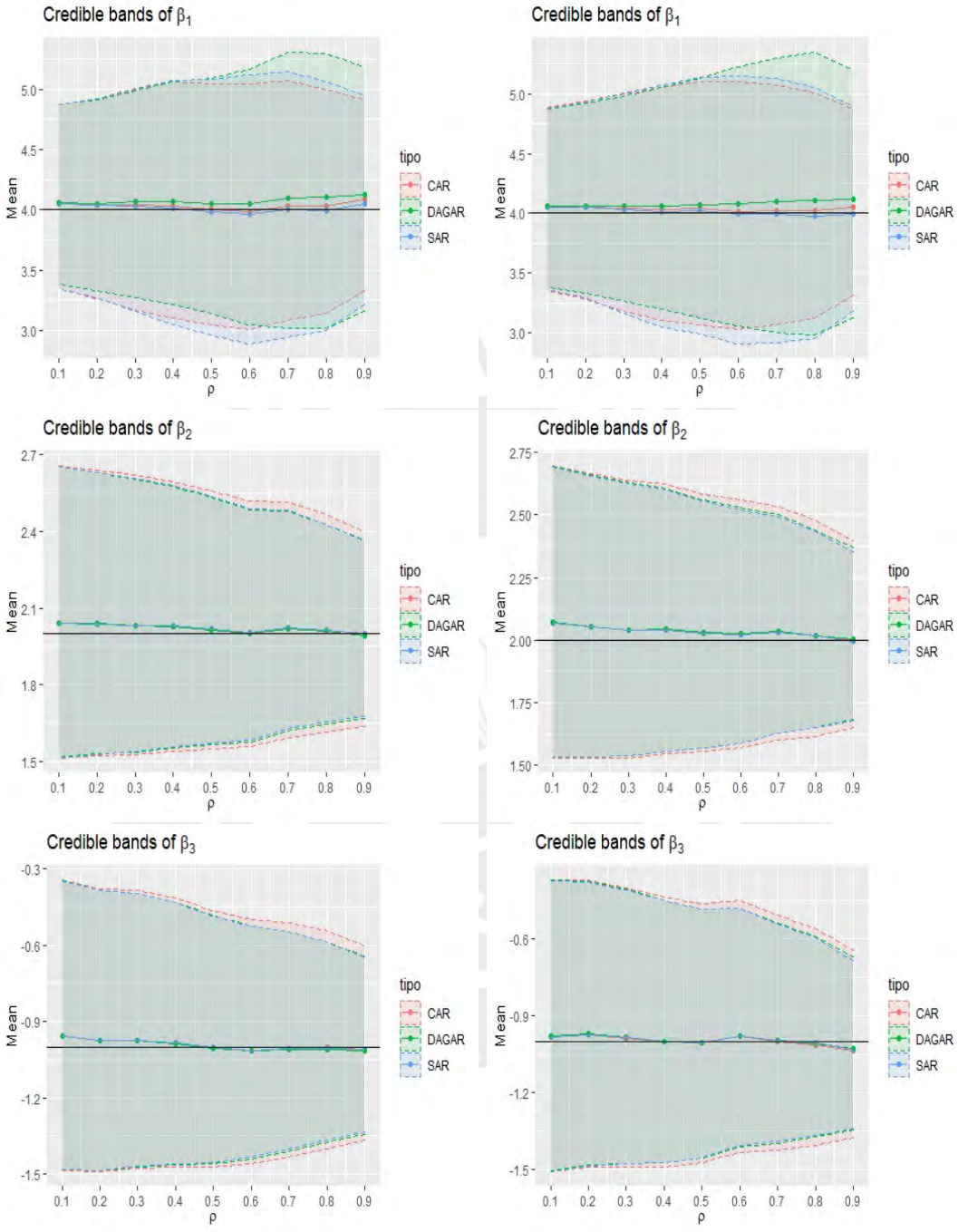


Figure 4.8: Credible bands of the parameters as a function of the true ρ (x-axis) for 100 replicated simulations of small (left) and large (right) areal data with Binomial distribution.

4.2.3 Parameter estimation with the Gamma distribution

The response variable Y_i for $i = 1, \ldots, n$ is simulated for each scenario from a Gamma distribution $Gamma(\mu_i, \kappa_i)$, where μ_i is sampled from:

$$\mu_i = \exp(\beta_0 + \beta_1 z_1 + \beta_2 z_2 + u_i).$$

It is conducted 100 replications for the eighteen scenarios (nine scenarios for different ρ values and two scenarios for $n = 100$ and $n = 900$). Then, the DAGAR, CAR and SAR models were fitted throughout INLA.

Figure 4.7 shows the mean of the 100 median posterior estimates and 95% credible bands intervals of ρ under the three models(CAR, SAR and DAGAR) and two scenarios, for small areal data, $n = 100$ areas (left panel) and large areal data, $n = 900$ areas (right panel). For both scenarios, estimates for ρ from the CAR model (red lines) are higher than the true value. For small and large areal data the 95% confidence bands of CAR models (red bands) slightly cover the true ρ values. The SAR models generally perform better in this respect with less estimation bias (blue line), particularly for higher ρ. The 95% confidence bands for the SAR models (blue bands) cover the true value of ρ for small and large areal data. Finally, the DAGAR models generally perform much better than CAR and SAR models, with much less estimation bias (green line) for all ρ values. The 95% confidence bands of the DAGAR models (green bands) always cover the true value of ρ when the data is small or large.

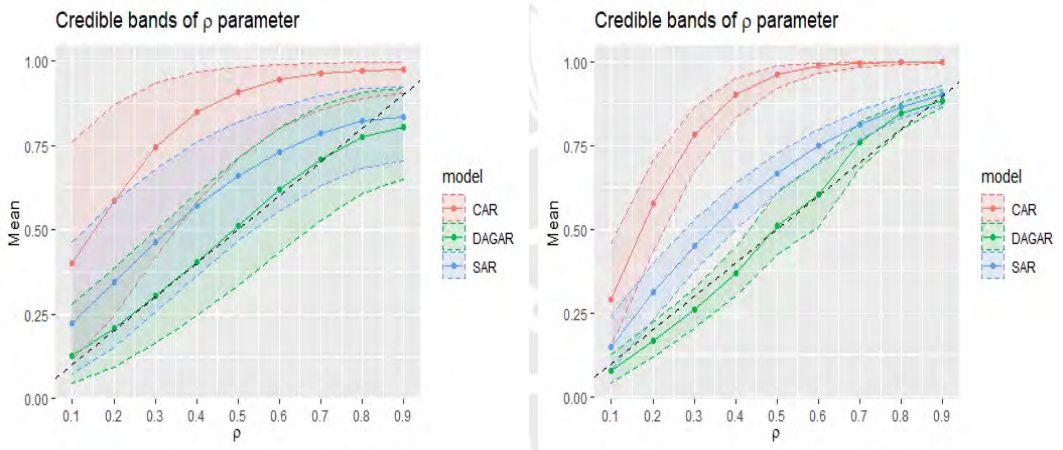


Figure 4.9: Estimate and credible intervals of ρ as a function of the true ρ (x-axis) for the simulation small and large datasets.

Figure 4.10 shows the mean of median posterior estimates and credible bands for regression coefficients $\beta_0, \beta_1, \beta_2$ over the 100 replications of the three models for small areal data (left column) and large areal data (right column). For small and large areal data, the 95% confidence bands for all models cover the true value of the parameters. In general, the confidence bands of all models are smaller as the value of ρ increases. And in some cases, the 95% confidence bands are a little bit smaller for SAR and DAGAR models. The spatial marginal precision τ_u has different interpretation for CAR, SAR and DAGAR models, therefore the estimations for this parameter are not compared.

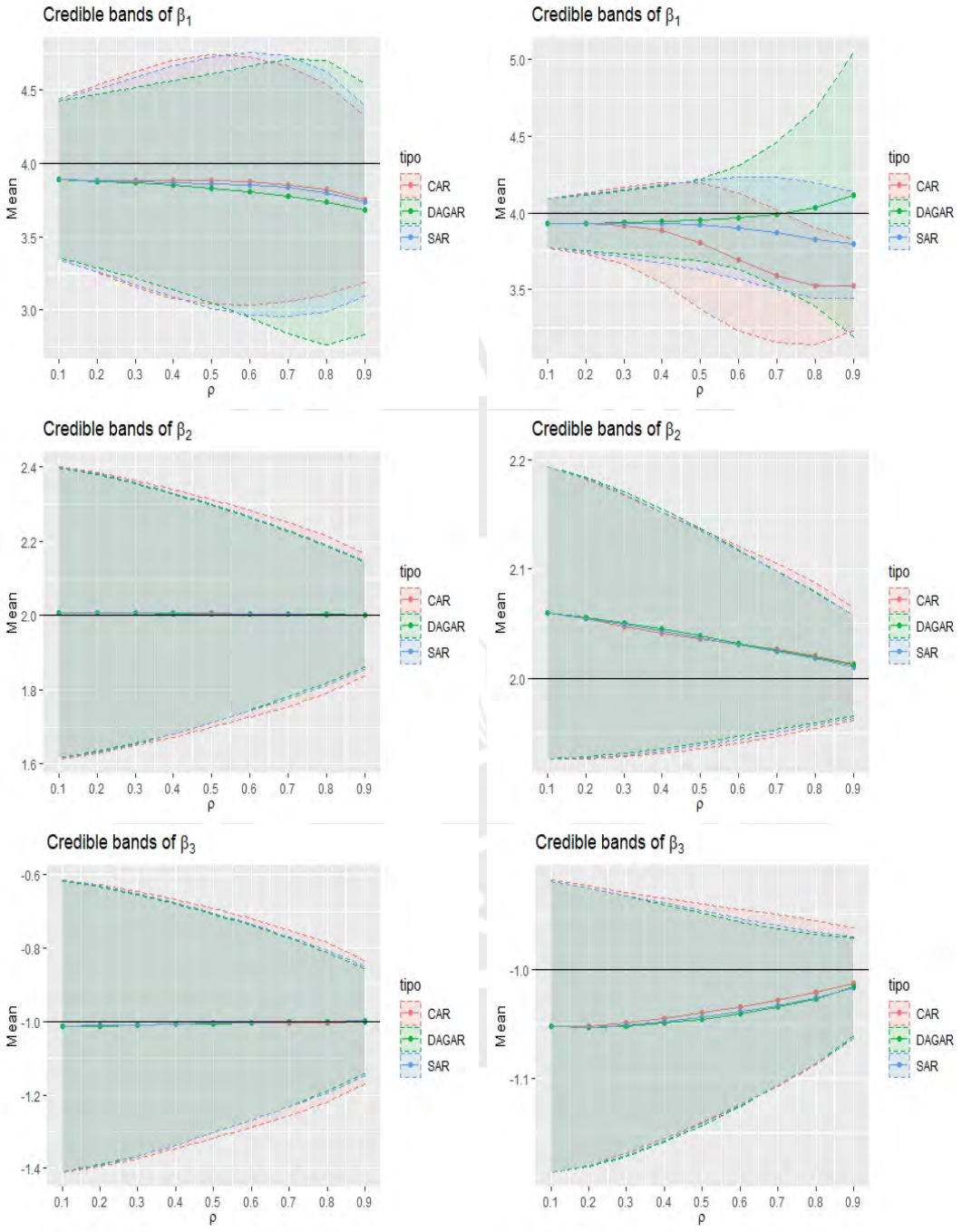


Figure 4.10: Credible intervals of the parameters as a function of the true ρ (x-axis) for the simulation small areal data analysis with Gamma distribution.

Chapter 5

Applications

5.1 Application 1: Sudden infant death syndrome in North Carolina

About three thousand of babies in the United States unexpectedly die each year. Sudden unexpected infant deaths include sudden infant death syndrome (SIDS), that is, deaths from unknown causes. In this chapter, it is analyzed SIDS data in North Carolina (NC). The data is available in the `spData` package in R. The data consist of counts of live births, counts of deaths and counts of SIDS in $n = 100$ counties of NC from 1974 to 1978. In particular, the statistical analysis presented in this section has as main goals: (i) to adapt to the characteristics of the sampled data; (ii) to provide important insights on the spatial distribution of SIDS; and (iii) to propose a model for estimation of SIDS in the NC.

Figure 5.1 shows the map of the proportion of non-white births, which is used as a covariate. Figure 5.2 shows the map of the counts of SIDS.

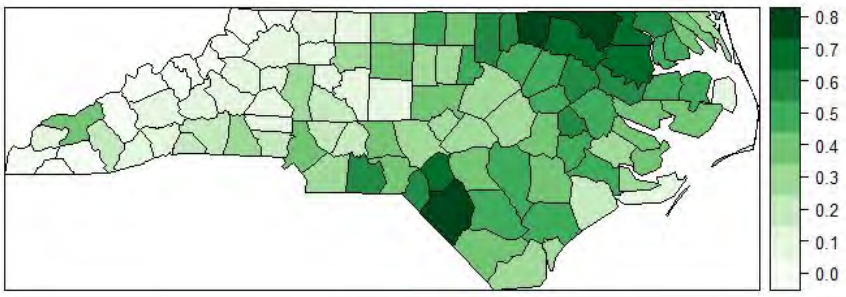


Figure 5.1: Proportion of non-white births in North Carolina dataset.

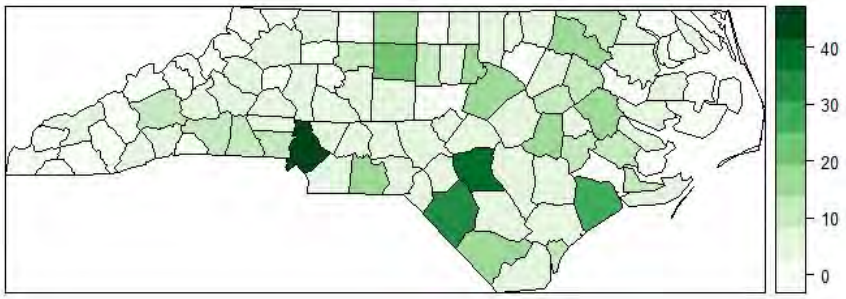


Figure 5.2: Counts of SIDS in North Carolina dataset.

Through the counties of North Carolina and their location, it is feasible to construct

a graph considering a neighbor county whether both counties share common border lines (Figure 5.3). The adjacency matrix for this graph is shown in Figure 5.4. The black squares represent that two counties are neighbors, while the green squares represent that two counties are not neighbors. This adjacency matrix is used to fit the CAR model.

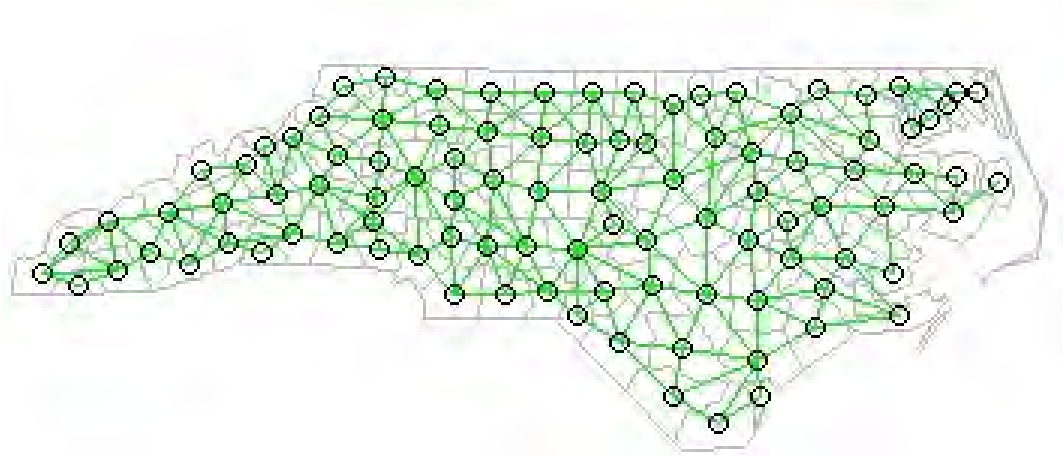


Figure 5.3: Graph representation of counties map in North Carolina

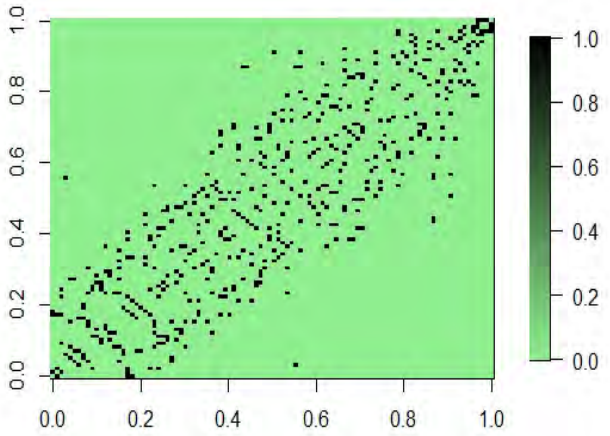


Figure 5.4: Adjacency Matrix of counties in North Carolina

Let Y_i be a random variable that represents the counts of SIDS for an area $i = 1, \ldots, 100$. Given a latent field $\boldsymbol{x}$ and hyperparameters $\boldsymbol{\theta}$ as, it is assumed that Y_i are conditional independent, that is, $Y_i|\boldsymbol{x}, \boldsymbol{\theta} \overset{ind}{\sim} \text{Poisson}(\lambda_i)$, where the mean of SIDS is λ_i and

$$\log(\lambda_i) = \beta_0 + \beta_1 z_1 + u_i,$$

where λ_i is the expected counts of SIDS at area i, β_0 and β_1 denote the regression coefficients, z_i is the proportion of non-white births in the county i and u_i represents the spatial random effect of the county $i = 1, 2, \ldots, n$. The specific definition of $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^T$ depends on

the model fitted. In particular, the CAR-Poisson and DAGAR-Poisson models are proposed to fit these data. Then the CAR-Poisson and DAGAR-Poisson models can be defined by the latent Gaussian field $\boldsymbol{x} = \{\boldsymbol{\beta}, \boldsymbol{u}\}$ and hyperparameters $\boldsymbol{\theta} = (\tau_u, \rho)$.

Bayesian inference was carried out through INLA for Poisson-DAGAR and Poisson-CAR models. In order to compare the time requirements of MCMC for DAGAR models, it was also implemented for the Poisson-DAGAR model. Posterior inference of MCMC was based upon one chain of 10000 iterations (with a burn-in of 5001 iterations). MCMC for Poisson-DAGAR model takes 200.2894 seconds. Using INLA the Poisson-DAGAR model runs in a lower time.

Table 5.1 presents the selection criteria of the fitted models. The WAIC and LPML values indicate that the Poisson-DAGAR model have a better goodness of fit than the Poisson-CAR model. Further, RMSEE is also in favor of Poisson-DAGAR model. Based on these considerations, we conclude that the Poisson-DAGAR model presents the best goodness of fit to these SIDS data.

	Poisson-DAGAR INLA	Poisson-CAR INLA
WAIC	**472.6759**	475.5959
LPML	**-4.0715**	-4.0355
RMSEE	**0.9072**	0.9568
Time (sec)	**16.2**	22.6

Table 5.1: The selection criteria for the models proposed, total run time in seconds and RMSEE.

The posterior parameters estimates of Poisson-CAR and Poisson-DAGAR models are reported in Table 5.2. First it should be pointed out the different estimates and confidence intervals of ρ from both models, although the Poisson-CAR model suggest strong spatial correlation, the Poisson-DAGAR model suggest moderate spatial correlation. It have seen consistently from the simulation study that when the underlying spatial correlation is moderate, the DAGAR model performs better than the CAR models. This result is confirmed through the criteria assessment that choose the Poisson-DAGAR model as the preferred one for fitting. All the regression coefficient credible intervals do not cover zero. Specifically, for the Poisson-DAGAR model the posterior mean of the proportion of non-white births (0.3) indicates that a greater count of births implies higher counts of SIDS.

	Poisson-DAGAR INLA	Poisson-CAR INLA
β_0	0.7637 (0.1213, 1.4102)	0.3990 (-0.3128, 1.0234)
β_1	1.5047 (0.0116, 2.9469)	1.7619 (0.4371, 3.1117)
ρ	0.4046 (0.1866, 0.7132)	0.9370 (0.5617, 0.9937)
τ_u	0.8601 (0.4566, 1.4973)	1.3986 (0.9297, 2.0840)

Table 5.2: Summary statistics: posterior median and 95% credible interval for the hyperparameters for each model.

Figure 5.5 shows the posterior mean estimates of counts of SIDS in North Carolina from Poisson-DAGAR model (upper panel) and Poisson-CAR model (lower panel). After comparing Figure 5.2 and Figure 5.5, it is confirmed that the Poisson-DAGAR model estimates

quite good the counts of SIDS in North Carolina.

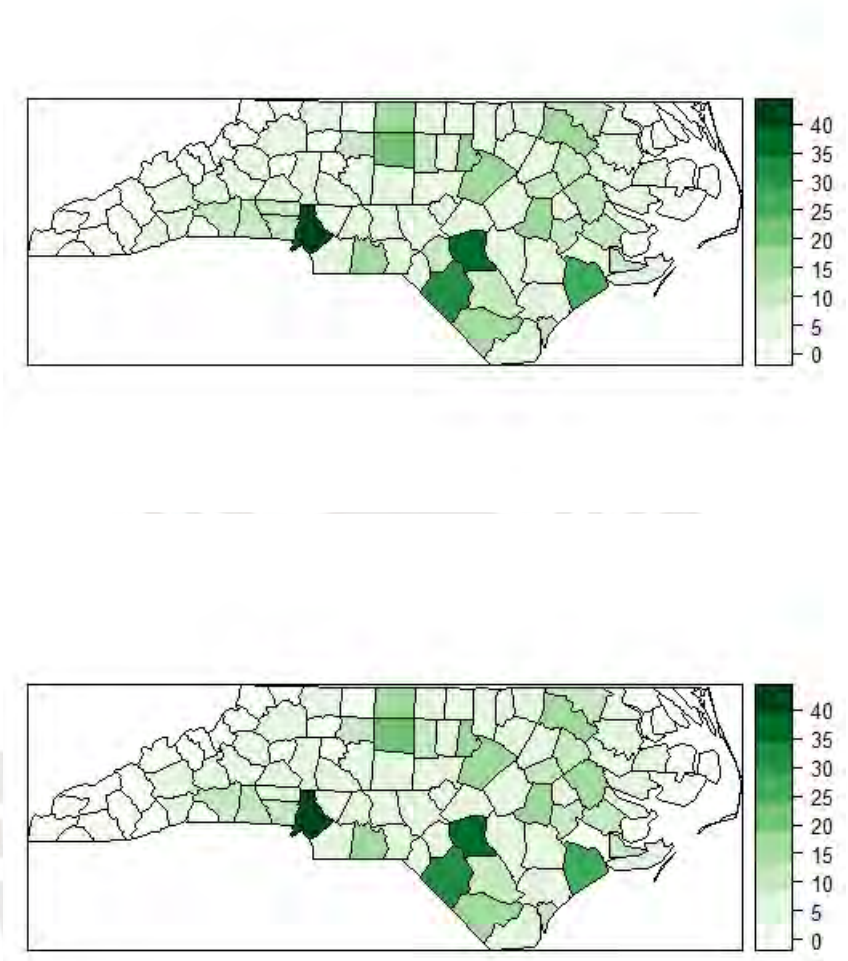


Figure 5.5: Posterior mean estimates of counts of SIDS from Poisson-DAGAR model (upper panel) and Poisson-CAR model (lower panel).

5.2 Application 2: Covid-19 in Peru

The pandemic of coronavirus disease 2019 (Covid-19) has spread to more than 150 countries around the world. Peru exceeded 1.4 millions reported Covid-19 cases before to 2021-02-09. The main goal of this application is to study the spatial distribution of counts of covid-19 in districts of Peru.

The data analyzed in this study are available at: https://www.datosabiertos.gob.pe/dataset/casos-positivos-por-covid-19-ministerio-de-salud-minsa and contains the cases of positive Covid-19 test result, either polymerase chain reaction (PCR) or antigen (PC), from 2020-03-06 to 2021-02-09. Peruvian socio-economic indicators are collected by the Instituto Nacional de Estadıstica e Informática (INEI), and are available in https://www.inei.gob.pe. All the variables in the data are summarized in Table 5.3.

Table 5.3: Summary of variables in the coronavirus dataset in Peru

Variable	Description
UUID	Person id number
Departament	Departament of the infected person.
Province	Province of the infected person.
District	District of the infected person.
Test	Type of test. It can be PCR or PC.
Date	Date reported.
TotPovRate	Rate of total poverty by district.
HouNoSan	Rate of Households without access to sanitation by district.
Severity	Index of severity by district.

Figure 5.6 shows the daily counts of covid-19 in Peru. It can be observed that the first wave begin in april 2020 and finish in october 2020, and the country experienced another wave of Covid-19 in january 2021. This reflects the limitations of the state to contain the spread of the virus.

In addition, Figure 5.7 shows a map of confirmed cases of Covid-19 in $n = 1845$ districts of Peru. It is observed that the districts in the coast has more confirmed cases of covid-19, followed by the the districts in the jungle. Further, this map shows some evidence of spatial autocorrelation between the counts of covid-19 in the districts. It means that the higher (lower) the confirmed cases of covid-19 in some district, the higher (lower) the confirmed cases of covid-19 in near districts. In fact the Moran I test (0.5257) and Geary C test (0.7813) proved this statement, since the null hypothesis of no evidence of spatial autocorrelation were rejected with $p - value < 2.2e - 16$ and $p - value = 0.003777$, respectively.

In order to fit areal data models, the adjacency matrix $\boldsymbol{W}$ for these data was built assuming that two districts are neighbors if they share somo geographical limit. The adjacency matrix for these data is shown in Figure 5.8, each black square represents that two districts are neighbors.

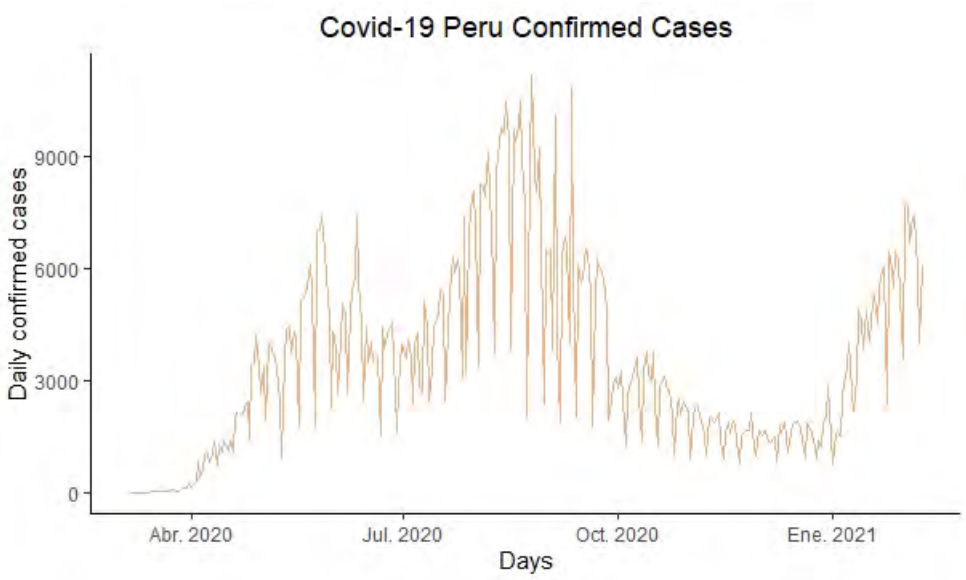


Figure 5.6: Number of daily confirmed cases of Covid-19 in Peru.

The Poisson-CAR, Poisson-DAGAR and zero-inflated Poisson DAGAR (ZIP-DAGAR) models are proposed to fit these data. These models are defined as follows:

- Poisson-CAR model: Let Y_i be a random variable that represents the counts of covid-19 in a district $i = 1, \ldots, 1845$. It is assumed that $Y_i|\boldsymbol{x}, \boldsymbol{\theta} \overset{ind}{\sim} \text{Poisson}(\lambda_i)$, where the mean of covid-19 is λ_i and

$$\log(\lambda_i) = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3 z_{3i} + u_i,$$

  where λ_i is the expected counts of covid-19 at district i, β_0, β_1, β_2 and β_3 denote regression coefficients, z_{1i}, z_{2i} and z_{3i} are the covariates rate of total poverty, rate of householdswithout access to sanitation and the index of sevirity for the i-th district and u_i represents the spatial random effect of the i-th district. The spatial structured random effect $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^\top$ follows a CAR structure. From these definitions, the latent Gaussian field is $\boldsymbol{x} = \{\boldsymbol{\beta}, \boldsymbol{u}\}$ and the vector of hyperparameters is $\boldsymbol{\theta} = (\tau_u, \rho)$.

- Poisson-DAGAR model: Let Y_i be a random variable that represents the counts of SIDS in a district $i = 1, \ldots, 1845$. It is assumed that $Y_i|\boldsymbol{x}, \boldsymbol{\theta} \overset{ind}{\sim} \text{Poisson}(\lambda_i)$, where the mean of SIDS is λ_i and

$$\log(\lambda_i) = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3 z_{3i} + u_i,$$

  where λ_i,β_0, β_1, β_2, β_3, z_{1i}, z_{2i} and z_{3i} are defined in the same way as for the Poisson-CAR-model. Moreover, u_i represents the spatial random effect of the district $i = 1, 2, \ldots, n$ and $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^T$ follows a DAGAR structure. From these definitions, the latent Gaussian field $x = \{\beta_0, u, \}$ and hyperparameters $\theta = (\tau_u, \rho)$.

- ZIP-DAGAR model: This model is proposed because in 15% of districts there were no confirmed cases of covid-19. Let Y_i be a random variable that represents the counts

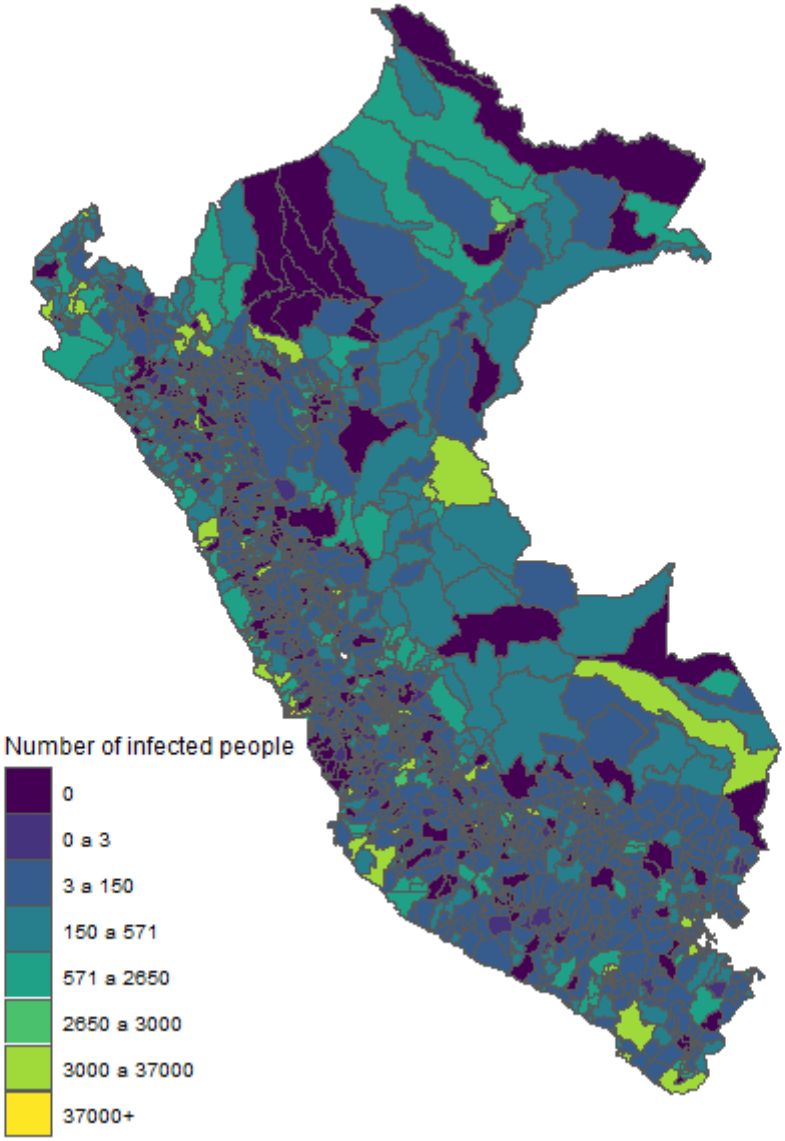


Figure 5.7: Number of confirmed cases of Covid-19 in each district of Peru.

of SIDS for an area $i = 1, \ldots, n$. It is assumed that $Y_i|x, \theta \overset{ind}{\sim} \text{ZIP}(p_i, \lambda_i)$, where p_i is the probability of zero counts of covid-19 is a district i and λ_i is the mean of SIDS of district i. The pdf mixture of Y_i is defined as:

$$\pi(y_i|p_i, \lambda_i) = p_i\delta_0 + (1 - p_i)h(y_i|\lambda_i)I[y_i > 0]$$

where $h(.)$ is the pdf of a Poisson distribution. Further,

$$\text{logit}(p_i) = \beta_0^{(1)} + \beta_1^{(1)}z_{1i} + \beta_2^{(1)}z_{2i} + \beta_3^{(1)}z_{3i} + u_i^{(1)},$$

$$\log(\lambda_i) = \beta_0^{(2)} + \beta_1^{(2)}z_{1i} + \beta_2^{(2)}z_{2i} + \beta_3^{(2)}z_{3i} + u_i^{(2)},$$

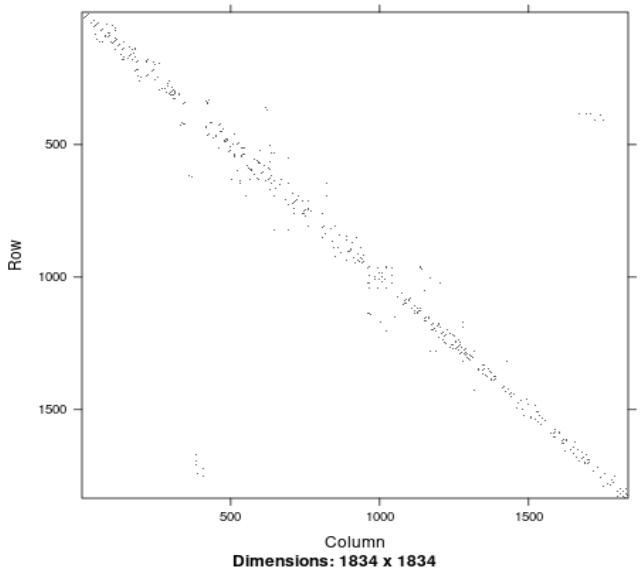


Figure 5.8: Adjacency Matrix of districts in Peru

where $\beta_0^{(k)}$, $\beta_1^{(k)}$, $\beta_2^{(k)}$ and $\beta_3^{(k)}$ for $k = 1, 2$, denote the regression coefficient for each linear predictor, $z_{1i}^{(k)}$, $z_{2i}^{(k)}$ and $z_{3i}^{(k)}$ represent the rate of total poverty, rate of householdswithout access to sanitation and the index of sevirity in the district i for each linear predictor, and $u_i^{(k)}$ represents the spatial random effect of the district $i = 1, 2, \ldots, n$ for each linear predictor. Then $\boldsymbol{u}^{(k)} = (u_1^{(k)}, u_2^{(k)}, \ldots, u_n^{(k)})^T$ follows a DAGAR structure. From these definitions, the latent Gaussian field is

$$x = \{\beta_0^{(1)}, \beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, u^{(1)}, \beta_0^{(2)}, \beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)} u^{(2)}\}$$

and the vector of hyperparameters is

$$\theta = (\tau_u^{(1)}, \rho^{(1)}, \tau_u^{(2)}, \rho^{(2)}).$$

Bayesian inference was carried out through INLA for Poisson-DAGAR, Poisson-CAR and ZIP-DAGAR models. In order to compare the time requirements of MCMC for DAGAR models, the MCMC algorithm was also implemented for the Poisson-DAGAR model. Posterior inference of MCMC was based upon one chain of 1000 iterations (with a burn-in of 5000 iterations). MCMC for the Poisson-DAGAR model takes 10800 seconds Using INLA all the fitted models run in a lower time (Table 5.4). This emphasizes the computational advantage of using the INLA method in comparison with traditional MCMC methods for large spatial data.

Table 5.4 also presents the selection criteria of the fitted models with INLA. The computational cost for the Poisson-DAGAR model was the lower one. The WAIC and LPML values indicate that the Poisson-DAGAR model has a better goodness of fit followed by the Poisson-CAR model. Further, RMSEE (0.6) is similar for Poisson-DAGAR model and Poisson-DAGAR model. Our previous simulation studies showed that when the spatial cor-

relation is small the DAGAR model performs better than the CAR model, since the CAR model usually overestimate the ρ value. In fact, the posterior mean estimate of ρ for the DAGAR model suggests a weak spatial correlation (0.1912), while the CAR model suggests a moderate spatial correlation (0.5886), see Table 5.5. Further, it is known that only the Poisson-DAGAR model interprets correctly ρ. Based on these considerations, it is concluded that the Poisson-DAGAR model enjoys a better goodness of fit.

	Poisson-DAGAR INLA	ZIP-DAGAR INLA	Poisson-CAR INLA
WAIC	12646.72	13197.08	12659.47
LPML	-6.1421	-5.0448	-6.6084
Time (sec)	501	1143	550

Table 5.4: The selection criteria for the models proposed and total run time in seconds.

The posterior mean parameter estimates of the Poisson-DAGAR and Poisson-CAR models are reported in Table 5.5. The posterior summaries for regression coefficients are quite similar. The regression coefficient credible intervals for $\beta_0, \beta_1, \beta_2$ and β_3 do not contain the zero value, thus, the variables rate of poverty, rate of households without access to sanitation, and severity index are significant to explain the confirmed covid-19 cases by districts. In addition, for each percentage unit that the rate of poverty and the rate of households without access to sanitation are inscreased, the mean of covid-19 counts is reduced in 1.76% and 3.96%, respectively. While for each percentage unit that the index of severity is inscreased, the mean of covid-19 counts is increased in 5.35%.

As it was pointed out, the Poisson-DAGAR model suggest a weak spatial correlation ($\rho = 0.1912$). It have seen from the simulation study that when the spatial correlation is weak, the DAGAR model performs better than the CAR model. This result is confirmed through the criteria assessment. And the precision parameter τ_w is small (0.1641), which means that the spatial marginal variance of the spatial process is large, and it fully explain the spatial variability in the data. For interpreting the coeficients, there is a 5% increase in the mean of counts of covid-19 when increase one percent point of Severity. The marginal densities for all the parameters are presented in the appendix.

Figure 5.9 shows the posterior mean estimates of counts of covid-19 in Peru from the Poisson-DAGAR (upper panel) and Poisson-CAR (lower panel) models.

Finally, it was also tested the predictive efficiency of the DAGAR model using INLA.

	Poisson-DAGAR INLA	Poisson-CAR INLA
β_0	6.4436 (5.9813, 6.9035)	6.6301 (6.1977, 7.0612)
β_1	-0.0178 (-0.0310, -0.0047)	-0.0193 (-0.0325, -0.0062)
β_2	-0.0404 (-0.0461, -0.0346)	-0.0408 (-0.0463, -0.0352)
β_3	0.0521 (0.0066, 0.0977)	0.0590 (0.0122, 0.1057)
ρ	0.1912 (0.1624, 0.2248)	0.5886 (0.4529, 0.7088)
τ_w	0.1641 (0.1497, 0.1797)	0.1680 (0.1548, 0.1815)

Table 5.5: Summary statistics: posterior median and 95% credible interval for the hyperparameters of each model.

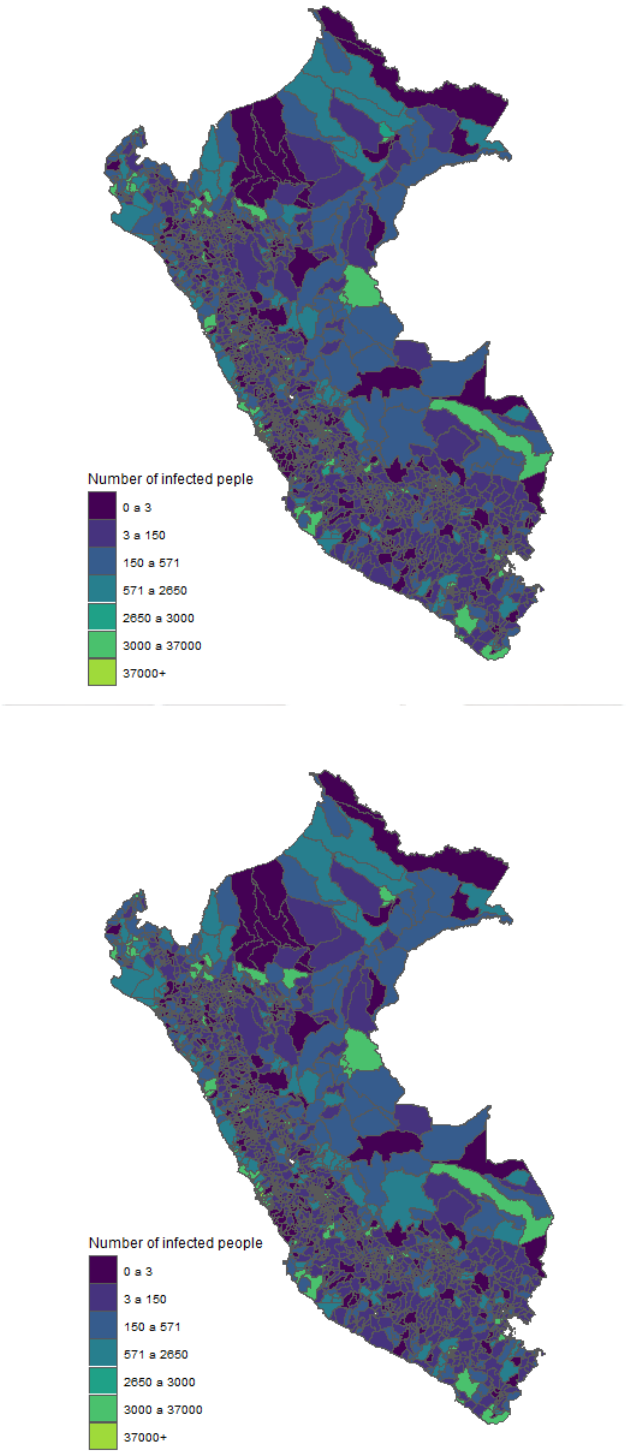


Figure 5.9: Predicted number of infected peopple in Peru

In order to predict the counts of Covid-19 in the Kelluyo district. Therefore, the Kelluyo observation in the dataset was removed and then predict it with the Poisson-DAGAR model.

From table 5.6, the Kelluyo district has a population of more than 22 thousand people and 28 of them are infected with the virus. Moreover, according to the dataset, Kelluyo has been identified with ID 47.

ID	District	Province	Departament	Population	Number of cases
47	Kelluyo	Chucuito	Puno	22766	28

Table 5.6: Kelluyo information for Covid-19.

The posterior predictive distribution is given by:

$$\tilde{\pi}(y_{47}|y_{-47}) = \int \tilde{\pi}(y_{47}|\lambda_{47})\tilde{\pi}(\lambda_{47}|y_{-47})d\lambda_{47},$$

Fig. 5.10 shows the posterior predictive density of Kelluyo, $\tilde{\pi}_{47}|y_{-47}$. The red vertical line indicates the observation y_{47} that was removed.

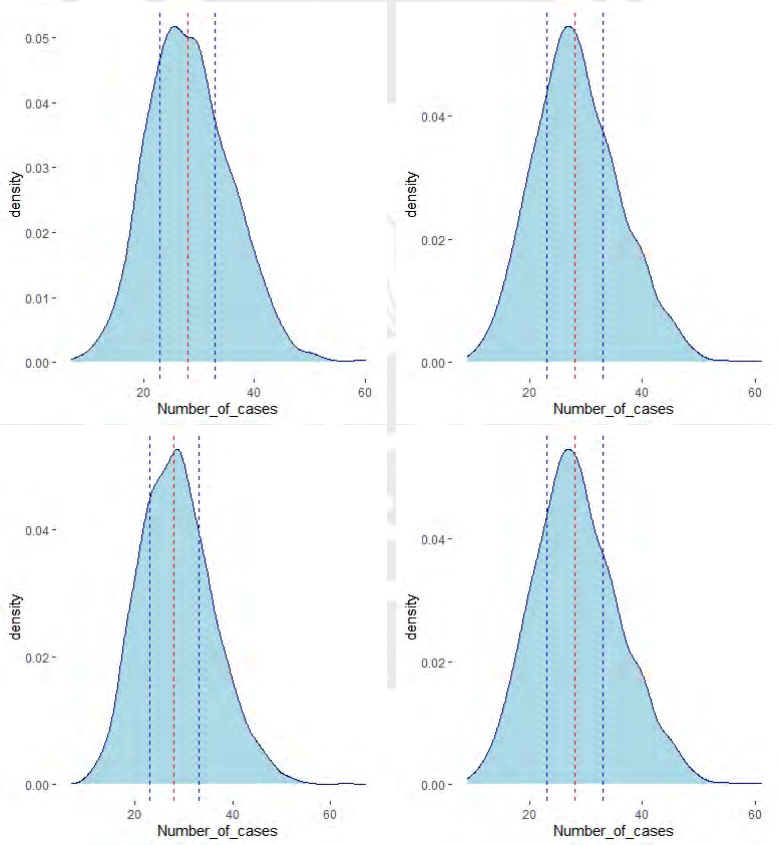


Figure 5.10: Posterior predictive density of Kelluyo $\tilde{\pi}_{47}|y_{-47}$.

Chapter 6

Conclusions and future works

The main goal of this thesis was to implement aproximate Bayesian inference for DAGAR models through INLA. DAGAR models with Gaussian distribution and non gaussian distributions were implemented for simulation studies with small and large areal data and a wide band of scenarios under weak and strong spatial autocorrelation. This analysis showed the efficiency of the INLA algorithm for DAGAR models. Using INLA each simulation for $n = 900$ took less than two minutes, while the MCMC approach took more than two hours. The results also showed the main advantages of DAGAR models, in terms of interpretability of the spatial autocorrelation parameter. Specially, when the spatial autocorrelation parameter estimate is moderate, because it is evident that in this case the DAGAR models are much better.

The performance of the DAGAR model using INLA was also assessed through two applications. The first application studied sudden infant death syndrome data (SIDS) in counties of North Caroline. The proposed models for these data were the DAGAR and CAR models for Poisson distributions. Although the data is small, these data is studied because the CAR model estimates a high spatial autocorrelation parameter, and previous studies has shown that this model usually estimates high values for this parameter. In fact, it was found out that the DAGAR model fits better these data. This result agrees with the statement that DAGAR models performs better than CAR models, specially when the spatial autocorrelation is moderate.

On the other hand, the second application studies a large dataset of covid-19 confirmed cases in districts of Peru. First, the efficiency of running the DAGAR model using INLA was huge in terms of time requirements. The DAGAR model for a Poisson distribution was better in terms of goodness of fit, in addition, the DAGAR model enjoys a better interpretability, in terms of the spatial autocorrelation parameter and precision parameter. The results showed that the confirmed cases of covid-19 are similar between districts, this pattern is natural due to the behaviour of the virus, that is extremely contagious. Morevover, the results also showed that the higher the poverty of a district, the lower the cases of covid-19. In fact, poor districts of Peru are isolated and probably people do not travel outside their district frequently. Finally, the higher the severity index, an indicator of inequality between the poor people in a district, the higher the confirmed cases of covid-19. From an health and social point of view, the modeling is relevant because it allows authorities to analyze and focus their attention on districts with high incidence of covid-19, since they could be places in

emergency situation.

Finally, as future works, the DAGAR model can be extended to implement a spatio-temporal model using INLA. This extension would make feasible to fit large data, in terms of spatial locations as well as times, in a reasonable time. Due to the flexibility of INLA, can also be fitted gaussian or non-gaussian distributions, zero-inflated models, generalized additive models, among others. Moreover, the covid-19 data could also be fitted with this model in order to study the temporal evolution of the spatial pattern of the virus. Although fitting covid-19 data is extremely difficult, this kind of modeling can be used to find out which covariates and factors can specifically explain of the virus in , as well as to understand better the spatio-temporal distribution of the virus in our country.

Appendix A

A.1 Application 2: covid-19 data in Peru

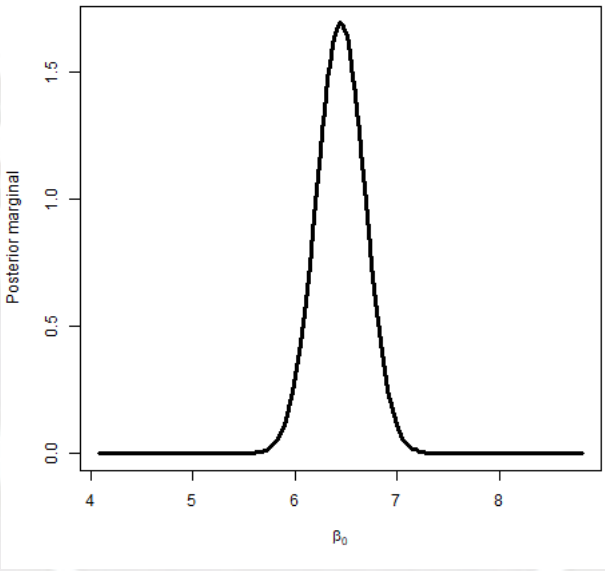


Figure A.1: Posterior marginal distribution of β_0 of the Poisson DAGAR model fitted.

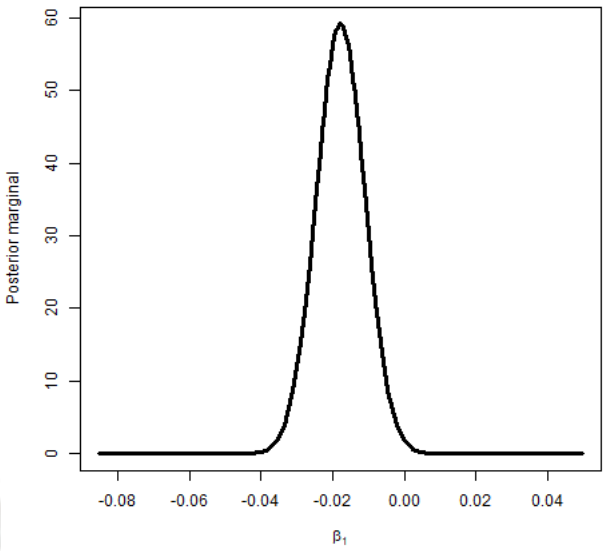


Figure A.2: Posterior marginal distribution of β_1 of the Poisson DAGAR model fitted.

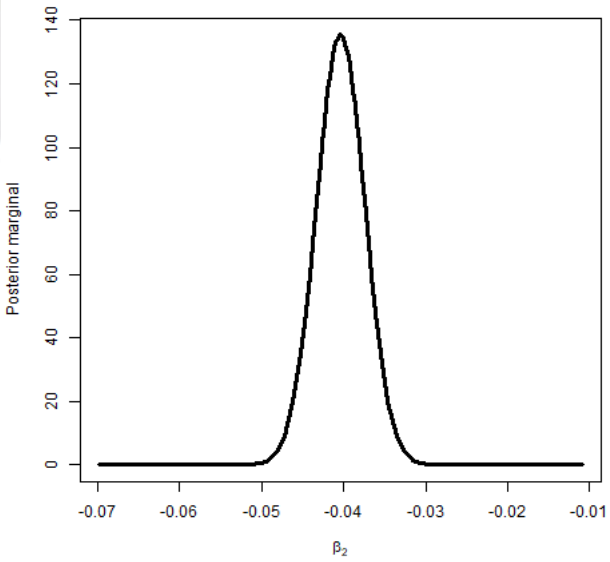


Figure A.3: Posterior marginal distribution of β_2 of the Poisson DAGAR model fitted.

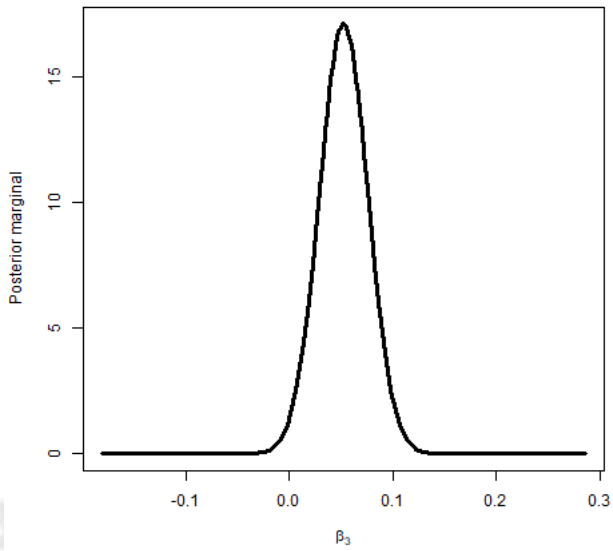


Figure A.4: Posterior marginal distribution of β_3 of the Poisson DAGAR model fitted.

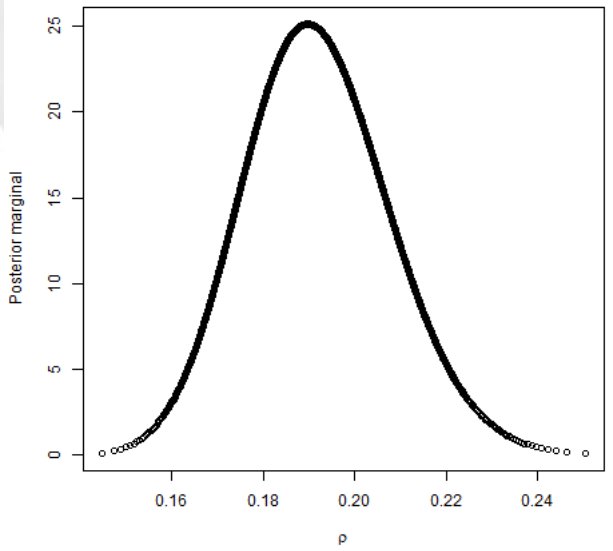


Figure A.5: Posterior marginal distribution of ρ of the Poisson DAGAR model fitted.

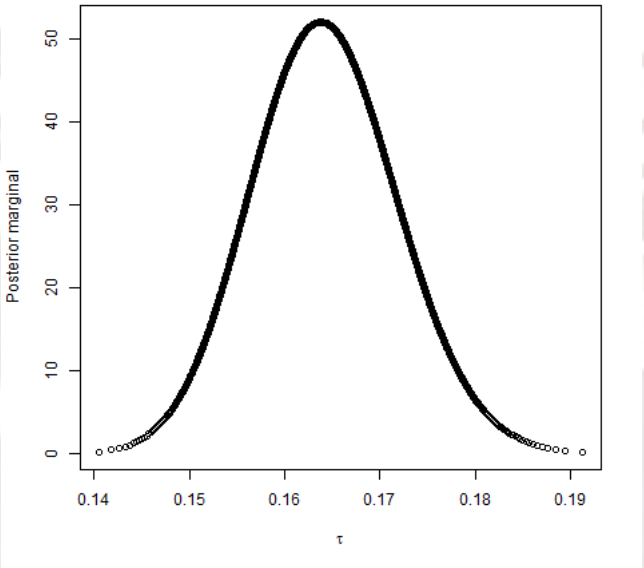


Figure A.6: Posterior marginal distribution of τ of the Poisson DAGAR model fitted.

Appendix B

B.1 Implementation of SAR latent model in R-INLA:

```
'inla.rgeneric.SAR.model' <- function(
  cmd = c("graph", "Q", "mu", "initial", "log.norm.const",
          "log.prior", "quit"),
  theta = NULL) {

  interpret.theta <- function() {
    return(list(prec = exp(theta[1L]),
                rho = 1 / (1 + exp(-theta[2L]))))
  }

  graph <- function(){
     return(Q())
  }

  Q <- function() {
    require(Matrix)
    param <- interpret.theta()
    return(param$prec * (Diagonal(nrow(Minc), x = 1) -
    param$rho)%*% t(Diagonal(nrow(Minc), x = 1) - param$rho ))
  }

  mu <- function() {
    return(numeric(0))
  }

  log.norm.const <- function() {
    return(numeric(0))
  }
```

```
  log.prior <- function() {
    ## this one I have not checked.
    param = interpret.theta()
    res <- dgamma(param$prec, 1, 5e-05, log = TRUE) + log(param$prec) +
      log(1) + log(param$rho) + log(1 - param$rho)

    return(res)
  }

  initial <- function() {
    return(c(0, 0))
  }

  quit <- function() {
    return(invisible())
  }

  if (is.null(theta))
    theta <- initial()

  res <- do.call(match.arg(cmd), args = list())
  return(res)
}
```

B.2 Implementation of CAR latent model in R-INLA:

```
'inla.rgeneric.CAR.model' <- function(
  cmd = c("graph", "Q", "mu", "initial", "log.norm.const",
          "log.prior", "quit"),
  theta = NULL) {

  #Internal function
  interpret.theta <- function() {
    return(
      list(prec = exp(theta[1L]),
           rho = 1 / (1 + exp(-theta[2L])))
    )
  }

  graph <- function(){
    require(Matrix)
```

```
    return(Diagonal(nrow(W), x = 1) + W)
  }

  Q <- function() {
    require(Matrix)

    param <- interpret.theta()

    return(param$prec * (Diagonal(nrow(Minc), x = 1) - param$rho) )
  }

  mu <- function()
  {
    return(numeric(0))
  }

  log.norm.const <- function() {
    return(numeric(0))

  }

  log.prior <- function() {
    param = interpret.theta()

    res <- dgamma(param$prec, 1, 5e-05, log = TRUE) + log(param$prec) +
      log(1) + log(param$rho) + log(1 - param$rho)

    return(res)
  }

  initial <- function() {
    return(c(0, 0))
  }

  quit <- function() {
    return(invisible())
  }

  res <- do.call(match.arg(cmd), args = list())
  return(res)
}
```

B.3 Implementation of DAGAR latent model in R-INLA:

```
##### DAGAR precision #####
dagarprec <- function(rho,Minc,ni,maxn, neimat, nijvec, intersectmat) {
  n=length(ni)
  u=rho^2
  sumfuncvec=rep(0, maxn)
  for(i in 1:maxn)  sumfuncvec[i]=i/(1-u+i*u)
  cumsumfuncvec=rep(0, maxn)
  cumsumfuncvec[1]=sumfuncvec[1]
  for(i in 2:maxn) cumsumfuncvec[i]=cumsumfuncvec[i-1]+sumfuncvec[i]
  Qd=matrix(0,n,n)
  for(i in 1:n){
    s=0
    if(ni[i]>0) for(ck in 1:ni[i]){
      k=neimat[i,ck]
      s=s+ u*cumsumfuncvec[ni[k]]/(ni[k]*(ni[k]+1))
      Qd[i,k]=Qd[i,k]-rho
    }
    Qd[i,i]=(1-u)+u*ni[i]/2+s
    if(i < n){
      for(j in (i+1):n){
        t=0
        jc=0
        counter=(i-1)*n+j
        nij=nijvec[counter]
        if(nij>0){
          jointn=intersectmat[counter,1:nij]
          for(ck in 1:length(jointn))
          {
            k=jointn[ck]
            t=t+1/(2*(ni[k]+1))+(ni[k]-
            cumsumfuncvec[ni[k]])/(ni[k]*(ni[k]+1)*(ni[k]-1))
          }
        }
        Qd[i,j]=Qd[i,j]+t
        Qd[j,i]=Qd[j,i]+t
      }
    }
  }
  Qd=Qd/(1-rho^2)
```

```
    return(Qd)
  }

  Q <- function() {
    require(Matrix)

    param <- interpret.theta()

    return(param$prec * dagarprec(param$rho,Minc,ni,maxn, neimat,
    nijvec, intersectmat))
  }

  mu <- function()
  {
    return(numeric(0))
  }

  log.norm.const <- function() {
    return(numeric(0))

  }

  log.prior <- function() {
    param = interpret.theta()

    res <- dgamma(param$prec, 1, 5e-05, log = TRUE) + log(param$prec) +
      log(1) + log(param$rho) + log(1 - param$rho)

    return(res)
  }

  initial <- function() {
    return(c(0, 0))
  }

  quit <- function() {
    return(invisible())
  }

  res <- do.call(match.arg(cmd), args = list())
  return(res)
}
```

Bibliography

Banerjee, S., Carlin, B. P. y Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society. Series B (Methodological)* **36**(2): 192–236.
URL: *http://www.jstor.org/stable/2984812*

Blangiardo, M. y Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*, John Wiley & Sons, Ltd.

Cressie, N. (1993). *Statistics for Spatial Data*, Wiley Classics Library.

Datta, A., Banerjee, S., Finley, A. O. y Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets, *Journal of the American Statistical Association* **111**(514): 800–812.

Datta, A., Banerjee, S., Hodges, J. S. y Gao, L. (2019). Spatial disease mapping using directed acyclic graph auto-regressive (dagar) models, *Bayesian Analysis* **14**(4): 1221–1244.
URL: *https://doi.org/10.1214/19-BA1177*

Geisser, S. y Eddy, W. F. (1979). A predictive approach to model selection, *Journal of the American Statistical Association* **74**(365): 153–160.

Gelman, A., Hwang, J. y Vehtari, A. (2013). Understanding predictive information criteria for bayesian models, *Statistics and Computing* **24**(6): 997–1016.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications, *Biometrika* **57**(1): 97–109.

Held, L., Schrödle, B. y Rue, H. (2009). Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA, *Statistical Modelling and Regression Structures*, Physica-Verlag HD, pp. 91–110.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. y Teller, E. (1953). Equation of state calculations by fast computing machines, *The Journal of Chemical Physics* **21**(6): 1087–1092.

Rue, H. y Held, L. (2005). *Gaussian Markov Random Fields*, Chapman and Hall/CRC.

Rue, H., Martino, S. y Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(2): 319–392.
URL: *https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00700.x*

Tierney, L. y Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association* **81**(393): 82–86.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory.

Whittle, P. (1954). On stationary processes in the plane, *Biometrika* **41**(3-4): 434–449. **URL:** *https://doi.org/10.1093/biomet/41.3-4.434*