Bayesian spatio-temporal conditional overdispersion models for count data: Proposals, estimation and applications

Mabel Morales Otero

Doctorado en Economía:
Instrumentos del Análisis Económico

Director: Prof. Vicente Núñez Antón

Departmento de Métodos Cuantitativos
Facultad de Economía y Empresa
Universidad del País Vasco / Euskal Herriko Unibertsitatea
Noviembre de 2022

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Bayesian spatio-temporal conditional overdispersion models for count data: Proposals, estimation and applications

Mabel Morales Otero

Doctoral Programme in Economics:
Tools of Economic Analysis

Supervisor: Prof. Vicente Núñez Antón

Department of Quantitative Methods
Faculty of Business and Economics
University of the Basque Country
Spain
November 2022

Acknowledgements

First of all, I would like to express my most profound gratitude to professor Vicente Núñez Antón, for the opportunity of doing my Ph.D. under his supervision. Thank you for your guidance through this period, for all the knowledge you have shared with me, for the scientific advice and for your patience. It has been a true privilege for me and I really hope to have the possibility to keep working with you for many more years.

I also feel especially grateful to María Durbán Reguera, Edilberto Cepeda Cuervo, Christel Faes and Virgilio Gómez Rubio, who have all remarkably contributed to this work and who have also had a great positive impact on my own personal research career.

I would also like to thank Christel Faes and the Center for Statistics for hosting me during my Ph.D. stay in Hasselt University. To Thomas Neyens and the Spatio-temporal Epidemiological & Environmental Modelling Group, thank you for allowing me to participate in your regular meetings and your very interesting seminars.

To Pilar González Casimiro and the Institute of Public Economy, thank you for generously allowing me to make use of the institute's facilities during all this time. In addition, I am also grateful to María Jesús Compadre Fernández and Aleida Cobas Valdés for their help and the time we spent together.

Thanks should also go to the Departament of Quantitative Methods and to the directors of the Doctoral Programme, Marta Escapa García and María Paz Moral Zuazo, for their invaluable help in many situations.

I would also like to thank Vicente Núñez Antón, Marta Regúlez Castillo and Jorge Virto Moreno for offering me the possibility to collaborate with them in some of their classes, providing me with a very useful teaching experience.

To my thesis external reviewers, Thomas Neyens and Rolando de la Cruz Mesía, thank you so much for writing such comprehensive reports on this thesis. Your comments have been of great value.

To my thesis committee members, María Durbán Reguera, Pilar González Casimiro, Thomas Neyens, Carmen Armero Cervera, Jesus Orbe Lizundia and Geert Molenberghs, I am sincerely grateful and deeply honoured for your willingness to evaluate my work.

This thesis has been developed under the research grant BES-2017-079940 (Ayudas para contratos predoctorales para la formación de doctores) funded by Ministerio de Ciencia e Innovación (MCIN, Spain) / Agencia Estatal de Investigación (AEI) (10.13039/501100011033) and by European Social Fund (ESF) "ESF Investing in your future."

Additionally, this research has also been partially funded by Ministerio de Ciencia e Innovación (MCIN, Spain), Agencia Estatal de Investigación (AEI) (10.13039/501100011033) and Fondo Europeo de Desarrollo Regional (FEDER) "Una manera de hacer Europa" under the I+D+i research grant PID2020-112951GB-I00 and by Ministerio de Economía y Competitividad (Spain), Agencia Estatal de Investigación (AEI) and the Fondo Europeo de Desarrollo Regional (FEDER), under research grant MTM2016-74931-P (AEI/FEDER, UE).

Last but not least, I would like to thank both my Cuban and Basque families for their continuous support, especially Julen, who has been there for me at all times during the years of my Ph.D. studies, celebrating my accomplishments, but also helping me to face the difficult moments.

A mi madre, mis abuelos, mis tíos, mis primos, Vladimir, les debo todo, gracias por hacer de mí la persona que soy hoy. Abo, allá donde estés, sé que estás orgulloso de mí.

Mabel Morales Otero

Contents

Chapter 1

Introduction

Generalized linear models (GLM) are commonly used to model the response variable when working with count data (McCullagh and Nelder, 1989). However, count data regression models commonly exhibit overdispersion, a phenomenon generated when the data shows a larger variance than the one that would be expected from the specification of the model itself (Hinde and Demétrio, 1998). This situation is also known as extra-Poisson or extra-binomial variation when the response variable is assumed to follow either a Poisson or a binomial distribution, respectively. Fitting a model without taking into account the presence of overdispersion may result in standard errors underestimation, among other problems, which can lead to an inferential process that may be incorrect (Hinde and Demétrio, 1998). One of the possible causes for overdispersion is the presence of correlation among the values of the variable under study for the different units considered in the specific data set being analysed (Hinde, 1982). This is specially common with spatial data, where observations in locations that are closer in space tend to show similar values, a phenomenon known as spatial autocorrelation (Getis, 2008). Consequently, these issues must be taken into account in order to obtain reliable inference processes for the estimated parameters in the proposed model.

Overdispersion has been extensively studied in the literature, particularly for the case of response variables which follow a Poisson or a binomial distribution. A common way to deal with overdispersion in these cases is to modify the mean-variance relation for the model by scaling it with a dispersion parameter larger than one, and then use a quasi-likelihood approach for the estimation (McCullagh and Nelder, 1989). For overdispersed count data following a Poisson distribution, perhaps the most popular model is the negative binomial model, initially proposed by Margolin, Kaplan and Zeiger (1981), who applied their proposed model to data from a mutagenicity assay for the Salmonella bacteria. They were able to model the variability and, hence, the precision, when replicating plate environments by means of the extra-Poisson variation parameter. The normal or log-normal Poisson models, proposed by Hinde (1982), included a random effect following a normal distribution in the linear predictor, with a variance given by the extra parameter, so that it allowed for the possible existing overdispersion in the model specification itself. This model has been used for the analysis of cancer death

rates by Breslow (1984), where he proposed two estimation methods by weighted least squares in three steps. Results indicated that there was a clear evidence of variability in the data that the Poisson model failed to capture, an issue that was resolved with the fitting of the normal Poisson model with the extra overdispersion parameter included in it.

Williams (1975) proposed the beta-binomial model for the case of binomial responses. This model assumed that the probability of success from the binomial distribution is a random variable following a beta distribution, which includes an additional parameter to model the extra-binomial variation. Williams (1982) also modified the logistic linear model by incorporating an extra-binomial variation component and proposed a methodology for computing the maximum likelihood estimates of the regression parameters based on iterated re-weighted least squares. He applied his proposal to a data set where it was required to model the proportion of germinating seeds of two different types and root extracts, finding similar results when comparing his model's estimates to those from other models previously proposed in the literature. The normal binomial model can be obtained as an extension of this model by considering that the extra-binomial variation component follows a normal distribution (Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón, 2012).

Most overdispersion models assume the existence of constant dispersion in the data but, in a considerable number of cases, dispersion could behave differently. Therefore, models have been developed to allow for the dispersion to vary as a function of some explanatory variables. Hinde and Demétrio (1998) and McCullagh and Nelder (1989) mentioned the idea of the joint modelling of the mean and dispersion parameters. Double exponential families were introduced and applied to specific settings for the binomial and Poisson cases by Efron (1986). These distributions allow to model the mean of exponential family distributions, as well as to be able to capture the possible existing overdispersion in count data by specifying a regression model for the dispersion parameter. Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón (2012) proposed Bayesian extensions of overdispersion models, such as the negative binomial and the normal Poisson models for count data, where they assumed regression structures both for the mean and for the dispersion parameters, and, therefore, they were able to specify the so-called generalized overdispersion models.

Two of the most common model specifications to account for spatial dependence in the data are the Conditional Autoregressive (CAR) model (Besag, 1974) and the Simultaneous Autoregressive (SAR) model (Whittle, 1954). These models incorporate the spatial correlation by assuming a conditional covariance structure specified by means of a spatial neighbourhood matrix. Wall (2004) examined the correlation structures that these models imply in detail by fitting them to SAT scores from students in the USA, which led to finding conterintuitive results given that the assumed CAR and SAR structures for this data implied spatial correlations among some states that did not correspond to reality or to the specific data set characteristics.

Besag (1974) introduced autoregressive models for count data, such as the auto-binomial and auto-Poisson models, among others, by following the SAR model structure.

These models, however, present the disadvantage that they can only account for negative spatial autocorrelation in the data (Kaiser and Cressie, 1997). The spatial autoregressive Poisson model was proposed by Lambert, Brown and Florax (2010), who assumed a SAR model's structure for the mean of the response variable, for which a two step limited information maximum likelihood estimation method was also proposed. They analysed a data set concerning firm births in the different states in the USA for the period 2000-2004, and concluded that this proposed model and estimation method allowed them to better understand how geographical determinants can affect the firm births under study.

In the context of Bayesian regression models for count data, spatial dependence is commonly addressed by specifying a hierarchical model, including a set of random effects in the linear predictor of a GLM, for which spatially correlated prior distributions are usually assumed. In this way, the CAR model's structure has been extended by the well known Besag-York-Mollié (BYM) model (Besag, York and Mollié, 1991), where the spatial dependence in the data is incorporated by means of a spatially correlated prior with intrinsic CAR distribution, and an extra-variability is also included with a set of uncorrelated random effects. This model, in particular, is widely used for estimating relative risks in small areas in disease mapping (Lawson, 2008). A comparison between this model and alternative models applied in this area can be found in Best, Richardson and Thomson (2005). Another distribution also frequently found in the literature for spatial random effects is the one proposed by Leroux, Lei and Breslow (2000). Alternative variants of prior distributions that have been proposed can be also consulted in Lee (2013).

More recently, a new parametrization of the BYM model has been developed with the proposal of the so-called BYM2 model, which allows for a more comprehensible interpretation of the parameters and, in addition, it offers some specific improvements over the BYM model (Riebler et al., 2016). Morris et al. (2019) applied the BYM2 model to a data set of motor vehicle crashes from 2005 to 2014, where school-age pedestrians resulted injured in the city of New York, and explored socio-demographic factors related to their occurrence. They concluded that the BYM2 model offered a good fit to this data and was also able to identify areas with increased risk of pedestrian injuries in crashes, and to establish positive relationships between this risk and the number of people who commute to work by walking, bicycle and public transport.

In order to be able to address the issue of spatial autocorrelation in the data, Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) proposed the spatial conditional models for overdispersed spatial count data, where they assumed that part of the overdispersion can be explained by the spatial neighbourhood structure in the proposed models. This model accounts for the spatial dependence in the data by incorporating a spatial term in the linear predictor with a parameter that estimates the intensity of the spatial association. By adopting a Bayesian framework, the authors applied these models to infant mortality data from Colombia and to mother's postnatal period screening test, obtaining a good fit and being able to model the positive spatial dependence, as well as accounting for data overdispersion. Additionally, to allow for nonconstant overdispersion, extensions of the generalized overdispersion models in Quintero-Sarmiento, Cepeda-Cuervo and Núñez-

Antón (2012) were also proposed by incorporating spatial neighbourhood structures in the regressions for the mean and the variance and, thus, proposing the so-called generalized spatial conditional overdispersion models.

Regarding spatio-temporal data, besides the possible existence of spatial autocorrelation, temporal dependence might also be present (Cressie and Wikle, 2011). This type of data is quite common to be found in disease mapping (Lawson, 2008), a discipline that aims to study the geographic distribution of diseases and forecast their behaviour and spread. Some of the most popular models for spatio-temporal data include those proposed by Bernardinelli et al. (1995), who included the temporal correlation in the regression structure by means of a parametric trend. In addition, Knorr-Held (2000) formulated non parametric models, where he considered random effects with spatial and temporal structures, as well as interactions between them.

Congdon and Southall (2005) performed an analysis of infant mortality rates in the north of England between the years 1921 and 1970 and their relationship with some socio-economic factors. The authors fitted several Poisson models to these data, including random effects with different spatial and temporal structures. These models allowed them to identify the characteristics which aggravated the inequalities in the infant mortality rates among the districts under study.

Carroll et al. (2016) illustrated a new approach for Bayesian spatio-temporal model selection within the disease mapping context. The authors proposed a multivariate space time mixture model, where they included weights that allowed them to determine if the most appropriated linear predictor was the one that included a spatial or a temporal term, or a mixture of both. By means of a simulation study they found that their proposed model also offered the best fit, besides being the least vulnerable to misspecification issues, and also allowing them to accommodate a large number of linear predictors. In addition, they illustrated its usefulness by fitting it to a data set concerning the incidence of skin melanoma in the counties of Georgia, USA, from 1999 to 2007.

Gómez-Rubio et al. (2019) proposed a Bayesian hierarchical model for the joint analysis of multiple diseases, controlling the correlation in the data with weights for the shared and specific effects, which allow to identify diseases with similar spatial and temporal patterns. They analysed causes of death by oral cavity, esophagus and stomach cancer in Spain, by province, from 1996 to 2004. By fitting their proposed model, they were able to identify areas with increased risk, and they also found evidence of the existence of a strong correlation between the spatial distribution of oral and esophagus cancer.

When fitting Bayesian models, it is not always possible to obtain analytically closed form expressions for the posterior distribution (Gelman et al., 2013). Therefore, it is necessary to be able to approximate this distribution by computational methods. One of the most extended examples of these approximation methods is the Markov chain Monte Carlo (MCMC) approach, an algorithm that generates simulated samples from the posterior distributions of the regression parameters by the Gibbs sampling method, a special case of the Metropolis-Hastings algorithm (Geman and Geman, 1984;

Gamerman and Lopes, 2006; Gelman et al., 2013). An alternative to these computational algorithms, such as the MCMC methods in the Bayesian estimation mentioned above, is the integrated nested Laplace approximation (INLA) algorithm (Rue, Martino and Chopin, 2009), which is based on the deterministic numerical computation of posterior distributions approximations.

There are several software packages available for the estimation of Bayesian models, such as, for example, OpenBUGS (Lunn et al., 2009), JAGS (Plummer, 2021) and the CARBayes R package (Lee, 2013), using the MCMC approach and, the R-INLA R package (Lindgren and Rue, 2015), which uses the INLA approach. Carroll et al. (2015) studied models to account for spatial dependence for Poisson distributed count data in disease mapping, and their performance was compared for the two software packages OpenBUGS and R-INLA. More specifically, the authors fitted some models such as the Poisson, normal Poisson, and BYM models, among others, to simulated count data for which spatial correlation structures were implemented in different ways. Reported results concluded that there were substantial differences in the two implementations, specially for the estimation of the random effects, but they managed to find a good fit in both cases by varying some of the prior settings. Moreover, they highlighted the much shorter computation time that R-INLA required for the fitting of a model when compared to OpenBUGS. In Vranckx, Neyens and Faes (2019), an extensive comparison of the fitting of the BYM model with OpenBUGS, CARBayes, R-INLA and other available software packages was also reported by fitting data from young people receiving diabetes medication in Belgian municipalities for the year 2014. Although no covariates were used in their analysis, the authors were able to identify locations with increased relative risk by fitting the BYM model to the data set under study.

One of the main interests in this thesis lies in presenting and illustrating a variety of applications of the generalized spatial conditional overdispersion models of Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) within different contexts. In these applications, we demonstrate their flexibility, clear and direct interpretation, as well as their straightforward implementation in some of the available software packages for Bayesian modelling. Furthermore, we also focus in developing novel and useful extensions for these models, such as extensions for modelling spatio-temporal count data, and investigate the variation in time of their spatial correlation. We also propose semiparametric extensions that would allow us to capture non linear relationships between the response and the explanatory variables and a geometric mean extension that uses the spatial lag of the logarithm of the response variable.

Among the data sets analysed, we can find some within the context of public health, such as the analysis of infant mortality rates and mother's postnatal period screening test in Colombia, the study of the impact of air pollution on respiratory health in Glasgow, UK, and the analysis of low birth weight in Georgia, USA. Additionally, within an epidemiological context, we analyse the spreading of COVID-19 in the municipalities of Flanders, Belgium. Throughout this work, we compare the results we obtain when fitting these models and propose extensions with the ones from some of models most frequently applied in the spatial and spatio-temporal count data modelling literature,

such as the BYM (Besag, York and Mollié, 1991), BYM2 (Riebler et al., 2016) and the Knorr-Held (2000) models. In addition, we also consider new alternative ways to implement the fitting of the generalized spatial conditional overdispersion models, by proposing an algorithm to fit double hierarchical generalized linear models in INLA.

In Chapter 2, we perform a review of regression models for spatial count data with overdispersion, such as the spatial conditional overdispersion models (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018), the BYM (Besag, York and Mollié, 1991) and BYM2 (Riebler et al., 2016) models, particularly focusing on the cases where the response variable follows a Poisson or a binomial distribution. In order to be able to compare their performance, we apply them to the study of infant mortality rates and to mother's postnatal period screening test in Colombia. We present the most relevant results and also include a series of posterior predictive checks, maps of the predictions obtained and marginal effects of the covariates present in the different models.

In Chapter 3, we propose semiparametric extensions of the generalized spatial conditional overdispersion models of Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), where we consider that the relationship between a given variable and the conditional mean of the response is given by a smooth function, allowing us to relax their linearity assumptions. We illustrate these proposals by fitting them to the study of the infant mortality rates and the mother's postnatal period screening test in Colombia, where we investigate the possible existence of such non linear relationships.

In Chapter 4, we propose the spatio-temporal conditional models as extensions of the spatial conditional overdispersion models of Cepeda-Cuervo, Córdoba and Núñez-Antón (2018). We apply these proposals for Poisson distributed responses to the respiratory hospital admissions in Glasgow and, for the binomial case, to the Georgia low birth weight data. In both scenarios, we compare the performance of such models with the Knorr-Held (2000) models. In addition, we also propose the temporally varying spatial lag coefficient models, which allow the coefficient for the spatial term to vary with time.

In Chapter 5, we present a study of the geographical spread of COVID-19 cases in the municipalities of the Flanders region in Belgium during the period going from September 2020 to January 2021. In order to be able to fit this data, we consider the spatial conditional model of Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) and Morales-Otero and Núñez-Antón (2021). Furthermore, we also propose an extension of these models based on the geometric mean of the incidence rates. We compare the use of different weights matrices based on contiguity, distance, differences in covariates, and on the mobility of individuals from one municipality to another. In addition, we carry out a simulation study where we test the performance of such models when the spatial correlation is given by the mobility matrix.

In Chapter 6 we propose an alternative to the MCMC approach for fitting generalized overdispersion models. Moreover, this proposal can be used to fit double hierarchical generalized linear models (DHGLM) (Lee and Nelder, 2006) in INLA. In particular, we have developed an algorithm which combines INLA and the adaptive multiple importance sampling (AMIS) to fit DHGLM. In this chapter, we illustrate the proposed method with three simulation studies and by applying it to two real data examples cor-

responding to the analysis of infant mortality rates in Colombia and to the study of the effect of sleep deprivation on the reaction time on a number of subjects.

Chapter 2

Spatial conditional overdispersion models for Poisson and binomial responses

2.1 Introduction

As already discussed in the introduction, when modelling spatial count data, the possible existence of overdispersion and spatial correlation in the data under study are issues that need to be taken into account. For this purpose, there are several approaches and models that have been developed, such as the generalized overdispersion models proposed by Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón (2012), the spatial conditional overdispersion models proposed by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) and the Besag-York-Mollié (BYM) models in Besag, York and Mollié (1991), among others. In this chapter, we review these regression models for spatial count data with overdispersion, particularly focusing on the cases where the response variable follows either a Poisson or a binomial distribution. In addition, we provide a comparison among the performance of these models when applying them to real data examples.

In Sections 2.2 and 2.3, we describe the spatial conditional and generalized spatial conditional models to be fitted to Poisson and binomial responses, respectively. In Section 2.4, the BYM models, which will be compared to the spatial conditional models, are presented. In Section 2.5, we describe the methods that will be used for their estimation within a Bayesian context. In Section 2.6, we apply these models to infant mortality rates from Colombia and present the most relevant results. In Section 2.7, the results obtained when modelling the proportion of mothers who underwent a postnatal period screening test in Colombia are presented. In addition, in Sections 2.6 and 2.7 we have also performed and included the comparison of the models under consideration with the BYM models, widely used in the analysis of spatial count data. Finally, in Section 2.8, we conclude with a discussion on the findings.

We would like to mention that the work presented in Section 2.6 has been published by Morales-Otero and Núñez-Antón (2021).

2.2 Overdispersion models for Poisson responses

Let us assume that the random variables Y_i, for $i = 1, \ldots, n$, represent counts with corresponding means $\mathrm{E}(Y_i) = \mu_i$. The Poisson model generally assumes that $Y_i \sim \mathrm{Poi}(\mu_i)$, with variance $\mathrm{Var}(Y_i) = \mu_i$, so that the variance is equal to the mean, a property that is known as equidispersion. In a generalized linear model, the mean of the distribution depends on the explanatory variables through the following regression model, known as linear predictor:

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \tag{2.1}$$

where $g(\cdot)$ is a monotonic and differentiable link function, $\mathbf{x}_i$ is the $k \times 1$ vector of explanatory variables for the i-th observation and $\boldsymbol{\beta}$ is the $k \times 1$ vector of unknown regression parameters that need to be estimated. For the Poisson regression model, the natural logarithm is often chosen as the link function; that is, $g(\mu_i) = \log(\mu_i)$. Under these assumptions, overdispersion would occur when there is extra-Poisson variability in the data, so that $\mathrm{Var}(Y_i) > \mu_i$.

Two of the most common models used to accommodate overdispersion in count data for the case of response variables following a Poisson distribution are the normal Poisson and the negative binomial models. In the normal Poisson model, the overdispersion is corrected with the inclusion of a random effect, assumed to be normally distributed, in the linear predictor. In this way, the normal Poisson can be written as:

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \nu_i, \tag{2.2}$$

where $\mathbf{x}_i$ and $\boldsymbol{\beta}$ are as before, and $\nu_i \sim N(0, \tau)$, with $\tau > 0$, for $i = 1, \ldots, n$. In this model, $(Y_i|\nu_i)$, for $i = 1, \ldots, n$, follows a Poisson distribution with conditional mean $\mathrm{E}(Y_i|\nu_i) = \mu_i$. Although the distribution of Y_i does not have a closed form expression (Hinde, 1982), when the variance τ of the random effects is small enough, the random variables $Y_i,\ i = 1, \ldots, n$, can be considered as mixed Poisson variables, with mean and variance that can be approximated by $\mathrm{E}(Y_i) \approx \mu_i$ and $\mathrm{V}ar(Y_i) \approx \mu_i + \tau\mu_i^2$ (Dean, 1992). The dispersion parameter τ allows for modelling the possible existing overdispersion and, in addition, it also captures the variability unexplained by the covariates. That is, since $\tau > 0$, the variance is larger than that specified by the Poisson model, so that $\mu_i + \tau\mu_i^2 > \mu_i$. We believe it is important to mention that the condition that the variance τ of the random effects should be small enough (Hinde, 1982 or Dean, 1992) is satisfied in the relevant cases in the application of this model in Section 2.6, so that the approximations mentioned above hold.

Another frequently used model to fit overdispersed count data is the standard negative binomial or NB2 model (Hilbe, 2011). One possible way to be able to generate this model is by considering a Poisson-gamma mixture. That is, if we assume that the random variables $m_i,\ i = 1, \ldots, n$, follow a gamma distribution, such that $m_i \sim \mathrm{G}(\tau, \tau)$, with $\tau > 0$ a parameter that needs to be estimated, and that the random count variables $Y_i,\ i = 1, \ldots, n$, conditioned on μ_i and the random variables m_i, follow a Poisson distribution with mean $\mathrm{E}(Y_i) = \mu_i m_i$, such that $Y_i \sim \mathrm{Poi}(\mu_i m_i)$, then the unconditional distribution of Y_i can be derived in the following way (Hilbe, 2011):

$$f(y_i) = \int_0^\infty f(y_i|m_i)f(m_i)dm_i = \binom{y_i + \tau - 1}{y_i} \left(\frac{\tau}{\tau + \mu_i}\right)^\tau \left(\frac{\mu_i}{\tau + \mu_i}\right)^{y_i}, \qquad (2.3)$$

which corresponds to the probability density for a negative binomially distributed count variable, so that $Y_i \sim \text{NB}(\tau/(\tau+\mu_i), \tau)$, $i = 1, \ldots, n$, with mean $\text{E}(Y_i) = \mu_i$ and variance $\text{Var}(Y_i) = \mu_i + \tau^{-1}\mu_i^2$. The dispersion parameter τ allows for the modelling of the extra-Poisson variability because, since we have that $\tau > 0$, then $\mu_i + \tau^{-1}\mu_i^2 > \mu_i$. Therefore, and from the above, for the negative binomial regression model, the linear predictor is specified for the mean μ_i, so that:

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \qquad (2.4)$$

where $\mathbf{x}_i$ and $\boldsymbol{\beta}$ are as before.

2.2.1 Spatial conditional overdispersion models for Poisson responses

One of the reasons for the existence of overdispersion in spatial data may be the possible existing spatial correlation between the responses corresponding to the different adjacent locations. Hence, it can be assumed that a portion of the overdispersion can be explained by taking into account this spatial correlation. Thus, the spatial conditional overdispersion regression models proposed by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) assumed a specific spatial structure for the variable under study. That is, they assumed that Y_i, for $i = 1, \ldots, n$, conditioned on the values in all of the neighbours of the i-th region, except for the i-th region itself (i.e., $Y_{\sim i}$), follows a conditional overdispersed distribution denoted by $f(y_i|y_{\sim i})$, for $i = 1, \ldots, n$. In this distribution, the conditional mean follows a given regression structure that includes some covariates affecting the response variable, as well as its spatial lags, together with a spatial parameter that allows to account for the intensity of the spatial dependence that is present in the data. In the case where the conditional distribution follows one of the aforementioned models, this model leads to the spatial conditional Poisson, negative binomial and normal Poisson regression models, respectively.

The spatial distribution is commonly specified by means of a neighbourhood structure, defined, for a sample of n regions, by a $n \times n$ spatial weights matrix, denoted by $\mathbf{W} = [w_{ij}]$, where its elements, w_{ij}, are the weights to be specifically used to model the strength of the dependence between the i-th and the j-th regions. These elements are given by the contiguity criteria chosen by the researcher, which can be based on the boundaries of the regions or on the distance from one spatial location to the others, or by any other alternative criteria previously proposed in the literature. It is commonly assumed that, $w_{ij} = 1$, if region i is adjacent or a neighbour to region j, and $w_{ij} = 0$, otherwise. First order contiguity can be specified, for example, when we use the criteria that regions i and j are neighbours if they share at least one point in their boundaries. Second order could also be considered if we extend the criteria by considering that i and j are neighbours if they share a common neighbour.

This weights matrix is usually standardized by rows, so that, if region i is adjacent to region j, then $w_{ij} = 1/n_i$, where n_i is the number of neighbours region i has. Along these lines, if $\mathbf{y}$ is the $n \times 1$ vector of observations for a response variable Y, then the spatial lag of Y is defined as the product of the $1 \times n$ vector corresponding to the i-th row of the weights matrix $\mathbf{W}$, $\mathbf{W}_i$, and the vector $\mathbf{y}$; that is, $\mathbf{W}_i\mathbf{y}$, a product representing the averaged values of the considered variable in the neighbouring locations for the i-th region.

In this work, we only assume first order adjacency among regions and, in addition, that the spatial weights matrix is standardized by rows. However, we believe it is important to mention that, in Chapter 5, we present a detailed description of the different weights matrices that can be specified, and we also propose some alternatives to the traditional weights matrices that can be found in the literature. Moreover, in Chapter 5, we illustrate the performance of the spatial conditional overdispersion models when considering each of these weights matrices, offering a comparison that would allow the researchers to choose the structure that best accommodates the spatial underlying process of the data under study.

The spatial conditional Poisson model is specified by assuming that the conditional distribution of the variable under study follows a Poisson distribution, that is $(Y_i|Y_{\sim i}) \sim \text{Poi}(\mu_i)$, with conditional mean $\text{E}(Y_i|Y_{\sim i}) = \mu_i$, so that its corresponding regression model, with the previously described spatial association dependence already incorporated, can be specified as:

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{y}, \tag{2.5}$$

where $\mathbf{x}_i$ and $\boldsymbol{\beta}$ are as before, ρ is the parameter incorporating the first order spatial association, $\mathbf{W}_i$ is the i-th row of the $n \times n$ weight matrix $\mathbf{W}$ that represents the spatial neighbourhood structure assumed in the model, and $\mathbf{y}$ is the vector of dimension $n \times 1$ for the observed values for the response variable under study.

The spatial conditional normal Poisson model assumes that the distribution of the variable under study, conditioned on its neighbours excluding the i-th region itself, $Y_{\sim i}$, and the normally distributed random effect $\nu_i \sim N(0, \tau)$, with $\tau > 0$, follows a Poisson distribution. That is, $(Y_i|Y_{\sim i}, \nu_i) \sim \text{Poi}(\mu_i)$, with conditional mean $\text{E}(Y_i|Y_{\sim i}, \nu_i) = \mu_i$, so that its corresponding regression model, with the previously described spatial association dependence already incorporated, can be specified as:

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{y} + \nu_i, \tag{2.6}$$

where $\mathbf{x}_i$, $\boldsymbol{\beta}$, ρ, $\mathbf{W}_i$ and $\mathbf{y}$ are as before.

In the same way, the spatial conditional negative binomial model can be also specified if we assume that the response variable under study, conditioned on $Y_{\sim i}$, follows a negative binomial distribution. That is, $(Y_i|Y_{\sim i}) \sim \text{NB}(\tau/(\tau + \mu_i), \tau)$, with conditional mean $\text{E}(Y_i|Y_{\sim i}) = \mu_i$, and with a regression structure given by equation (2.5).

2.2.2 Generalized spatial conditional overdispersion models for Poisson responses

We believe it is important to mention that, in the spatial conditional normal Poisson and the spatial conditional negative binomial models, a portion of the overdispersion that may have been generated by the possible existing spatial correlation in the data is considered to be incorporated into the model by using the specified neighbourhood spatial structure, given by the product between the spatial weights matrix and the vector of responses (i.e., by incorporating spatial lags of the variable under study), $\mathbf{W}_i\mathbf{y}$. The remaining unexplained overdispersion in the data will be modelled by means of the dispersion parameter τ. However, we should note that these models assume a constant overdispersion, and there are cases where the dispersion in the data can vary among groups or observations. The generalized overdispersion models proposed by Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón (2012) introduced Bayesian extensions of the standard overdispersion models, where regression structures are assumed both for the mean and for the dispersion parameter. Their model allowed for the dispersion to vary as a function of some explanatory variables, so that dispersion can vary for the different regions or observations in the study.

In this sense, generalized overdispersion models offer a reasonable and well justified proposal for fitting count data with overdispersion. However, for the case of spatial count data, they do not provide information or incorporate into the model the possible existing spatial dependence in the data set under study, which clearly motivates the inclusion of the spatial dependence structure in these models. Along these lines, the generalized spatial conditional overdispersion regression models (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018) assumed that the spatial count variable in their model, $Y_i,\ i = 1, \ldots, n$, conditioned on the values in all of its neighbours, except the i-th region itself (i.e., $Y_{\sim i}$), follows an overdispersed conditional distribution $f(y_i|y_{\sim i}),\ i = 1, \ldots, n$, with conditional mean and dispersion parameter following specific regression structures that include some covariates affecting the response variable and the spatial lags of the variable of interest.

If we now consider the case where Y_i follows a Poisson distribution with mean μ_i, and assume the normal Poisson model with mean structure given by equation (2.6), for the generalized spatial conditional normal Poisson model, we will have that the conditional mean and variance components in the random effect distribution, will be specified by regression structures given by:

$$\begin{aligned} \log(\mu_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + \rho_1 \mathbf{W}_i\mathbf{y} + \nu_i, \ \text{ with } \ \nu_i \sim N(0, \tau_i), \ \text{ for } \ \tau_i > 0 \ \text{ and} \\ \log(\tau_i) &= \mathbf{z}_i^\top \boldsymbol{\gamma} + \rho_2 \mathbf{W}_i\mathbf{y}, \end{aligned} \tag{2.7}$$

where $\mathbf{x}_i$, $\boldsymbol{\beta}$, $\mathbf{W}_i$ and $\mathbf{y}$ are as before, and ρ_1 and ρ_2 are the parameters that explain the spatial association in the mean and dispersion structures, respectively. In addition, $\mathbf{z}_i$ is the $q \times 1$ vector of explanatory variables for the i-th observation and $\boldsymbol{\gamma}$ is a vector of dimension $q \times 1$ containing the unknown regression parameters that need to be estimated. The generalized spatial conditional negative binomial model can be specified in the same way, by assuming that $Y_i,\ i = 1, \ldots, n$, conditioned on $Y_{\sim i}$, follows a negative binomial

distribution, with dispersion parameter $\tau_i > 0$, and assuming the following regression structures for both the conditional mean and dispersion parameter:

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \rho_1 \mathbf{W}_i \mathbf{y} \quad \text{and} \quad \log(\tau_i) = \mathbf{z}_i^\top \boldsymbol{\gamma} + \rho_2 \mathbf{W}_i \mathbf{y}, \tag{2.8}$$

where $\mathbf{x}_i$, $\boldsymbol{\beta}$, $\mathbf{W}_i$, $\mathbf{y}$, ρ_1, ρ_2, $\mathbf{z}_i$ and $\boldsymbol{\gamma}$ are as before.

2.3 Overdispersion models for binomial responses

Let Y_i, for $i = 1, \ldots, n$, be random variables and suppose there are n clusters (or groups), with n_i observations (trials) each. A binomial regression model is specified when we assume that these variables follow a binomial distribution $Y_i \sim \text{Bin}(n_i, \pi_i)$, with π_i being the probability of success of a trial in cluster i. The mean and variance of Y_i would be given by $\text{E}(Y_i) = \mu_i = n_i\pi_i$ and $\text{Var}(Y_i) = n_i\pi_i(1-\pi_i)$, respectively, for $i = 1, \ldots, n$. Then, a regression model is specified for the mean, or the probability of success, by means of a logistic function so that:

$$\log\left(\frac{\mu_i}{n_i - \mu_i}\right) = \text{logit}(\pi_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \tag{2.9}$$

where $\mathbf{x}_i$ is the $k \times 1$ vector of explanatory variables for the i-th observation and $\boldsymbol{\beta}$ is the $k \times 1$ vector of unknown regression parameters that need to be estimated.

When fitting a binomial regression model, overdispersion can arise when the variance of the data is larger than the variance specified by this model, that is, when $\text{Var}(Y_i) > n_i\pi_i(1-\pi_i)$. In order to overcome this problem, several alternatives have been proposed in the literature, such as the beta binomial model (see, for example, Griffiths, 1973; Williams, 1975 and Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón, 2012) and the normal binomial model (see, for example, Williams, 1982 and Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón, 2012), among others.

In the beta binomial model, we assume that the response variable Y_i, conditioned on a random variable π_i^* follows a binomial distribution, so that $(Y_i|\pi_i^*) \sim \text{Bin}(n_i, \pi_i^*)$. Then, it is assumed that π_i^* follows a Beta distribution, that is $\pi_i^* \sim \text{Beta}(\tau\pi_i, \tau(1-\pi_i))$, with an unknown parameter $\tau > 0$, which needs to be estimated. Here, the unconditional distribution of Y_i is beta binomial, $Y_i \sim \text{BB}(n_i, \pi_i, \tau)$ with $\text{E}(Y_i) = n_i\pi_i$ and $\text{Var}(Y_i) = n_i\pi_i(1-\pi_i)[1+\tau(n_i-1)]$, where π_i follows the regression structure:

$$\text{logit}(\pi_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \tag{2.10}$$

with $\mathbf{x}_i$ and $\boldsymbol{\beta}$ as before. Note that the parameter τ allows us to model the overdispersion that might be present in the data that the binomial model is not able to account for.

In order to specify the normal binomial model, a normally distributed random effect, that is $\nu_i \sim N(0, \tau)$, $\tau > 0$, is included in the regression model for the mean of a binomial model, in this way:

$$\text{logit}(\pi_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \nu_i, \tag{2.11}$$

where $\mathbf{x}_i$ and $\boldsymbol{\beta}$ are as before. Here, for small values of τ, the unconditional mean and variance of Y_i can be approximated as $\mathrm{E}(Y_i|\nu_i) = n_i\pi_i$ and $\mathrm{Var}(Y_i) \approx n_i\pi_i(1-\pi_i)[1 + \tau(n_i-1)\pi_i(1-\pi_i)]$ respectively (Williams, 1982 and Dean, 1992). In the same way as with the beta binomial, the normal binomial model allows us to capture the extra binomial variability in the data by means of the dispersion parameter τ.

2.3.1 Spatial conditional overdispersion models for binomial responses

Let us assume that the variables Y_i, for $i = 1, \ldots, n$, represent counts on n regions, with n_i trials for each region. The spatial conditional overdispersion models for responses following a binomial distribution (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018) can be specified if we assume that the variables Y_i, conditioned on the set of values they take in all the neighbouring locations except for the i-th area itself, denoted by $Y_{\sim i}$, for $i = 1, \ldots, n$, follow an overdispersed distribution denoted by $f(y_i|y_{\sim i})$. When this conditional distribution is one of those that we have previously mentioned in Section 2.3, we can obtain the spatial conditional binomial, beta binomial and normal binomial regression models, respectively. The conditional mean of this distribution is then modelled via a regression structure that includes some covariates affecting the response variable and spatial terms, together with a spatial parameter that allows to account for the strength of the spatial dependence that is present in the data.

To be able to capture the spatial association, Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) proposed the use of a spatial term given by the $n \times 1$ vector $\mathbf{A}$, with elements defined the following way: $A_i = \frac{\hat{\pi}_{\sim i}}{1-\hat{\pi}_{\sim i}}$, where $\hat{\pi}_{\sim i} = \frac{\mathbf{W}_i\mathbf{y}}{\mathbf{W}_i\mathbf{n}}$ with $\mathbf{W}_i$ being the i-th row of spatial weights matrix $\mathbf{W}$, corresponding to the i-th observation, and $\mathbf{y}$ and $\mathbf{n}$ being the vectors of observations of the response variable and the number of trials, respectively. Here, the spatial weights matrix $\mathbf{W}$ is defined in the same way as in Section 2.2.1, for the spatial conditional models for Poisson distributed response variables.

The spatial conditional normal binomial model is specified by assuming that, given a set of normally distributed random effects $\nu_i \sim N(0, \tau)$, $\tau > 0$, the conditioned variables $(Y_i|Y_{\sim i}, \nu_i)$ follow a binomial distribution, so that $(Y_i|Y_{\sim i}, \nu_i) \sim \mathrm{Bin}(n_i, \pi_i)$, where π_i follows the regression model:

$$\mathrm{logit}(\pi_i) = \mathbf{x}_i^\top\boldsymbol{\beta} + \rho A_i + \nu_i, \tag{2.12}$$

with $\mathbf{x}_i$ and $\boldsymbol{\beta}$ as before, ρ is the unknown spatial parameter that needs to be estimated and A_i is the i-th element of the spatial term $\mathbf{A}$.

In the same way, we can obtain the spatial conditional beta binomial model by assuming a beta binomial distribution on the conditioned variables $(Y_i|Y_{\sim i})$. That is $(Y_i|Y_{\sim i}) \sim \mathrm{BB}(n_i, \pi_i, \tau)$, $\tau > 0$, where π_i follows the regression model:

$$\mathrm{logit}(\pi_i) = \mathbf{x}_i^\top\boldsymbol{\beta} + \rho A_i, \tag{2.13}$$

with $\mathbf{x}_i$, $\boldsymbol{\beta}$, ρ and A_i as before.

2.3.2 Generalized spatial conditional overdispersion models for binomial responses

As we have already mentioned in Section 2.2.2, generalized overdispersion models were proposed by Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón (2012) in order to provide more flexibility to overdispersion models by allowing the dispersion parameter to vary across the observations or regions. These models can be specified if we assume two regression structures, one for the mean and another for the dispersion parameter, which can include some given covariates or, in the case of the generalized spatial conditional models, also spatial terms.

More specifically, the generalized spatial conditional normal binomial model is defined so that $(Y_i|Y_{\sim i}, \nu_i) \sim \text{Bin}(n_i, \pi_i)$ where the random effects $\nu_i \sim N(0, \tau_i)$, $\tau_i > 0$, for $i = 1, ..., n$, depend on a different variance τ_i for each region. Here, we will have that the conditional probability of success and the dispersion parameter, which is the variance of the random effect, follows the regression structures:

$$\text{logit}(\pi_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \rho_1 A_i + \nu_i \quad \text{and} \quad \log(\tau_i) = \mathbf{z}_i^\top \boldsymbol{\gamma} + \rho_2 A_i, \tag{2.14}$$

with $\mathbf{x}_i$, $\boldsymbol{\beta}$ and A_i as before and ρ_1 and ρ_2 are the parameters that explain the spatial association in the mean and dispersion models, respectively. In addition, $\mathbf{z}_i$ is the $q \times 1$ vector of explanatory variables for the i-th observation and $\boldsymbol{\gamma}$ is a vector of dimension $q \times 1$ containing the unknown regression parameters that need to be estimated.

In the same way, the generalized spatial conditional beta binomial model can be defined if we assume that $Y_i \sim \text{BB}(n_i, \pi_i, \tau_i)$, where the probability of success and the dispersion parameter follow these regression equations respectively:

$$\text{logit}(\pi_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \rho_1 A_i \quad \text{and} \quad \log(\tau_i) = \mathbf{z}_i^\top \boldsymbol{\gamma} + \rho_2 A_i, \tag{2.15}$$

with $\mathbf{x}_i$, $\boldsymbol{\beta}$, A_i, ρ_1, ρ_2, $\mathbf{z}_i$ and $\boldsymbol{\gamma}$ as before.

2.4 Besag-York-Mollié (BYM) model

The Besag-York-Mollié (BYM) model (Besag, York and Mollié, 1991) is a Bayesian hierarchical model widely used in the literature for fitting spatial count data, particularly in the field of disease mapping (Lawson, 2008). It is an extension of the generalized linear model that includes both a spatially structured and an unstructured random effect in the regression model structure. If we let Y_i, $i = 1, \ldots, n$, represent counts for the n different regions, the BYM model is specified by assuming that the variable under study follows a Poisson distribution with mean $\text{E}(Y_i) = \mu_i$, having a mean regression structure given by:

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \nu_i + \eta_i, \tag{2.16}$$

with $\mathbf{x}_i$ and $\boldsymbol{\beta}$ as before, ν_i is a normally distributed random effect, so that $\nu_i \sim N(0, \tau)$, with $\tau > 0$ being an unknown variance parameter that needs to be estimated, and η_i is

an intrinsic conditional autoregressive (CAR) (Besag, York and Mollié, 1991) distributed random effect, so that:

$$(\eta_i|\eta_{\sim i}, \mathbf{W}, \tau_\eta) \sim N\left(\frac{\sum_{j=1}^{n} w_{ij}\eta_j}{\sum_{j=1}^{n} w_{ij}}, \frac{\tau_\eta}{\sum_{j=1}^{n} w_{ij}}\right), \tag{2.17}$$

where $\eta_{\sim i}$ represents the set of values of all neighbours of the i-th region, except for the i-th region itself, $\mathbf{W}$ is the spatial weights matrix and $\tau_\eta > 0$ is an unknown variance parameter that needs to be estimated. Given that this model is able to account for spatial dependence and also for the extra-variability in the data set not explained by the covariates, it has a considerable potential to motivate and justify its use for the analysis of spatial count data. However, since it is only possible to obtain information from the data from the sum of the two random effects, but not from each of the individual components separately, its use in this specific context has been questioned because of the possible identifiability problems that it may present (Eberly and Carlin, 2000).

Some authors (e.g., Riebler et al., 2016) have addressed the issue of identifiability. They proposed the BYM2 model, an extension of the BYM model that scales the spatial component and the unstructured component, so that the mean regression structure can be written as:

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \frac{1}{\sqrt{\tau_s}}\left(\sqrt{1-\phi_s}\nu_i + \sqrt{\phi_s}\eta_i\right), \tag{2.18}$$

where the random effects ν_i and η_i are as in the BYM model, but with a scaled variance approximately equal to one, and τ_s and ϕ_s are unknown parameters to be estimated. Here, the precision parameter τ_s captures the variance contribution from the sum of the two random effects and the mixing parameter ϕ_s controls for the variance contribution of the spatially structured component $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$, whereas the variance contribution of the unstructured random component $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)^\top$ is explained by $1 - \phi_s$.

The main advantage of the BYM2 model is precisely the possibility it offers to be able to separately capture the impact of the spatial dependence and the effect of the variability or the overdispersion present in the data. The priors for these hyperparameters are defined by means of the penalized complexity priors developed by Simpson et al. (2017). The complexity prior for the parameter τ_s can be specified by assuming the probability statement that $\text{Prob}(1/\sqrt{\tau_s} > U) = \alpha$, and, for the parameter ϕ_s, that $\text{Prob}(\phi_s < U) = \alpha$, with U and α being fixed values that depend on the specific application under consideration. The use of these priors has been proved to be a suitable choice in Bayesian spatial models and, especially, for the BYM2 model, mainly due to the fact that they favour less complex models and allow for a clearer interpretation of the parameters (Bakka et al., 2018).

Although the BYM model is mostly used to fit Poisson distributed data, it is also possible to model binomially distributed responses with it (Lee, 2013). The spatially structured and the unstructured random effects would be included in the logistic regression model for the probability of success (Blangiardo and Cameletti, 2015). Let us assume that for n areas, the random variables Y_i, for $i = 1, \ldots, n$, with n_i trials on each area, follow a binomial distribution, $Y_i \sim \text{Bin}(n_i, \pi_i)$ with π_i being the probability of

success of a trial in the i-th area. Then, the BYM model would be specified if we assume the following regression structure for π_i:

$$\text{logit}(\pi_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \nu_i + \eta_i, \tag{2.19}$$

where $\mathbf{x}_i$, $\boldsymbol{\beta}$, ν_i and η_i are as before.

2.5 Bayesian estimation

As we have previously mentioned, models studied here will be estimated by using a Bayesian approach. That is, we assume that we have a sample of n independent observations, y_i, for $i = 1, \ldots, n$, from the variable $\mathbf{Y}$, and we wish to estimate a parameter θ. We will consider it as a random variable and express our beliefs about this parameter via a prior distribution $p(\theta)$. The information available in the data about θ will be included in the likelihood function $L(\mathbf{y}|\theta)$, which is the joint distribution of the sample, so that, if $y_i,\ i = 1, \ldots, n$, given the parameter θ, are independent and have a probability density function given by $f(y_i|\theta)$, then $L(\mathbf{y}|\theta) = \prod_{i=1}^{n} f(y_i|\theta)$. In Bayesian inference, we use this information to update our knowledge using the Bayes theorem, thus, being able to obtain a posterior distribution for the parameter given the data, $p(\theta|\mathbf{y}) \propto L(\mathbf{y}|\theta)p(\theta)$. In the case of spatial conditional regression models, Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) considered the variables $(Y_i|Y_{\sim i})$, conditioned on the assumed spatial neighbourhood structure, following an overdispersion distribution such as the ones mentioned in Sections 2.2 and 2.3, and the parameter θ, to be estimated, to be independent. Therefore, under these independence assumptions, the likelihood function can be obtained in the usual way and, thus, the Bayesian inference process is valid.

In Bayesian analysis, vague or noninformative prior distributions for the parameters are usually specified in order to minimize the possible impact of prior information, compared to the likelihood of the data, on the posterior inference. For the regression coefficients of explanatory variables, typically normal prior distributions with zero mean and large variances are considered. In this case, we assume that $\beta_j \sim N(0, 1 \times 10^5),\ j = 1, \ldots, k$. In most software packages available for Bayesian inference approaches, the prior distribution for the variance component in the normal distribution is implemented on its inverse instead, which is usually labelled as the precision parameter (i.e., ψ), so that $\psi = 1/\tau$, if τ is the variance component parameter. For this precision parameter several prior distributions have been proposed in the literature for Bayesian hierarchical models (Gelman, 2006), with the gamma distribution being the most commonly used. In this way, it is assumed that $\psi \sim G(\alpha_1, \alpha_2)$, with α_1 and α_2 being fixed and user-specified parameters. The choice of these values, α_1 and α_2, is a crucial issue that needs to be addressed in a careful manner, mainly because inference can be sensitive to their selection, specially when the data set does not have a large number of observations available (Gelman, 2006). Specific values of $\alpha_1 = \alpha_2 = 0.001$ for this prior distribution are often employed in many applications (Lawson, 2008), so that $\psi \sim G(0.001, 0.001)$, which, given that its mean is equal to 1 and its variance equal to 1000, a large value, it can be considered as a vague prior. Alternative frequently used

values that can be found in the literature are, $\alpha_1 = 1$ and $\alpha_2 = 0.01$, in Vranckx, Neyens and Faes (2019), $\alpha_1 = 0.05$, $\alpha_2 = 5 \times 10^{-4}$ in Best, Richardson and Thomson (2005), $\alpha_1 = 1$ and $\alpha_2 = 0.5$ in Carroll et al. (2015), and $\alpha_1 = \alpha_2 = 1 \times 10^{-4}$ in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), among others. Nevertheless, the choice of these parameters must be based on their adequacy to the specific application considered and its adverse effects on the posterior inference should be appropriately assessed and studied.

Following these guidelines, for our Bayesian analysis, we will assume noninformative prior distributions with zero mean and large variance for the regression parameters, as well as for the spatial lag parameters included in the proposed models. For the inverse of the dispersion parameters, $\psi = 1/\tau$, we will specify gamma distributions with large variances, so that $\psi \sim G(\alpha, \alpha)$, with α being a very small value. Estimation will be carried out in OpenBUGS, JAGS and also in R-INLA for some specific cases.

Model selection will be performed by using the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) and the Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010), also known as the Widely Applicable Information Criterion, where the models with the lowest values for these criteria would be considered as the best fitting ones. On the one hand, DIC is based on the posterior distribution of the deviance statistic, a measure of the model's fit, and it is penalized by the effective number of parameters, which is a measure representing the complexity of the model. On the other hand, WAIC is based on the logarithm of the pointwise posterior predictive density and receives a penalty specified by a different definition of the effective number of parameters. This criterion has become very popular in the last few years, since it is considered as a fully Bayesian approach (Gelman et al., 2013; Gelman, Hwang and Vehtari, 2014). Given that each of these two measures has its own advantages and drawbacks (Gelman, Hwang and Vehtari, 2014), we will include both of them in our analysis, since we believe the information provided by one can be complemented by the other. Moreover, besides these information criteria values, we will also take into account the predictive accuracy of the fitted models to select the best fitting ones by performing posterior predictive checks on each one of the fitted models.

In general, when comparing fitted models with an information criterion, such as the DIC or WAIC, differences of more than 10 units between the values obtained for two specific models under comparison are often considered as substantial, whereas differences of less than 2 units are mostly considered as not significant (Spiegelhalter et al., 2002; Burnham and Anderson, 2002). In the second case, the most parsimonious model is typically chosen as the best fitting one (Lawson, 2008), which may be given by the model with the simplest structure or where fewer covariates were included. In this thesis, we will follow this approach. However, we should take into account that, in some cases, we will prefer the model that makes more sense within the specific context of the application under consideration.

2.6 Application to infant mortality in Colombia

The data we will analyse here has been obtained from the National Statistics Department of Colombia and corresponds to 32 departments or geographical units (areas or regions) in this country. For each geographical unit, the available variables are: the number of children under one year of age who died in 2005 (i.e., variable ND), the total number of births in 2005 (i.e., variable NB), an index representing the percentage of the population not having their basic services satisfactorily attended for the same year (i.e., variable IBN), the amount of resources (in thousands of dollars) for academic achievement or education and integral attention for young children provided by the government per household in the year 2005 (i.e., variable Rec), the percentage of women over the age of 18 who had suffered physical violence from their current partners (i.e., variable Viol), the percentage of young people (i.e., between 18 and 24 years) who were able to opt for a higher educational level (i.e., variable HE), and the percentage of children under one year of age who received the third dose of the polio vaccine in the year 2004 (i.e., variable Vac).

A similar version of this data has been previously analysed by Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón (2012) and Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), where the authors fitted their proposed generalized spatial conditional models to analyse mortality rates for children under five years of age, considering the response variable to be the number of children under five years of age who died in each department in Colombia from 2000 up to 2005. They found evidence of overdispersion and positive spatial autocorrelation in the data set they analysed, and they were able to capture these specific features in the data set under study with the fitting of the proposed models, where positive significant relations between the variables IBN and the number of deaths for children under five years of age, as well as negative significant relations between the variable Rec and mortality rates were found. We would like to indicate that the data set here was directly obtained from the National Statistics Department of Colombia because we were unable to have access to the one previously analysed by other authors and, thus, our analysis is not applied to the same data set.

In this section, we study the mortality rates for children under one year of age in the year 2005, and fit the models previously discussed in Section 2.2. The explanatory variables that we will include in the study constitute relevant socio-economic indicators that can considerably affect infant mortality (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018). In order to specify noninformative prior distributions for the parameters in our Bayesian analysis, we assume independent normal distributions, $N(0, 1\times10^5)$, for all the regression parameters; that is, $\beta_j \sim N(0, 1\times10^5)$, $j = 1, \ldots, k$, as well as for the spatial association parameter ρ. As for the inverse of the dispersion parameters τ, $\psi = 1/\tau$, based on the sensitivity analysis performed in Section 2.6.2, gamma $G(1\times10^{-4}, 1\times10^{-4})$ distributions were assumed. When running the implemented software programs in OpenBUGS, and after 10000 iterations, a burn in period of 5000 samples and considering a thinning parameter of 10, the MCMC chains showed strong signs of convergence for all of the parameters included in the proposed models.

With the available information in the data set under study, we can approximate infant mortality rates as the number of children under one year of age who died in the year 2005 per 1000 born alive in each of the departments in Colombia. In this way, we can obtain a new variable (i.e., the variable Rates) as:

$$\text{Rates}_i = \frac{\text{ND}_i}{\text{NB}_i} \times 1000, \ i = 1, \ldots, n \tag{2.20}$$

In order to better understand the data under study, Table 2.1 includes a summary of some descriptive statistics for the available variables in the Colombian mortality rates data set. Here, we could highlight the fact that the variable Rec has a considerably large range, going from 8.02 to 274.1, with a large standard deviation of 64.56. As the mean and the median are close to around 80, there are reasons to believe that this variable could be left skewed. However, as they are relatively close, this might be due to the possible existence of outliers in the right part of the distribution. For example, the largest value of this variable, that is 274.1, corresponds to the Antioquía department, which is the second most populated region after Bogotá. Given this fact, and the specific socio-demographic characteristics of this region, we do not believe that this value would affect the analysis. We will examine this variable in more detail in Section 3.4.1 in the next chapter.

Table 2.1: Descriptive statistics for the variables available in the study of infant mortality in Colombia.

	Rates	**ND**	**NB**	**Viol**	**IBN**	**Rec**	**HE**	**Vac**
Median	23.17	472.50	22574.50	35.87	35.55	81.19	13.85	68.80
Mean	25.82	542.88	26731.59	34.73	37.99	84.29	15.69	69.32
SD	10.37	444.27	27567.57	5.64	17.15	64.56	10.52	10.96
Minimum	14.25	16.00	408.00	22.58	9.20	8.02	1.30	42.70
Maximum	49.33	1802.00	115890.00	44.69	79.20	274.10	52.20	94.70

At this point, it could be important to mention that the sample size of the dataset under analysis could be considered as relatively limited, since it only includes 32 observations. In any case, this does not violate the assumptions of the models considered. Furthermore, it has been demonstrated by the authors who have already analysed similar versions of this data, that this fact does not affect the proposed methodology. Moreover, we will also address this issue in Section 3.6 in the next chapter.

Figure 2.1 shows the spatial distribution of the variable Rates, representing an approximation of infant mortality rates in each department of Colombia for the year 2005. In the map, there are clear indicators of spatial association in the data, as regions with similar values of the variable appear to be grouped together in space. Departments located in the east of the country, belonging to the natural region called the Amazonia, which is located in the Amazonian rain forest, show large values of mortality rates. In

addition, for departments located in the central part of the country, surrounding the capital, Bogotá, smaller rates can be observed.

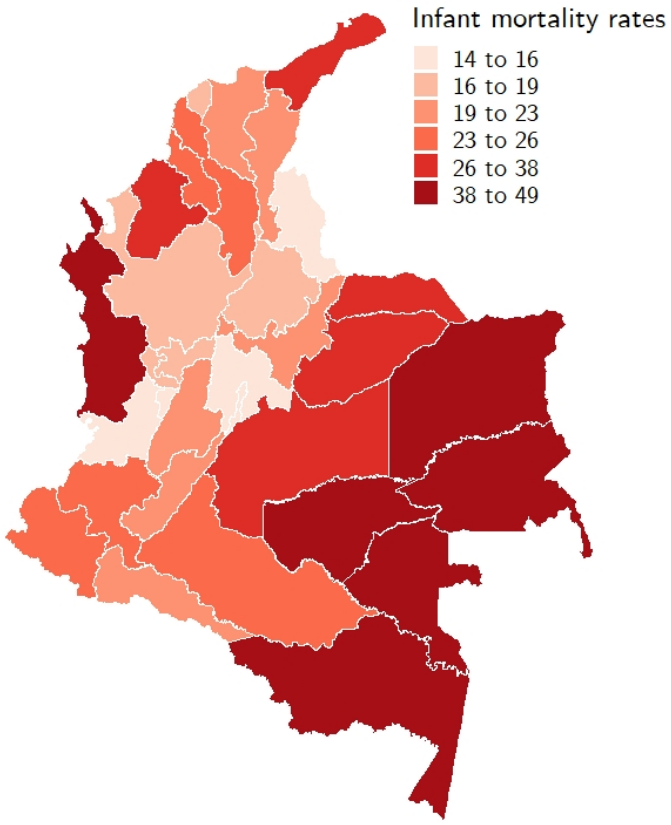


Figure 2.1: Spatial distribution of infant mortality rates in Colombia for the year 2005, by department, obtained as the number of children under one year of age who died in the year 2005 per 1000 born alive.

It is important to mention that it has seemed reasonable for authors that have previously analysed similar data sets to assume that, with regard to infant mortality rates, regions closer in space share somehow similar socio-demographic and economic characteristics, i.e. they can present similar values of infant mortality rates. As observed in the regional map shown in Figure 2.1, we have reasons to believe that this data set may also exhibit spatial dependence, which must be properly assessed and, if required, it should also be appropriately included in the proposed models to be fitted to this specific data set. An additional issue worth mentioning is the possibility that the aforementioned association may also be related to migration, either to other regions or to the capital. This is a topic that needs to be further explored and the appropriate data sets, if available, where infant mortality rates and migration are related, should be used. However,

to our knowledge, such data sets are not available at the moment and, therefore, this issue remains to be further investigated. Moreover, we believe it is out of the scope of the research proposals presented here. Therefore, we will specify a spatial neighbourhood structure, represented by the $n \times n$ first order spatial weights matrix $\mathbf{W} = [w_{ij}]$, standardized by rows, with $w_{ij} = 1/n_i$, if i is adjacent, and hence, a neighbour to region j; that is, if they share at least one point in their boundaries, and $w_{ij} = 0$, otherwise. Here, n_i is the number of neighbours for the i-th region. The spatial lag of the variable Rates is obtained by multiplying the i-th row of the spatial weights matrix $\mathbf{W} = [w_{ij}]$, $\mathbf{W}_i$, and the $n \times 1$ vector of observations for the variable Rates, that is $\mathbf{W}_i\mathbf{Rates}$.

In order to be able to assess the possible existence of spatial dependence in the data, we apply the global Moran's I test (Moran, 1948) to the variable Rates, obtaining a test statistic value of I=0.3490, providing a p-value=0.0017. Hence, for the specific data set under study, there is evidence against the null hypothesis that the values of the variable Rates are randomly distributed across the different regions in Colombia. This result can be better seen in a graphical display in Moran's scatterplot, shown in Figure 2.2, where the variable Rates is plotted against the spatial lag $\mathbf{W}_i\mathbf{Rates}$, and where a red line represents the estimated linear regression line fitted for these two variables. The slope of this regression is precisely Moran's I statistic and, as can be seen in Figure 2.2, the values of the variable for each department appear to have a positive significant correlation with the averaged values of the variable in the adjacent regions. Hence, there is substantial evidence for the possible existence of positive spatial autocorrelation in the data that needs to be accounted for in the models to be fitted in the following sections.

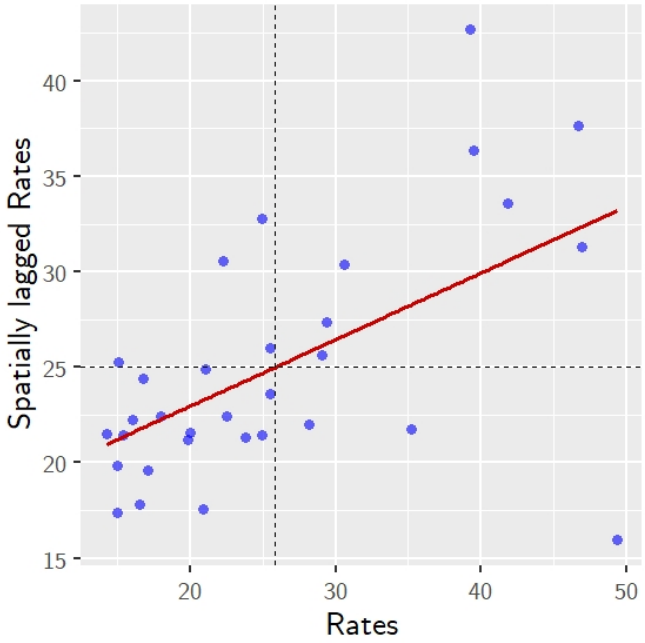


Figure 2.2: Moran's scatterplot for the variable Rates.

To further confirm the previous conclusions, we have also applied Geary's C test (Geary, 1954) to the variable Rates, obtaining a value for the C statistic of C=0.5490, with a p-value=0.0017. Taking into account that significant values for the C statistic of 1 indicate no spatial autocorrelation for this test, and that values between 0 and 1 indicate positive spatial autocorrelation in the variable under study, the results obtained for our case (i.e., C=0.5490) suggest further evidence that the variable is not randomly distributed across the regions, but positively correlated.

2.6.1 Fitting of the spatial conditional overdispersion models

We assume that the number of children under one year of age who died in 2005 in the i-th region; that is, the variable ND_i, follows a Poisson distribution with mean μ_i. As we are interested in modelling the mortality rates, we will consider that $\text{E}(\text{ND}_i) = \mu_i$, with $\mu_i = \text{NB}_i\lambda_i$, where NB_i is the total number of births in 2005 (i.e., the variable NB), and λ_i represents the corresponding mortality rate. Therefore, we include the logarithm of the total number of births $\log(\text{NB}_i)$ in the linear predictor as an offset variable, so that we can write this model as:

$$\log(\mu_i) = \log(\text{NB}_i) + \beta_0 + \beta_1\text{Viol}_i + \beta_2\text{IBN}_i + \beta_3\text{Rec}_i + \beta_4\text{HE}_i + \beta_5\text{Vac}_i \qquad (2.21)$$

Table 2.2 includes the corresponding parameter estimates for the fitting of the Poisson model with regression structure given in equation (2.21), together with its standard deviations and 95% credible intervals in parenthesis, when fitted to the Colombia infant mortality rates data set. For this model, the resulting information criteria values where DIC = 491.7 and WAIC = 524.0. Based on the estimations for the regression coefficients, we can observe that all the explanatory variables, except HE, are statistically significant, as none of their 95% credible intervals includes the value zero. However, these reported results should not be taken as an indication of a good model's fit, mainly because it could be a consequence of fitting the model without taking into account the possible existence of overdispersion. In fact, if we plot the estimated mean and variance obtained by the fitting of this model (see Figure 2.3), where the red line represents that its mean equals its variance, we notice that most of the estimated variance values from the Poisson model are larger than the mean, which may be a clear indication that the assumption of equidispersion of the Poisson distribution does not seem to hold for this data set.

Table 2.2: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the different models fitted to the Colombian infant mortality rates data set.

		Poisson	**Spatial conditional Poisson**	**Negative binomial**	**Spatial conditional negative binomial**	**Normal Poisson**	**Spatial conditional normal Poisson**
Intercept	Mean	-4.2960	-4.6131	-4.4647	-5.0672	-4.4330	-4.8410
	SD	(0.0902)	(0.1188)	(0.2109)	(0.2273)	(0.3171)	(0.2900)
	CI	(-4.4710,-4.1210)	(-4.8500,-4.3680)	(-4.9050,-4.1010)	(-5.5530,-4.4670)	(-5.0610,-3.8189)	(-5.4120,-4.3220)
Viol	Mean	0.0066	0.0065	0.0118	0.0112	0.0109	0.0092
	SD	(0.0015)	(0.0015)	(0.0056)	(0.0058)	(0.0060)	(0.0060)
	CI	(0.0036,0.0096)	(0.0035,0.0093)	(8.860e-04,0.0229)	(1.650e-04,0.0229)	(-9.898e-04,0.0230)	(-0.0019,0.0217)
IBN	Mean	0.0155	0.0165	0.0153	0.0167	0.0155	0.0159
	SD	(6.076e-04)	(6.419e-04)	(0.0024)	(0.0023)	(0.0025)	(0.0023)
	CI	(0.0143,0.0167)	(0.0153,0.0178)	(0.0104,0.0200)	(0.0121,0.0211)	(0.0104,0.0204)	(0.0113,0.0203)
Rec	Mean	-4.953e-04	-5.363e-04	-9.818e-04	-7.407e-04	-0.0011	-8.113e-04
	SD	(1.279e-04)	(1.278e-04)	(5.822e-04)	(5.615e-04)	(5.989e-04)	(5.656e-04)
	CI	(-7.454e-04,-2.460e-04)	(-7.881e-04,-2.819e-04)	(-0.0021,1.756e-04)	(-0.0018,3.747e-04)	(-0.0023,1.118e-04)	(-0.0020,2.760e-04)
HE	Mean	-0.0013	3.117e-04	-0.0058	3.353e-04	-0.0054	-7.355e-04
	SD	(9.218e-04)	(0.0010)	(0.0042)	(0.0045)	(0.0044)	(0.0047)
	CI	(-0.0031,4.562e-04)	(-0.0017,0.0023)	(-0.0143,0.0025)	(-0.0085,0.0091)	(-0.0140,0.0033)	(-0.0099,0.0091)
Vac	Mean	-0.0036	-0.0029	-0.0011	-5.550e-04	-0.0014	-0.0016
	SD	(9.417e-04)	(9.291e-04)	(0.0028)	(0.0028)	(0.0032)	(0.0031)
	CI	(-0.0054,-0.0017)	(-0.0047,-0.0011)	(-0.0065,0.0045)	(-0.0061,0.0051)	(-0.0076,0.0050)	(-0.0083,0.0043)
ρ	Mean	-	0.0095	-	0.0169	-	0.0149
	SD	-	(0.0025)	-	(0.0065)	-	(0.0068)
	CI	-	(0.0045,0.0144)	-	(0.0041,0.0295)	-	(6.637e-04,0.0290)
τ	Mean	-	-	41.5803	46.2603	0.0277	0.0243
	SD	-	-	(13.9969)	(15.4866)	(0.0102)	(0.0088)
	CI	-	-	(20.0300,73.6512)	(21.9997,82.7505)	(0.0138,0.0531)	(0.0121,0.0460)
	DIC	491.7	480.4	362.2	360.1	308.0	307.3
	WAIC	524.0	519.6	361.8	360.3	302.3	301.3

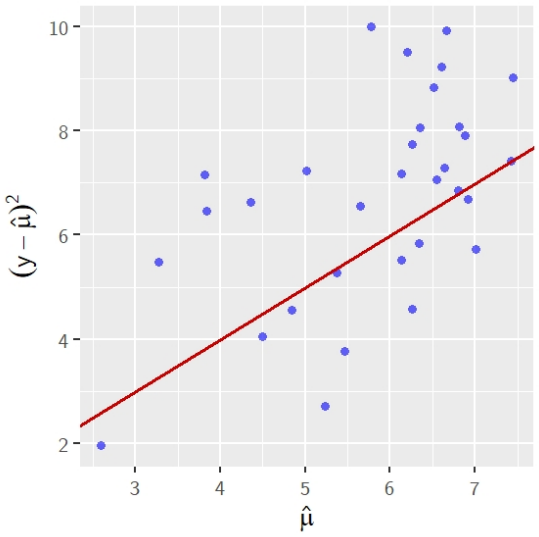


Figure 2.3: Estimated mean and variance scatterplot for the Poisson results after fitting model (2.21) to infant mortality rates in Colombia.

Therefore and based on the possible existence of overdispersion, we have fitted the overdispersion models mentioned in the previous section, starting with the normal Poisson model with regression structure given by equation (2.22), and the negative binomial model with regression structure as in the Poisson model in equation (2.21). These models show considerable improvements in their DIC and WAIC values when compared to those for the Poisson model. That is, resulting values were DIC = 308.0 and WAIC = 302.3 for the normal Poisson model, and DIC = 362.2 and WAIC = 361.8 for the negative binomial model. Moreover, if we carefully observe the results for the 95% credible intervals for the estimated coefficients after fitting the normal Poisson model, we notice that, except for IBN, all the other variables are not statistically significant. In the case of the fitting of the negative binomial model, only the variables IBN and Viol are statistically significant. These results are justified because these models have taken into account the overdispersion in the data set under study and, therefore, these credible intervals become wider than those obtained when fitting the Poisson model.

$$\begin{aligned}\log(\mu_i) = &\log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{HE}_i + \beta_5 \text{Vac}_i \\ &+ \nu_i, \quad \nu_i \sim N(0, \tau),\ \tau > 0\end{aligned} \tag{2.22}$$

In addition, to be able to account for the possible spatial dependence present in the data set under study, we have also fitted the spatial conditional Poisson and negative binomial models, with regression structures given by equation (2.23), where the spatial lag of the variable Rates is also included in these models. That is, we have that:

$$\begin{aligned}\log(\mu_i) = &\log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{HE}_i + \beta_5 \text{Vac}_i \\ &+ \rho \mathbf{W}_i \mathbf{Rates}\end{aligned} \tag{2.23}$$

Finally, we have also fitted the spatial conditional normal Poisson model with regression structure given in equation (2.24) and reported information criteria values of DIC = 307.3 and WAIC = 301.3, these being the lowest information criteria values obtained so far when compared to the rest of the models. In addition, the estimated value for the spatial lag parameter, ρ, was $\hat{\rho} = 0.0149(0.0068)$, with the value zero not included in its 95% credible interval, a clear evidence for the existence of positive spatial autocorrelation in the data. Moreover, the estimated dispersion parameter was $\hat{\tau} = 0.0243(0.0088)$, also indicating that, besides explaining the spatial dependence, this model also captures the overdispersion, allowing for an extra variation given that a specific random effect has been included for this purpose in this model.

$$\begin{aligned}\log(\mu_i) = &\log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{HE}_i + \beta_5 \text{Vac}_i \\ &+ \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \quad \nu_i \sim N(0, \tau),\ \tau > 0\end{aligned} \tag{2.24}$$

From the results reported in Table 2.2, the lowest DIC and WAIC values were obtained for the spatial conditional normal Poisson model, which was considered therefore the best fitting model. A variable selection process was performed afterwards for this model by fitting reduced versions and comparing the information criteria values obtained for the different models. The results for some of the fitted models shown in Table 2.3 indicate that the model with the lowest information criteria values, i.e., DIC = 306.8 and WAIC = 301.2, is the one which includes the variables Viol, IBN and Rec. However, in this model, the variables Viol and Rec are not statistically significant, since zero is contained in their 95% credible interval for the estimated coefficients.

The model including the variables IBN and Rec with regression structure given by equation (2.25), which provided information criteria values of DIC = 307.4 and WAIC = 302.3, could be a good candidate for fitting the data under study. In this model, according to its 95% credible intervals, all of the estimated coefficients are statistically significant. The estimated coefficient for the variable IBN was $\hat{\beta}_1 = 0.0167(0.0018)$, which indicates that, according to the model fitted to this data, infant mortality rates have a statistically significant positive association with the percentage of people with unsatisfied basic needs. With regard to the estimated coefficient for the variable Rec, it was $\hat{\beta}_2 = -0.0011(4.982\text{e}-04)$, a fact that could be an indication of a statistically significant negative association between infant mortality rates and the amount of resources provided for academic achievement. In addition, the estimated coefficient for the spatial term has a positive value of $\hat{\rho} = 0.0151(0.0061)$, again a clear evidence of the presence of positive spatial autocorrelation in the data.

$$\log(\mu_i) = \log(\text{NB}_i) + \beta_0 + \beta_1 \text{IBN}_i + \beta_2 \text{Rec}_i + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i,\ \nu_i \sim N(0, \tau),\ \tau > 0 \tag{2.25}$$

Table 2.3: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for some of the reduced versions of the spatial conditional normal Poisson model fitted to the Colombian infant mortality rates data set.

		Intercept	**Viol**	**IBN**	**Rec**	**Vac**	ρ	τ
DIC= 307.3	Mean	-4.8298	0.0087	0.0165	-9.389e-04	-0.0017	0.0145	0.0230
WAIC= 301.7	SD	(0.3218)	(0.0055)	(0.0018)	(5.362e-04)	(0.0029)	(0.0061)	(0.0084)
	CI	(-5.4881,-4.2110)	(-0.0019,0.0197)	(0.0127,0.0201)	(-0.0020,1.010e-04)	(-0.0072,0.0042)	(0.0022,0.0264)	(0.0111,0.0442)
DIC= 307.8	Mean	-4.5858	-	0.0165	-0.0011	-8.340e-04	0.0153	0.0247
WAIC= 302.4	SD	(0.2861)	-	(0.0019)	(5.353e-04)	(0.0030)	(0.0063)	(0.0086)
	CI	(-5.1450,-4.0230)	-	(0.0127,0.0204)	(-0.0022,-9.296e-05)	(-0.0069,0.0051)	(0.0028,0.0277)	(0.0125,0.0450)
DIC= 307.5	Mean	-5.1301	0.0116	0.0158	-	-9.405e-04	0.0183	0.0250
WAIC= 301.8	SD	(0.2881)	(0.0054)	(0.0019)	-	(0.0030)	(0.0059)	(0.0085)
	CI	(-5.6960,-4.5760)	(0.0011,0.0223)	(0.0121,0.0196)	-	(-0.0068,0.0050)	(0.0068,0.0301)	(0.0125,0.0457)
DIC= 306.8	Mean	-4.9505	0.0083	0.0165	-8.274e-04	-	0.0150	0.0225
WAIC= 301.2	SD	(0.2588)	(0.0055)	(0.0018)	(5.446e-04)	-	(0.0060)	(0.0080)
	CI	(-5.4700,-4.4500)	(-0.0026,0.0194)	(0.0128,0.0200)	(-0.0019,2.409e-04)	-	(0.0030,0.0270)	(0.0113,0.0421)
DIC= 307.8	Mean	-4.8702	-	0.0157	-	3.069e-04	0.0207	0.0285
WAIC= 302.0	SD	(0.2594)	-	(0.0020)	-	(0.0031)	(0.0061)	(0.0095)
	CI	(-5.3800,-4.3700)	-	(0.0118,0.0197)	-	(-0.0057,0.0064)	(0.0090,0.0324)	(0.0147,0.0505)
DIC= 307.4	Mean	-4.6479	-	0.0167	-0.0011	-	0.0151	0.0233
WAIC= 302.3	SD	(0.1660)	-	(0.0018)	(4.982e-04)	-	(0.0061)	(0.0080)
	CI	(-4.9660,-4.3240)	-	(0.0132,0.0202)	(-0.0021,-1.824e-04)	-	(0.0033,0.0272)	(0.0119,0.0430)
DIC= 306.9	Mean	-5.1912	0.0114	0.0159	-	-	0.0184	0.0241
WAIC= 301.4	SD	(0.2136)	(0.0053)	(0.0018)	-	-	(0.0059)	(0.0082)
	CI	(-5.6120,-4.7650)	(0.0010,0.0221)	(0.0124,0.0193)	-	-	(0.0067,0.0298)	(0.0123,0.0437)
DIC= 308.2	Mean	-4.8478	-	0.0156	-	-	0.0208	0.0272
WAIC= 303.4	SD	(0.1439)	-	(0.0019)	-	-	(0.0059)	(0.0091)
	CI	(-5.1390,-4.5740)	-	(0.0119,0.0194)	-	-	(0.0094,0.0323)	(0.0143,0.0490)

It is worth mentioning that the model providing information criteria values of DIC = 306.9 and WAIC = 301.4, which corresponds to the model containing the variables IBN and Viol, with regression structure given by equation (2.26), should also be considered as a good candidate for fitting the data set under study where, based on their 95% credible intervals, all of the estimated coefficients are statistically significant. The estimated coefficient for the variable IBN has a value of $\hat{\beta}_2 = 0.0159(0.0018)$, a very similar value to that obtained in the previous model for this specific coefficient. In addition, the estimated coefficient for the variable Viol was $\hat{\beta}_1 = 0.0114(0.0053)$, implying that infant mortality rates may have a statistically significant positive relation with the percent of women who suffered any type of physical abuse from their current partner. The spatial coefficient estimated value was $\hat{\rho} = 0.0184(0.0059)$, which represents further evidence of the presence of positive spatial autocorrelation in the data.

$$\log(\mu_i) = \log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i,\ \nu_i \sim N(0, \tau),\ \tau > 0 \quad (2.26)$$

In order to provide some information that could be useful for the readers about the size of the effects of the explanatory variables in the models, we have computed estimations of the marginal effects at the means for these variables in the spatial conditional normal Poisson models in equations (2.25) and (2.26) and included the results in Table 2.4. For the variable IBN in the model in equation (2.25), this estimated marginal effect was 3.9981e-04, which means that with an increment of 1 percentage point in the variable IBN, the mortality rate would be increased by a 0.039981%, when the other variables are set at their mean value. Although this represents a very small effect on the infant mortality rate, according to the credible interval (3.1469e-04,4.8580e-04), it is indeed significant. Moreover, small effects, but significant according to their credible intervals, were also obtained for the variable Rec in this model and for the variables Viol and IBN for the model in equation (2.26).

Table 2.4: Marginal effects at the means for the spatial conditional normal Poisson models in equations (2.25) and (2.26) fitted to the Colombian infant mortality rates data set.

		Mean	SD	95% CI
Model in equation (2.25)	**IBN**	3.9981e-04	(4.3043e-05)	(3.1469e-04,4.8580e-04)
	Rec	-2.6978e-05	(1.2012e-05)	(-5.0980e-05,-4.2588e-06)
Model in equation (2.26)	**Viol**	2.7189e-04	(1.2728e-04)	(2.4744e-05,5.2951e-04)
	IBN	3.7891e-04	(4.2576e-05)	(2.9550e-04,4.6250e-04)

In order to assess the convergence of the MCMC chains for the fitted models, we have computed the effective sample size for estimating the means, (i.e., N_{eff}), and the potential scale reduction factor (i.e., $\hat{\text{R}}$) for each parameter's MCMC chain. For brevity

of exposition, results are reported in Table 2.5 only for the models in equations (2.25) and (2.26).

The values N_{eff} represent the equivalent number of independent samples in each parameter's MCMC chain. It is often considered that a minimum of 100 independent simulations is a sufficient number for performing reasonable posterior inference (Gelman et al., 2013). Therefore, a desirable value for N_{eff} would be any number of at least 100. We have simulated 3 Markov chains, with 10000 iterations and discarded the first 5000 samples from each one, which leaves us with a total of 15000 samples. Taking this into account and considering that the values of N_{eff} for the chains of all the estimated parameters obtained from the fitting of models in equations (2.25) and (2.26), reported in Table 2.5, are larger than 1000, we can conclude that we have enough simulations to correctly approximate the target distribution and that the correlation in the chains should not affect posterior inference on the parameters. Moreover, the values $\hat{R}$ are an estimation of the potential scale reduction factor (Gelman and Rubin, 1992). Values closer to 1 indicate convergence of the chain, whereas when values larger than 1.1 are obtained, it is considered that further simulations need to be computed in order to improve posterior inference. In this specific application, all of the $\hat{R}$ values are approximately one, which suggests that the chains for all of the parameters have successfully converged to the target distribution.

Table 2.5: Convergence diagnostics for the spatial conditional normal Poisson models in equations (2.25) and (2.26) fitted to the Colombian infant mortality rates data set.

		Intercept	**Viol**	**IBN**	**Rec**	ρ	τ
Model in equation (2.25)	N_{eff}	13000	-	15000	11000	7300	1500
	$\hat{R}$	1.0011	-	1.0010	1.0011	1.0012	1.0023
Model in equation (2.26)	N_{eff}	2700	9600	3400	-	7800	13000
	$\hat{R}$	1.0016	1.0011	1.0015	-	1.0012	1.0011

As a summary of this section, we believe it may be convenient to mention that the results obtained from the fitting of these models in equations (2.25) and (2.26) are consistent with the previous results obtained in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) for a similar data set. This fact is an illustration of the usefulness of the spatial conditional overdispersion models for being able to explain spatial dependence and overdispersion in applications for real data sets.

2.6.2 Sensitivity analysis for the precision of the prior distributions

The prior distribution $G(\alpha, \alpha)$, assumed for the inverse of the variance parameter τ for the random effects, that is, for the precision parameter $\psi = 1/\tau$, can have a significant effect on the inferential process, so inference may be quite sensitive to the choice of the fixed parameters α in this gamma distribution (Gelman, 2006). Therefore, in order to better assess this effect and select prior distributions where it is limited or at least

controlled for, we have performed a sensitivity analysis for the best fitting model in the previous section, the spatial conditional normal Poisson model with regression structure given by equation (2.25), considering different possible values for α in the gamma prior distribution $\psi \sim G(\alpha, \alpha)$ for the precision parameter in the random effects, from $\alpha = 0.1$ to $\alpha =$ 1e-08.

Results are included in Table 2.6, where we can observe that, for values of $\alpha =$ 1e-04, as well as for smaller values, there were only small differences in the estimates in the third decimal place for a few cases. If we also look at the posterior marginal densities for the estimated precision parameter ψ (see right panel in Figure 2.4), no changes are observed when setting the values for α in the prior distributions for the last five values considered here (i.e., from $\alpha =$ 1e-04 up to $\alpha =$ 1e-08). Therefore, we believe the choice of the value $\alpha =$ 1e-04 is well justified and, in addition, it does not appear to represent any undesirable influence on the inferential process considered here.

Table 2.6: Posterior means for parameter estimates together with standard deviations, DIC and WAIC values, for the spatial conditional normal Poisson model in the analysis of the Colombian infant mortality rates data set with different prior distributions for the precision parameter of the random effects.

	Intercept	**IBN**	**Rec**	ρ	τ	DIC	WAIC
$\alpha = 0.1$	-4.6488	0.0167	-0.0012	0.0153	0.0332	306.5	298.2
	(0.1956)	(0.0022)	(6.020e-04)	(0.0073)	(0.0104)		
$\alpha = 0.01$	-4.6455	0.0167	-0.0011	0.0151	0.0245	307.2	301.6
	(0.1727)	(0.0018)	(5.154e-04)	(0.0064)	(0.0083)		
$\alpha = 0.001$	-4.6469	0.0167	-0.0011	0.0151	0.0234	307.4	302.3
	(0.1680)	(0.0018)	(5.017e-04)	(0.0062)	(0.0080)		
$\alpha =$ 1e-04	-4.6479	0.0167	-0.0011	0.0151	0.0233	307.4	302.3
	(0.1660)	(0.0018)	(4.982e-04)	(0.0061)	(0.0080)		
$\alpha =$ 1e-05	-4.6480	0.0167	-0.0011	0.0151	0.0233	307.4	302.3
	(0.1661)	(0.0018)	(4.981e-04)	(0.0061)	(0.0080)		
$\alpha =$ 1e-06	-4.6480	0.0167	-0.0011	0.0151	0.0233	307.4	302.3
	(0.1660)	(0.0018)	(4.981e-04)	(0.0061)	(0.0080)		
$\alpha =$ 1e-07	-4.6480	0.0167	-0.0011	0.0151	0.0233	307.4	302.3
	(0.1660)	(0.0018)	(4.981e-04)	(0.0061)	(0.0080)		
$\alpha =$ 1e-08	-4.6480	0.0167	-0.0011	0.0151	0.0233	307.4	302.3
	(0.1660)	(0.0018)	(4.981e-04)	(0.0061)	(0.0080)		

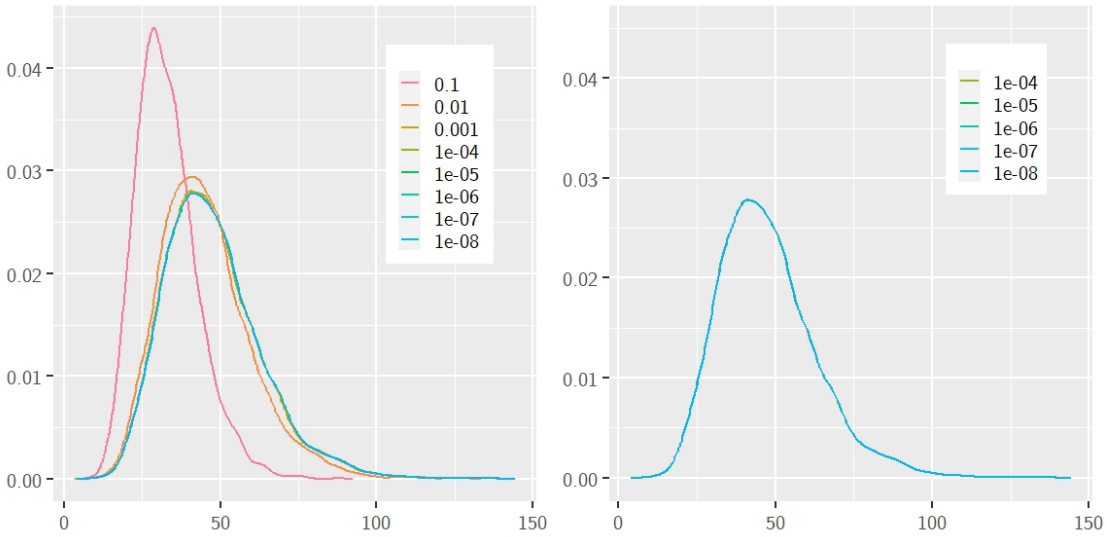


Figure 2.4: Posterior marginal distributions for the precision parameter $\psi = 1/\tau$, the inverse of the variance parameter τ for the random effects, for different values of α, where $\psi \sim G(\alpha, \alpha)$.

2.6.3 Fitting of the generalized spatial conditional normal Poisson model

For all of the aforementioned fitted models, it is assumed that the dispersion parameter (i.e., τ) is constant, a fact that could not always be a reasonable assumption. Hence, we will allow the dispersion parameter to vary with some explanatory variables by considering the so-called generalized spatial conditional models (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018). After fitting the generalized spatial conditional normal Poisson model for the infant mortality data in Colombia and, performing a variable selection process, we have concluded that the best fitting model was the one where the mean model contains the spatial lag and the dispersion model includes the variable IBN, so that:

$$\begin{aligned} \log(\mu_i) &= \log(\text{NB}_i) + \beta + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \quad \nu_i \sim N(0, \tau_i),\ \tau_i > 0, \\ \log(\tau_i) &= \gamma_0 + \gamma_1 \text{IBN}_i \end{aligned} \tag{2.27}$$

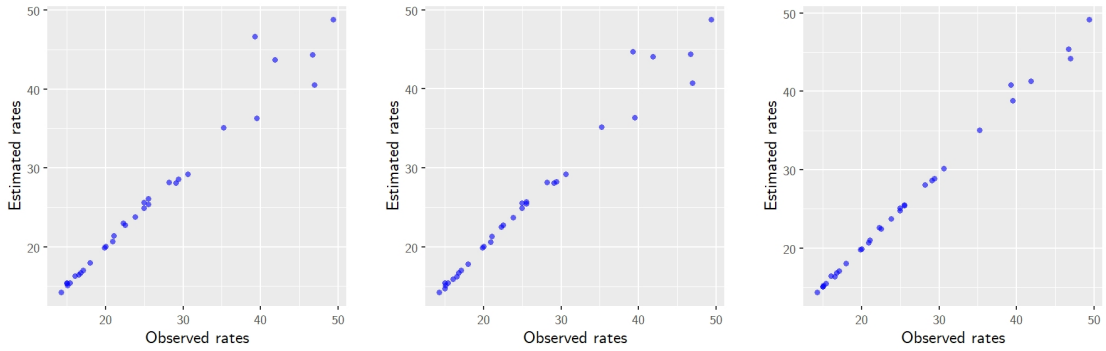
Table 2.7 reports the results from the fitting of the model in equation (2.27). The DIC = 308.0 and WAIC = 299.1 values from the fitting of this model are quite similar to those obtained when fitting the spatial conditional normal Poisson model and, therefore, this does not result in a significant improvement in the model's fitting. However, we believe that there are some very interesting features that can be noticed from these two fittings. On the one hand, the mean of the estimated coefficient for the variable IBN in the dispersion model was $\hat{\gamma}_1 = 0.0430(0.0145)$ and, according to its 95% credible interval, this variable is statistically significant, which could indicate that the dispersion varies

according to the variable IBN in such a way that in regions where the percentage of the population not having basic services satisfactorily attended is larger, the dispersion increases. On the other hand, the spatial parameter estimate in the mean model has a mean with a positive value of $\hat{\rho} = 0.0431(0.0093)$ and, in addition, according to its 95% credible interval, it is also statistically significant. The significance of ρ constitutes a clear evidence for the presence of positive spatial autocorrelation, which is appropriately captured by the fitting of this model. Taking these facts into consideration, the fitting of these generalized spatial conditional models would be well justified. Besides being able to model the spatial dependence in the data, it also explains their nonconstant dispersion.

Table 2.7: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values, for the generalized spatial conditional normal Poisson model fitted to the Colombian infant mortality rates data set.

	μ **Intercept** - β	ρ	τ **Intercept** - γ_0	**IBN**
Mean	-4.9262	0.0431	-4.2186	0.0430
SD	(0.2275)	(0.0093)	(0.6165)	(0.0145)
CI	(-5.3620,-4.4780)	(0.0249,0.0608)	(-5.4200,-2.9998)	(0.0161,0.0727)
DIC= 308.0 WAIC= 299.1				

As part of the posterior predictive checks required to better assess the fit of a model, Figure 2.5 includes the scatterplots for the observed mortality rates versus the predicted mortality rates for some of the fitted models. From the plots in Figures 2.5(a) and 2.5(b), we can mention that the spatial conditional normal Poisson models in equations (2.25) and (2.26) show high accuracy in the prediction of mortality rates for observed rates under 40 whereas, for some of the values larger than 40, predictions slightly deviate from the observed values. The scatterplot included in Figure 2.5(c) for the generalized spatial conditional normal Poisson model in equation (2.27) shows a considerable improvement, mainly because values of the mortality rates larger than 40 are now more accurately predicted.



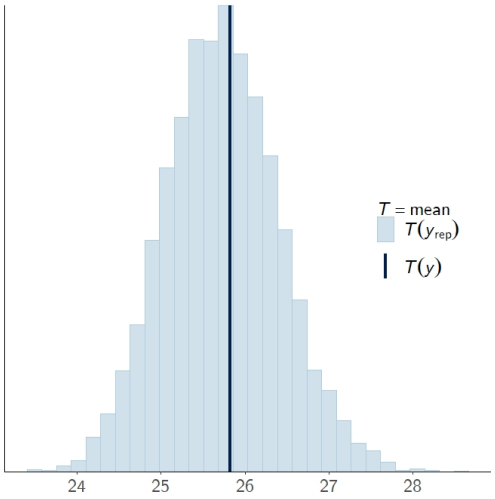
(a) Spatial conditional normal Poisson model in equation (2.25) fitted in OpenBUGS.

(b) Spatial conditional normal Poisson model in equation (2.26) fitted in OpenBUGS.

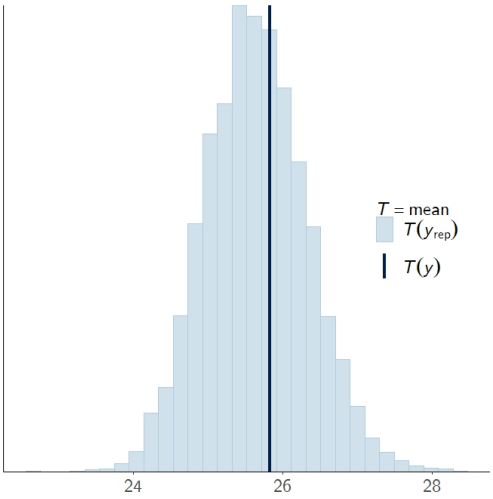
(c) Generalized spatial conditional normal Poisson model in equation (2.27) fitted in OpenBUGS.

Figure 2.5: Scatterplots for the observed versus the predicted rates obtained from some of the models fitted to the Colombian infant mortality rates data set, fitted in OpenBUGS.

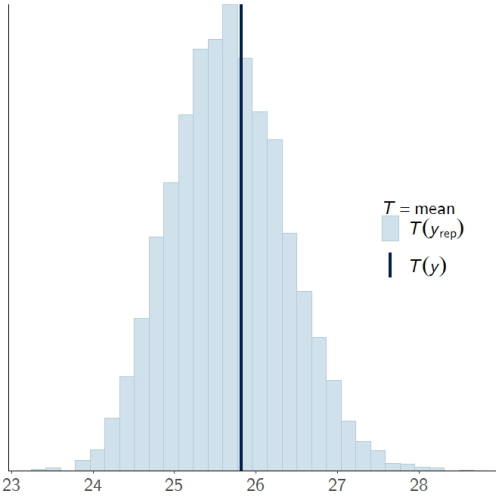
Finally, Figure 2.6 includes histograms of the posterior predictive distributions for the means over replicated simulations of the mortality rates, estimated from the fitting of the spatial conditional normal Poisson model in equation (2.25) (see Figure 2.6(a)) and equation (2.26) (see Figure 2.6(b)), and for the generalized spatial conditional normal Poisson model in equation (2.27) (see Figure 2.6(c)). Note that the dark blue vertical line represents the value of the mean for the observed data. As we can see from these figures, they show how the means of the replicated data vary with respect to the actual mean of the observed values, a fact suggesting that the considered models offer an adequate fit to the data set under study.



(a) Spatial conditional normal Poisson model in equation (2.25).



(b) Spatial conditional normal Poisson model in equation (2.26).



(c) Generalized spatial conditional normal Poisson model in equation (2.27).

Figure 2.6: Posterior predictive checks for some of the fitted models to the Colombia infant mortality rates data set.

2.6.4 Comparisons to the BYM model

In order to be able to compare the performance between the previous models and the BYM model, we have decided to select the best fitting model we have so far, which is the spatial conditional normal Poisson model, and compare its fitting to that of the BYM model. In order to do so, we have fitted the BYM model with regression structure given by:

$$\log(\mu_i) = \log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{Vac}_i + \nu_i + \eta_i, \qquad (2.28)$$

where $\nu_i,\ i = 1, \ldots, n$, is a set of normally distributed random effects so that $\text{Var}(\nu_i) = \tau$, with $\tau > 0$, and $\eta_i,\ i = 1, \ldots, n$, is a set of spatially structured random effects following an intrinsic CAR prior distribution, with variance parameter $\tau_\eta > 0$. As in the previous models, we assume noninformative normal priors for the regression parameters; that is, $N(0, 1 \times 10^{-5})$ and, for the precision parameters $\psi = 1/\tau$ and $\psi_\eta = 1/\tau_\eta$, we also assume noninformative gamma $G(1 \times 10^{-4}, 1 \times 10^{-4})$ prior distributions. Finally, for the BYM model, we assume the same first order neighbourhood structure as before.

The fitting of this model in OpenBUGS is very unstable, since it showed drastic changes if we slightly changed the assumed values for the prior distributions, especially in the estimated values for the variance parameter τ_η of the spatially structured random effects. In addition, in most cases, the effective number of parameters resulted in negative values, which could be a clear sign of conflict between the prior distribution and the data (Spiegelhalter et al., 2002).

For the reasons we have just described here, and in order to be able to compare the performance of both models when fitting them to the data set under study, we have fitted both the spatial conditional normal Poisson and the BYM models in R-INLA, by setting the same values for the prior distributions, so that more stable results were provided. In particular, the regression structure specified for the conditional mean of the spatial conditional normal Poisson model was:

$$\log(\mu_i) = \log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{Vac}_i + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \quad (2.29)$$

where $\nu_i,\ i = 1, \ldots, n$, is a set of normally distributed random effects with $\text{Var}(\nu_i) = \tau > 0$. Results for the estimation of the spatial conditional normal Poisson model in equation (2.29) are reported in Table 2.8, and those for the BYM model in equation (2.28) are included in Table 2.9, as well as estimations of reduced versions of these models.

Table 2.8: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the reduced versions of the spatial conditional normal Poisson model fitted in INLA to the Colombian infant mortality rates data set.

		Intercept	**Viol**	**IBN**	**Rec**	**Vac**	ρ	τ
DIC= 307.3	Mean	-4.8473	0.0087	0.0164	-8.938e-04	-0.0015	0.0148	0.0231
WAIC= 301.9	SD	(0.3241)	(0.0057)	(0.0018)	(5.425e-04)	(0.0029)	(0.0061)	(0.0082)
	CI	(-5.4930,-4.2117)	(-0.0024,0.0200)	(0.0127,0.0200)	(-0.0020,1.775e-04)	(-0.0073,0.0043)	(0.0028,0.0269)	(0.0114,0.0433)
DIC= 307.7	Mean	-4.5846	-	0.0165	-0.0011	-8.525e-04	0.0153	0.0243
WAIC= 302.3	SD	(0.2806)	-	(0.0019)	(5.295e-04)	(0.0030)	(0.0062)	(0.0084)
	CI	(-5.1419,-4.0332)	-	(0.0128,0.0202)	(-0.0022,-1.062e-04)	(-0.0067,0.0051)	(0.0031,0.0276)	(0.0123,0.0450)
DIC= 307.4	Mean	-5.1410	0.0116	0.0158	-	-8.776e-04	0.0185	0.0250
WAIC= 301.9	SD	(0.2810)	(0.0056)	(0.0019)	-	(0.0030)	(0.0058)	(0.0086)
	CI	(-5.7002,-4.5905)	(6.159e-04,0.0227)	(0.0121,0.0195)	-	(-0.0068,0.0051)	(0.0070,0.0301)	(0.0127,0.0461)
DIC= 306.9	Mean	-4.9464	0.0083	0.0166	-8.587e-04	-	0.0149	0.0224
WAIC= 301.5	SD	(0.2573)	(0.0055)	(0.0018)	(5.312e-04)	-	(0.0060)	(0.0078)
	CI	(-5.4577,-4.4400)	(-0.0026,0.0193)	(0.0130,0.0201)	(-0.0019,1.898e-04)	-	(0.0030,0.0267)	(0.0113,0.0416)
DIC= 307.9	Mean	-4.8797	-	0.0158	-	3.763e-04	0.0208	0.0285
WAIC= 302.3	SD	(0.2624)	-	(0.0020)	-	(0.0031)	(0.0060)	(0.0095)
	CI	(-5.4002,-4.3646)	-	(0.0118,0.0197)	-	(-0.0058,0.0066)	(0.0089,0.0328)	(0.0148,0.0516)
DIC= 307.3	Mean	-4.6489	-	0.0166	-0.0011	-	0.0153	0.0234
WAIC= 302.0	SD	(0.1671)	-	(0.0018)	(5.120e-04)	-	(0.0061)	(0.0080)
	CI	(-4.9791,-4.3188)	-	(0.0130,0.0202)	(-0.0021,-1.139e-04)	-	(0.0033,0.0274)	(0.0120,0.0429)
DIC= 307.1	Mean	-5.1920	0.0113	0.0160	-	-	0.0184	0.0241
WAIC= 301.6	SD	(0.2149)	(0.0054)	(0.0018)	-	-	(0.0058)	(0.0081)
	CI	(-5.6190,-4.7698)	(6.868e-04,0.0220)	(0.0124,0.0195)	-	-	(0.0071,0.0298)	(0.0124,0.0439)
DIC= 307.6	Mean	-4.8535	-	0.0157	-	-	0.0209	0.0272
WAIC= 302.3	SD	(0.1475)	-	(0.0019)	-	-	(0.0059)	(0.0089)
	CI	(-5.1451,-4.5630)	-	(0.0120,0.0195)	-	-	(0.0093,0.0326)	(0.0143,0.0490)

Table 2.9: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the reduced versions of the BYM model fitted in INLA to the Colombian infant mortality rates data set.

		Intercept	**Viol**	**IBN**	**Rec**	**Vac**	τ	τ_η
DIC= 307.9	Mean	-4.4748	0.0067	0.0173	-0.0010	-9.456e-04	0.0224	0.0260
WAIC= 302.7	SD	(0.3265)	(0.0073)	(0.0021)	(6.144e-04)	(0.0031)	(0.0141)	(0.0251)
	CI	(-5.1240,-3.8324)	(-0.0082,0.0207)	(0.0132,0.0213)	(-0.0022,1.819e-04)	(-0.0070,0.0051)	(0.0073,0.0601)	(0.0014,0.0910)
DIC= 308.0	Mean	-4.2825	-	0.0174	-0.0010	-4.280e-04	0.0185	0.0356
WAIC= 303.2	SD	(0.2439)	-	(0.0020)	(6.309e-04)	(0.0030)	(0.0123)	(0.0305)
	CI	(-4.7664,-3.8020)	-	(0.0133,0.0214)	(-0.0023,2.147e-04)	(-0.0062,0.0055)	(0.0052,0.0514)	(0.0044,0.1147)
DIC= 307.9	Mean	-4.5923	0.0058	0.0169	-	1.905e-04	0.0128	0.0668
WAIC= 302.8	SD	(0.3326)	(0.0082)	(0.0022)	-	(0.0030)	(0.0095)	(0.0477)
	CI	(-5.2487,-3.9345)	(-0.0106,0.0216)	(0.0125,0.0211)	-	(-0.0056,0.0060)	(0.0024,0.0377)	(0.0166,0.1946)
DIC= 307.5	Mean	-4.5267	0.0061	0.0174	-9.647e-04	-	0.0199	0.0302
WAIC= 302.3	SD	(0.2620)	(0.0071)	(0.0020)	(5.953e-04)	-	(0.0129)	(0.0267)
	CI	(-5.0379,-4.0010)	(-0.0084,0.0199)	(0.0133,0.0213)	(-0.0021,1.967e-04)	-	(0.0061,0.0544)	(0.0028,0.0991)
DIC= 307.9	Mean	-4.4211	-	0.0171	-	5.190e-04	0.0097	0.0742
WAIC= 303.1	SD	(0.2248)	-	(0.0021)	-	(0.0028)	(0.0076)	(0.0503)
	CI	(-4.8694,-3.9808)	-	(0.0129,0.0212)	-	(-0.0050,0.0062)	(0.0014,0.0295)	(0.0214,0.2088)
DIC= 307.6	Mean	-4.3162	-	0.0174	-9.695e-04	-	0.0170	0.0367
WAIC= 302.9	SD	(0.0889)	-	(0.0020)	(6.078e-04)	-	(0.0113)	(0.0301)
	CI	(-4.4904,-4.1392)	-	(0.0135,0.0213)	(-0.0022,2.002e-04)	-	(0.0047,0.0470)	(0.0055,0.1151)
DIC= 307.6	Mean	-4.5740	0.0057	0.0168	-	-	0.0114	0.0668
WAIC= 302.7	SD	(0.2781)	(0.0079)	(0.0021)	-	-	(0.0085)	(0.0467)
	CI	(-5.1153,-4.0184)	(-0.0102,0.0211)	(0.0126,0.0209)	-	-	(0.0020,0.0337)	(0.0175,0.1914)
DIC= 307.6	Mean	-4.3831	-	0.0170	-	-	0.0091	0.0718
WAIC= 303.8	SD	(0.0794)	-	(0.0020)	-	-	(0.0071)	(0.0475)
	CI	(-4.5384,-4.2244)	-	(0.0129,0.0210)	-	-	(0.0013,0.0275)	(0.0212,0.1992)

If we compare the results from the fitting of the spatial conditional normal Poisson model in OpenBUGS in Table 2.3 with the one in INLA in Table 2.8, we can see that reported results are quite similar. In some cases, the means of the estimated parameters only differ in the third decimal place and, only for very few cases, in the second decimal place. This also occurs for the estimations of the variance parameter for the random effects, where we can only observe differences for some cases in the third decimal place. Even though the estimations are very similar, it must be emphasized that the computation time that R-INLA used for the fitting of each one of the models reported in Table 2.8 was much smaller than the one used by OpenBUGS for the same model. With regard to the information criteria values for the fitted models (i.e., the DIC and WAIC values) reported in Table 2.9 for the BYM model, as well as those reported in Table 2.8 for the spatial conditional normal Poisson model, we can see that the WAIC values indicate a moderately better fit for the spatial conditional normal Poisson model whereas differences in the DIC values are minimal and do not favour any of these models in particular. Comparison between the performance of the spatial conditional normal Poisson and the BYM fitted models can be better explored in Figure 2.7 (Figures 2.7(a), 2.7(b), 2.7(c) and 2.7(d)), where scatterplots of the observed versus the estimated values obtained from the fitting of some of the reduced versions of these models are shown. In addition, there are certain points about the fitting of the BYM model that are worth being mentioned.

For the BYM models reported in Table 2.9, the estimated variance parameters for both the spatially correlated and the uncorrelated random effects show overall quite large standard deviations. For instance, in the case of the model with regression structure given by equation (2.28), the estimated variance parameters for the random effects were $\hat{\tau_\eta} = 0.0260(0.0251)$ and $\hat{\tau} = 0.0224(0.0141)$, making it difficult to interpret them, especially to explain the spatial dependence or the extra-variability present in the data. This also occurs for the model containing only the explanatory variables IBN and Rec, where the means of the estimated variance parameters were $\hat{\tau_\eta} = 0.0367(0.0301)$ and $\hat{\tau} = 0.0170(0.0113)$, respectively.

In any case, all of the BYM fitted models resulted in larger means for the variance parameter estimates τ_η for the spatially structured random effects than those obtained for the means for the estimates of the variance parameter τ for the uncorrelated effects. These results may be a sign indicating that fits of the BYM model for this data are probably giving more importance to the spatial structure assumed by the intrinsic CAR prior for the spatial effects than to the extra-variability represented by the unstructured effects, a fact that will be closer examined when fitting the BYM2 model in the next section. Nevertheless, we have not been able to obtain information about the strength, or even the type of the spatial association from the estimations in this model.

Another issue that we believe is important to mention relates to the interpretation of significance for the estimated coefficients of the regression parameters obtained when fitting the BYM models. According to the 95% credible intervals, the coefficient for the variable IBN is the only one that is statistically significant for each one of the reported models, whereas in the spatial conditional normal Poisson, the variables Rec and Viol

were statistically significant for some cases. Hence, regarding the explanatory variables, a direct interpretation of their statistical significance according to their credible intervals on infant mortality rates can only be made for the variable IBN.

We have also computed the marginal effects at the means for the explanatory variables in some of the fitted BYM models, which are reported in Table 2.10. All of the values reported there appear to be quite small, consistent with the values for the marginal effects at the means obtained from the fitting of the spatial conditional normal Poisson models in OpenBUGS, reported in Table 2.4 in Section 2.6.1. A difference to be noticed in this case is that, according to their 95% credible intervals, the effects for the variables Rec and Viol are not statistically significant for some of the models. For instance, the effect of the variable Viol is not significant in any of the models, and the effect of variable Rec is not significant for the BYM model.

Table 2.10: Marginal effects at the means for some of the models fitted in INLA to the Colombian infant mortality rates data set.

		Mean	SD	95% CI
Spatial conditional normal Poisson model in equation (2.25)	**IBN**	3.9685E-04	(4.3591E-05)	(3.2026E-04,4.7638E-04)
	Rec	-2.7963E-05	(1.0940E-05)	(-5.1158E-05,-9.1056E-06)
Spatial conditional normal Poisson model in equation (2.26)	**Viol**	2.6068E-04	(1.3455E-04)	(-9.9128E-06,4.9089E-04)
	IBN	3.7788E-04	(4.1759E-05)	(2.9543E-04,4.5364E-04)
BYM model including the variables IBN and Rec	**IBN**	4.1873E-04	(8.2243E-05)	(2.8326E-04,6.0400E-04)
	Rec	-2.2106E-05	(1.5282E-05)	(-5.3930E-05,6.3359E-06)
BYM model including the variables Viol and IBN	**Viol**	1.3015E-04	(1.7143E-04)	(-2.0407E-04,4.5079E-04)
	IBN	4.1280E-04	(9.3347E-05)	(2.6677E-04,6.2640E-04)

As a simple visual way of comparing the performance of the spatial conditional normal Poisson and the BYM models fitted in INLA, we can examine the scatterplots of the observed versus the predicted values obtained from the fitting of some of the reduced versions of these models, specifically comparing Figures 2.7(a) and 2.7(b) with Figures 2.7(c) and 2.7(d), respectively. These plots suggest that there are no major differences in the fitting of these two models to this data in terms of predictive accuracy.

From the issues mentioned above, we believe that, for the application considered here, the spatial conditional normal Poisson model may be a better option to model spatial count data following a Poisson distribution when compared to the BYM model. However, leaving aside the fact that there were some serious issues that emerged when fitting the BYM model with the MCMC approach, when fitting them in INLA the DIC and WAIC values were quite similar for both models. Moreover, the spatial conditional normal Poisson model provided information about the type and strength of the spatial

autocorrelation that is present in the data, information that could not be obtained from the fitting of the BYM model. Nevertheless, it is essential that we remember that the true model is not known, nor are the true values of the strength of the spatial association or the real overdispersion parameter.

2.6.5 Comparisons to the BYM2 model

Furthermore, we would like to explore the possibilities that the recently developed BYM2 model offers, given that it allows us to overcome the issue of nonidentifiability in the BYM model. In order to do so, we have fitted the BYM2 model with regression structure given by:

$$\begin{aligned}\log(\mu_i) = &\log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{Vac}_i \\ &+ \frac{1}{\sqrt{\tau_s}} \left(\sqrt{1-\phi_s}\nu_i + \sqrt{\phi_s}\eta_i \right),\end{aligned} \tag{2.30}$$

where $\nu_i,\ i = 1, \ldots, n$, is assumed to be a set of normally distributed random effects with a scaled variance approximately equal to one, and $\eta_i,\ i = 1, \ldots, n$, is a set of spatially structured random effects following an intrinsic CAR prior distribution, each one also with scaled variance approximately equal to one. The unknown precision parameter τ_s captures the variance contribution from the sum of the two random effects, and the mixing parameter ϕ_s controls for the variance contribution of the spatially structured effect $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$, whereas the variance contribution of the unstructured random effect $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_m)^\top$ is explained by $1 - \phi_s$.

As in the previous models, we assume noninformative normal priors for the regression parameters; that is, $N(0, 1 \times 10^{-5})$. For the precision and mixing parameters, complexity priors are specified when following the approach by Simpson et al. (2017). That is, we assume the probability statement that $\text{Prob}(1/\sqrt{\tau_s} > U) = \alpha$ for the parameter τ_s. By considering an upper bound for the marginal standard deviation of 0.2, and using the rule of thumb from the aforementioned proposed approach, we can set $U = 0.2/31$ and $\alpha = 0.01$. As for the mixing parameter ϕ_s, we can specify that $\text{Prob}(\phi_s < 0.5) = 2/3$, which would represent the initial assumption that the proportion of the variability captured by the unstructured random effect $\boldsymbol{\nu}$ is larger than the one explained by the spatially structured effect $\boldsymbol{\eta}$.

Table 2.11 reports the results for the estimation of the BYM2 model in equation (2.30), as well as some of its reduced versions. As can be seen, most of the estimations obtained for the regression parameters by fitting the BYM2 model are quite similar to those obtained by fitting the BYM model reported in Table 2.9 and, in addition, there are no improvements in the information criteria values for any of the fitted models. The parameter ϕ_s, associated to the amount of variability captured by the spatial structure considered in the model, explains more than 30% of the variability in all the fitted models. For instance, for the model in equation (2.30), this parameter's estimate was $\hat{\phi}_s = 0.3091(0.2219)$. In some cases, such as for the model including the variables IBN and Vac, and the model only including the variable IBN, the variance explained by the

spatial effect was more than 50% of the total variability. Therefore, in our view, the main advantage that the BYM2 model offers over the BYM model is the possibility of identifying the spatially structured and the overdispersion effects separately. Moreover, the results obtained show the importance of taking into account the spatial dependence in the data in the sense that it may be explaining a large portion of the overdispersion in the data.

In addition, Table 2.12 includes the estimated marginal effects at the means for some of the fitted BYM2 models. The values obtained for these effects are quite small and greatly resemble those obtained from the BYM models previously fitted and reported in Table 2.10.

Figures 2.7(e) and 2.7(f) show the scatterplots of the observed versus the predicted values of infant mortality rates obtained from the fitting of some of the reduced versions of the BYM2 model in equation (2.30). If we compare these plots to the ones from Figures 2.7(c) and 2.7(d), respectively, no significant differences can be seen. Hence, there are no improvements in terms prediction accuracy in the fitting of the BYM2 over the previously fitted BYM models that can be reported.

Finally, in order to illustrate the performance of the fitted models, Figure 2.8 includes maps of the observed infant mortality rates (i.e., variable Rates, see Figure 2.8(a)) and of the estimated mortality rates obtained by fitting some of the models considered here. As can be seen, the maps of the estimated mortality rates obtained by fitting the spatial conditional normal Poisson model in equation (2.25) (see Figure 2.8(b)) and equation (2.26) (see Figure 2.8(c)) in OpenBUGS, and for the generalized spatial conditional normal Poisson model in equation (2.27) (see Figure 2.8(d)) suggest that these three models, presented in Sections 2.6.1 and in 2.6.3, provided similar estimations of the mortality rates and, in addition, that their fitted values maps are almost identical to the map for the observed values in Figure 2.8(a).

This fact is consistent with their observed versus predicted mortality rates scatterplots (see Figure 2.7). Moreover, the maps for the spatial conditional normal Poisson models considered above, fitted in INLA (see Figures 2.8(e) and 2.8(f), respectively) also appear to be very similar to the map for the observed rates in Figure 2.8(a).

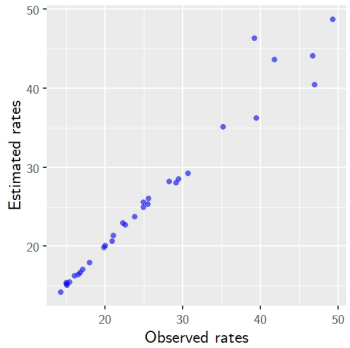
The scatterplots of the observed versus the predicted values of infant mortality rates obtained from the fitting of the BYM (see Figures 2.7(c) and 2.7(d)) and BYM2 models (see Figures 2.7(e) and 2.7(f)) showed no distinguishable differences in the prediction accuracy of infant mortality rates, if we compare them with those obtained from the fitting of the spatial conditional normal Poisson (see Figures 2.7(a) and 2.7(b)). However, the maps for the BYM model including the variables IBN and Rec (see Figure 2.8(g)), and for the BYM model including the variables Viol and IBN (see Figure 2.8(h)) differ from the map of the observed rates in some of the regions. These discrepancies are also displayed between the observed values (see Figure 2.8(a)) and the predicted values for infant mortality rates obtained from the fitting of the BYM2 models considered here (see Figures 2.8(i) and 2.8(j)). These facts seem to suggest that we have not been able to correctly predict the mortality rates in some cases with the fitted BYM and BYM2 models analysed here.

Table 2.11: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the reduced versions of the BYM2 model fitted in INLA to the Colombian infant mortality rates data set.

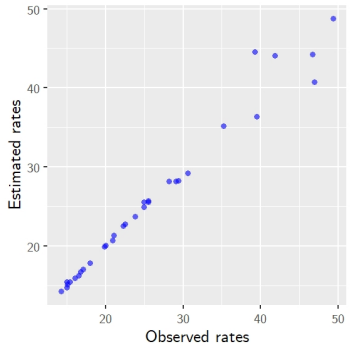
		Intercept	**Viol**	**IBN**	**Rec**	**Vac**	τ_s	ϕ_s
DIC= 307.9	Mean	-4.4738	0.0065	0.0174	-0.0010	-9.460e-04	41.1961	0.3091
WAIC= 302.8	SD	(0.3207)	(0.0071)	(0.0020)	(5.859e-04)	(0.0030)	(13.8614)	(0.2219)
	CI	(-5.1130,-3.8447)	(-0.0076,0.0203)	(0.0134,0.0213)	(-0.0022,1.373e-04)	(-0.0069,0.0051)	(20.1835,74.0994)	(0.0221,0.8088)
DIC= 308.0	Mean	-4.2823	-	0.0176	-0.0011	-4.666e-04	41.5126	0.3569
WAIC= 303.2	SD	(0.2433)	-	(0.0020)	(5.976e-04)	(0.0030)	(13.7300)	(0.2215)
	CI	(-4.7656,-3.8043)	-	(0.0137,0.0215)	(-0.0023,1.069e-04)	(-0.0063,0.0054)	(20.5930,74.0187)	(0.0398,0.8267)
DIC= 308.0	Mean	-4.6458	0.0071	0.0171	-	1.062e-04	36.8669	0.4602
WAIC= 302.9	SD	(0.3236)	(0.0078)	(0.0021)	-	(0.0030)	(11.9851)	(0.2285)
	CI	(-5.2871,-4.0082)	(-0.0085,0.0223)	(0.0129,0.0213)	-	(-0.0058,0.0061)	(18.4969,65.1399)	(0.0785,0.8865)
DIC= 307.5	Mean	-4.5308	0.0061	0.0175	-9.745e-04	-	42.3246	0.3228
WAIC= 302.4	SD	(0.2543)	(0.0069)	(0.0019)	(5.695e-04)	-	(13.9432)	(0.2248)
	CI	(-5.0311,-4.0251)	(-0.0077,0.0196)	(0.0136,0.0213)	(-0.0021,1.411e-04)	-	(21.0400,75.2899)	(0.0248,0.8194)
DIC= 307.9	Mean	-4.4388	-	0.0174	-	5.961e-04	36.9526	0.5309
WAIC= 303.1	SD	(0.2290)	-	(0.0021)	-	(0.0029)	(11.7748)	(0.2105)
	CI	(-4.8969,-3.9920)	-	(0.0133,0.0215)	-	(-0.0051,0.0064)	(18.7905,64.6248)	(0.1449,0.9035)
DIC= 307.7	Mean	-4.3186	-	0.0176	-0.0010	-	42.9158	0.3669
WAIC= 302.9	SD	(0.0875)	-	(0.0019)	(5.775e-04)	-	(13.9586)	(0.2228)
	CI	(-4.4904,-4.1447)	-	(0.0138,0.0214)	(-0.0022,9.322e-05)	-	(21.5381,75.8693)	(0.0429,0.8340)
DIC= 307.7	Mean	-4.6340	0.0070	0.0171	-	-	38.2786	0.4719
WAIC= 302.7	SD	(0.2675)	(0.0075)	(0.0020)	-	-	(12.3129)	(0.2283)
	CI	(-5.1549,-4.0989)	(-0.0081,0.0217)	(0.0131,0.0211)	-	-	(19.3509,67.2822)	(0.0845,0.8923)
DIC= 307.7	Mean	-4.3948	-	0.0173	-	-	38.4058	0.5440
WAIC= 303.1	SD	(0.0788)	-	(0.0020)	-	-	(12.1547)	(0.2094)
	CI	(-4.5496,-4.2380)	-	(0.0133,0.0212)	-	-	(19.6255,66.9478)	(0.1539,0.9095)

Table 2.12: Marginal effects at the means for some of the BYM2 models fitted in INLA to the Colombia infant mortality rates data set.

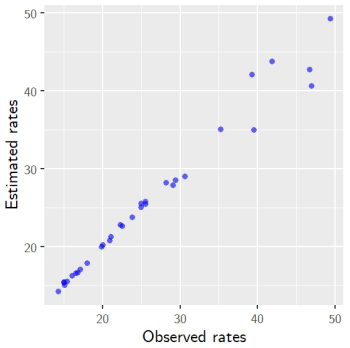
		Mean	SD	95% CI
BYM2 model including the variables IBN and Rec	**IBN**	4.1877E-04	(4.2850E-05)	(3.4240E-04,4.9881E-04)
	Rec	-2.4443E-05	(1.4208E-05)	(-4.9394E-05,7.8678E-07)
BYM2 model including the variables Viol and IBN	**Viol**	1.6102E-04	(1.6771E-04)	(-2.1457E-04,4.3999E-04)
	IBN	4.1087E-04	(5.0080E-05)	(3.1157E-04,5.0627E-04)



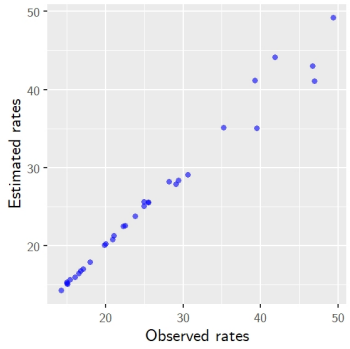
(a) Spatial conditional normal Poisson model in equation (2.25) fitted in INLA.



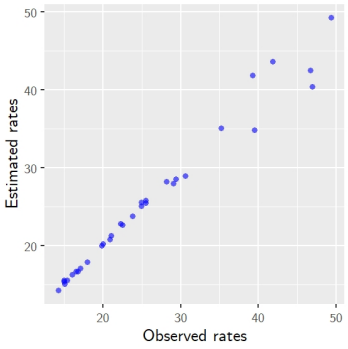
(b) Spatial conditional normal Poisson model in equation (2.26) fitted in INLA.



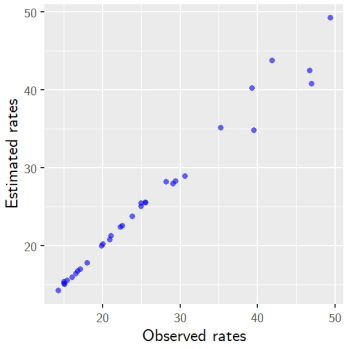
(c) BYM model including the variables IBN and Rec fitted in INLA.



(d) BYM model including the variables Viol and IBN fitted in INLA.

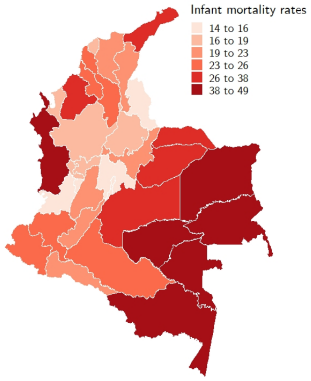


(e) BYM2 model including the variables IBN and Rec fitted in INLA.

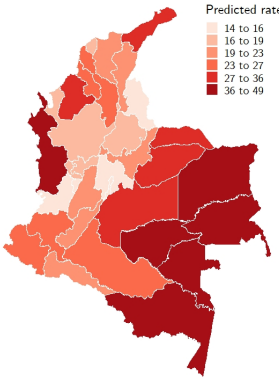


(f) BYM2 model including the variables Viol and IBN fitted in INLA.

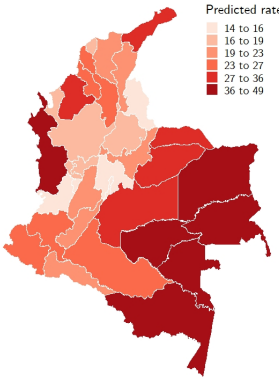
Figure 2.7: Scatterplots for the observed versus the predicted rates obtained from some of the models fitted to the Colombian infant mortality rates data set, fitted in INLA.



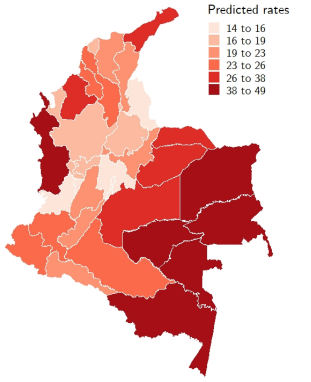
(a) Observed infant mortality rates.



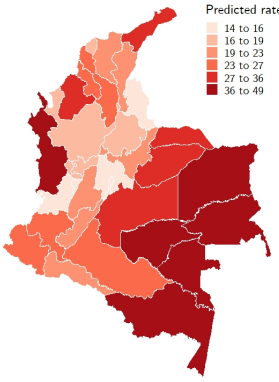
(b) Spatial conditional normal Poisson model in equation (2.25) fitted in OpenBUGS.



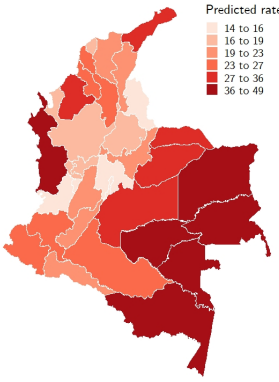
(c) Spatial conditional normal Poisson model in equation (2.26) fitted in OpenBUGS.



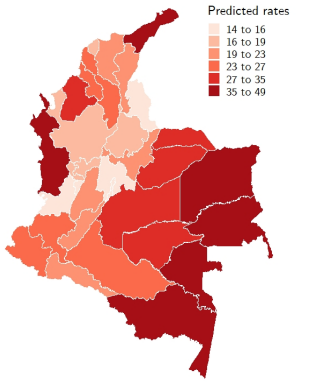
(d) Generalized spatial conditional normal Poisson model in equation (2.27) fitted in OpenBUGS.



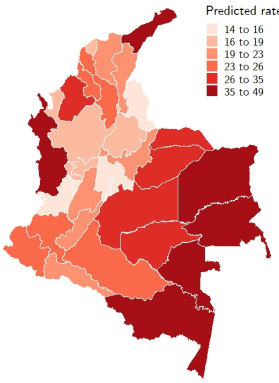
(e) Spatial conditional normal Poisson model in equation (2.25) fitted in INLA.



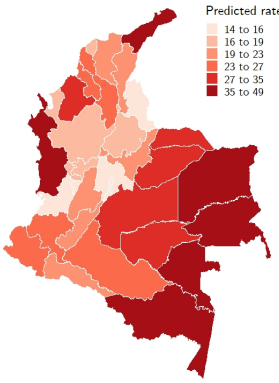
(f) Spatial conditional normal Poisson model in equation (2.26) fitted in INLA.



(g) BYM model including the variables the IBN and Rec fitted in INLA.

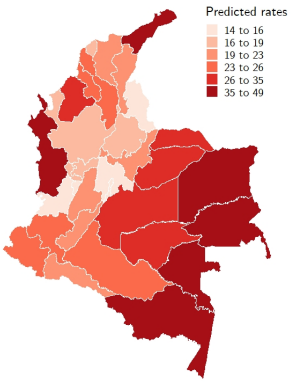


(h) BYM model including the variables the Viol and IBN fitted in INLA.



(i) BYM2 model including the variables the IBN and Rec fitted in INLA.

Figure 2.8: Maps for the observed and estimated mortality rates obtained from some of the models fitted to the Colombian infant mortality rates data set.



(j) BYM2 model including the variables the Viol and IBN fitted in INLA.

Figure 2.8: Maps for the observed and estimated mortality rates obtained from some of the models fitted to the Colombian infant mortality rates data set (Continued).

2.7 Application to mother's postnatal period screening test in Colombia

As in the the case of the study of infant mortality data from the previous section, the data that will be analysed here has also been extracted from the National Statistics Department of Colombia. It consists of observations of multiple variables for each one of the 32 departments or regions in that country, which include the number of women who gave birth to their last child between the years 1999 and 2005 and went through a postnatal period screening test (i.e., variable Nscreen), the number of women who has their last child after 1999 (i.e., variable NMothers), the percentage of women who had to pay for their postnatal check-up (i.e., variable Pay), the percentage of women over 18 years old who declared to have suffered physical violence from their current partners (i.e., variable Viol), the percentage of young people (between 18 and 24 years) who had access to a higher educational level (i.e., variable HE) and the percentage of the population that had basic services not being satisfactorily attended to for the year 2005 (i.e., variable IBN). These variables can be examined in more detail by consulting some of their descriptive statistics, shown in Table 2.13.

Table 2.13: Descriptive statistics for the variables available in the study of the mother's postnatal period screening test in Colombia.

	Nscreen	**NMothers**	**Viol**	**IBN**	**HE**	**Pay**
Median	38.50	56.00	35.87	35.55	13.85	5.12
Mean	118.06	191.09	34.73	37.99	15.69	6.93
SD	192.16	296.61	5.64	17.15	10.52	6.32
Minimum	12.00	22.00	22.58	9.20	1.30	0.78
Maximum	672.00	993.00	44.69	79.20	52.20	27.28

This data set is a variant of the one analysed by Quintero-Sarmiento, Cepeda-Cuervo

and Núñez-Antón (2012) and Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), where the authors fitted their proposed generalized spatial conditional models to the number of women who gave birth to their last child between the years 1999 and 2005 and went through a postnatal period screening test. In their analysis they found evidence of the existence of overdispersion and positive spatial autocorrelation, issues that were properly captured with the fitting of the proposed models. In addition, they identified significant relations between the response variable and the variables IBN and Pay, which were positive in the former case and negative in the latter.

In this section, we study the number of women who gave birth to their last child between the years 1999 and 2005 and went through a postnatal period screening test, represented by the variable Nscreen. We will assume a binomial distribution for this response variable and we will fit the spatial conditional overdispersion models, previously discussed in Section 2.3. The explanatory variables that we will include in the study constitute relevant socio-economic indicators that can have a considerable impact on this variable (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018).

In contrast to the software OpenBUGS, which was used for modelling infant mortality rates in Section 2.6, here the models will be implemented by using the software JAGS, which is also based on the MCMC approach. In the same manner as we did for the Poisson case and as it was already explained in Section 2.6, for our Bayesian framework, we will specify noninformative prior distributions for the parameters, assuming independent normal distributions, $N(0, 1 \times 10^5)$, for all the regression parameters; that is, $\beta_j \sim N(0, 1 \times 10^5)$, $j = 1, \ldots, k$, as well as for the spatial association parameter ρ. As for the inverse of the dispersion parameters τ, $\psi = 1/\tau$, gamma $G(1 \times 10^{-4}, 1 \times 10^{-4})$ distributions were assumed in accordance with the sensitivity analysis performed in Section 2.7.2. In any case, convergence of all the MCMC chains was achieved for all of the parameters included in the proposed models after 10000 iterations, a burn in period of 5000 samples and considering a thinning parameter of 10.

The spatial distribution of the variable corresponding to the number of mothers who went through a postnatal period screening test is shown in Figure 2.9. This map suggests the presence of spatial autocorrelation, since large values of the variable appear to be surrounded by similar values, which is also the case for small values. It is known that regions that are closer in space may also share similar socio-economic characteristics and, for the case for the proportion of mothers who underwent a postnatal period screening test in Colombia, it is reasonable to believe that regions closer in space could exhibit similar values of this variable. In fact, this hypothesis has been considered and investigated before by other authors that have analysed a similar version of this data set (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018). Consequently, this is an issue which will be needed to be taken into account when modelling this variable.

For this purpose, we will consider the same spatial neighbourhood structure that was defined in Section 2.6 for the study of infant mortality rates in Colombia. That is, a spatial weights matrix $\mathbf{W}$ following the first order contiguity criteria and standardized by rows. This matrix will be employed to obtain the spatial term which will be included in the regression structures for the spatial conditional models we will specify. We can define the spatial term corresponding to the i-th region in the following way:

$$A_i = \frac{\hat{\pi}_{\sim i}}{1 - \hat{\pi}_{\sim i}}, \text{ with } \hat{\pi}_{\sim i} = \frac{\mathbf{W}_i \mathbf{Nscreen}}{\mathbf{W}_i \mathbf{Nmothers}}, \tag{2.31}$$

where $\mathbf{W}_i$ the i-th row of spatial weights matrix $\mathbf{W}$ and **Nscreen** and **Nmothers** are the vectors of the observations of the response variable Nscreen and the variable Nmothers,

which represents the number of trials on each area, respectively. Note that $\mathbf{W}_i\mathbf{Nscreen}$ and $\mathbf{W}_i\mathbf{Nmothers}$ are the spatial lags of the variables Nscreen and Nmothers, respectively, for the i-th area.

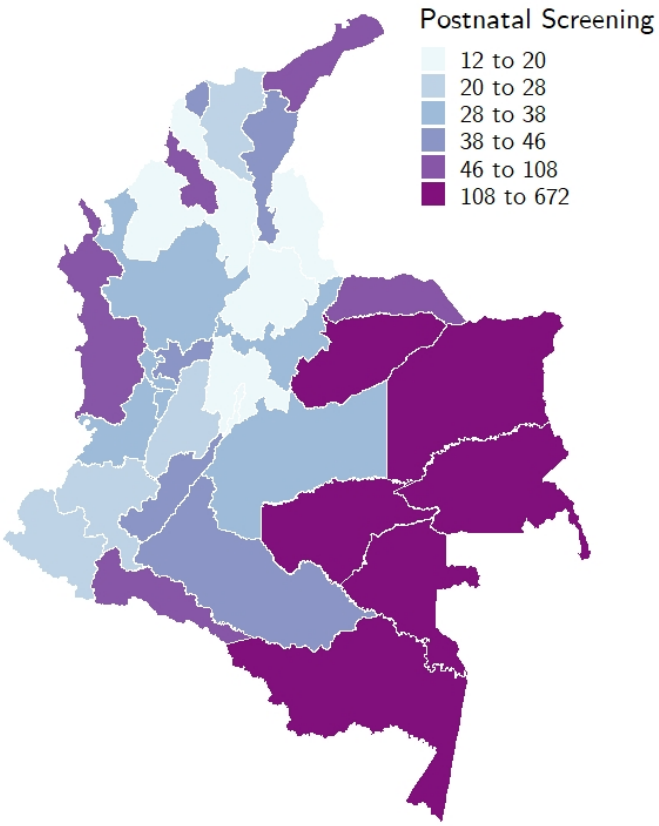


Figure 2.9: Spatial distribution of the number of women who gave birth to their last child between the years 1999 and 2005 and went through a postnatal period screening test in Colombia, by department.

2.7.1 Fitting of the spatial conditional overdispersion models

Our main objective in this study is to model the number of mothers who went through a postnatal period screening test (i.e., variable Nscreen) on each department of Colombia, taking into account overdispersion and the spatial association that might exist. We start by fitting a binomial model and considering that the variable Nscreen follows a binomial distribution, where the number of trials on each region is given by the number of women who gave birth in each of the regions (i.e., variable NMothers). That is, we assume that $\text{Nscreen}_i \sim \text{Bin}(\text{NMothers}_i, \pi_i)$, where π_i follows the regression model:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{HE}_i + \beta_4 \text{Pay}_i \tag{2.32}$$

The results for the fitting of this model to the mother's postnatal period screening test data in Colombia are included in Table 2.14, where the means of the parameter estimates, its standard deviations and its 95% credible intervals are included as well. The information criteria values obtained for this model were DIC = 269.3 and WAIC =

251.1. Based on the estimations for the regression coefficients for this model, we can see that the two variables Viol and Pay are significant, since the value zero is not contained in their 95% credible intervals, whereas the variables IBN and HE are not. However, in this model, the possible existence of overdispersion is not being taken into account and, hence, any inference process performed on the estimated parameters might not be reliable.

Therefore, with the aim of addressing the possible overdispersion, on the one hand, we have fitted the beta binomial model where we assume that π_i also follows the model in equation (2.32) and on the other hand, we have fitted the normal binomial model where π_i follows the regression structure:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{HE}_i + \beta_4 \text{Pay}_i + \nu_i,\ \nu_i \sim N(0, \tau),\ \tau > 0 \quad (2.33)$$

The information criteria values were DIC = 228.2 and WAIC = 204.9 for the normal binomial model and DIC = 233.5 and WAIC = 229.7 for the beta binomial model, a fact which suggests that the former model provides a better fit than the latter, and that the initial binomial model. Furthermore, it should be noted that, for both of these models, the only variable that seems to be statistically significant is Pay.

In order to be able to capture the possible existing spatial correlation in the data, we have also considered the spatial conditional binomial and beta binomial models, where the probability of success follows the regression model:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{HE}_i + \beta_4 \text{Pay}_i + \rho A_i \quad (2.34)$$

The spatial conditional binomial produced information criteria values of DIC = 264.3 and WAIC = 250.6 and, for the spatial conditional beta binomial these values were DIC = 231.1 and WAIC = 230.7. Furthermore, we have fitted the spatial conditional normal binomial model, where π_i follows the regression model:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{HE}_i + \beta_4 \text{Pay}_i + \rho A_i + \nu_i,\ \nu_i \sim N(0, \tau),\ \tau > 0 \quad (2.35)$$

For this model, the information criteria values were DIC = 224.1 and WAIC = 203.4, which are the smallest values obtained from all the models fitted so far. Therefore, we could consider that the best fitting model for this data set is the spatial conditional normal binomial model and, consequently, we have performed a variable selection process on this model by taking into account different combinations of the variables we have used as covariates. The resulting estimates and information criteria values for some of these model fittings are included in Table 2.15.

Table 2.14: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the different models fitted to the mother's postnatal period screening test in Colombia data set.

		Binomial	**Spatial conditional Binomial**	**Normal binomial**	**Spatial conditional normal binomial**	**Beta binomial**	**Spatial conditional beta binomial**
Intercept	Mean	2.0585	1.7271	2.0079	1.7248	2.0796	1.7643
	SD	(0.3021)	(0.3183)	(0.6133)	(0.6542)	(0.7664)	(0.6271)
	CI	(1.4525,2.6266)	(1.1212,2.3306)	(0.8613,3.2419)	(0.4011,2.8948)	(0.5076,3.5343)	(0.5409,3.0216)
Viol	Mean	-0.0325	-0.0336	-0.0301	-0.0321	-0.0324	-0.0333
	SD	(0.0078)	(0.0074)	(0.0173)	(0.0169)	(0.0204)	(0.0165)
	CI	(-0.0479,-0.0170)	(-0.0481,-0.0196)	(-0.0629,0.0036)	(-0.0660,5.533e-05)	(-0.0709,0.0089)	(-0.0667,-0.0026)
IBN	Mean	-0.0017	-0.0042	-0.0058	-0.0085	-0.0058	-0.0083
	SD	(0.0027)	(0.0028)	(0.0065)	(0.0067)	(0.0063)	(0.0067)
	CI	(-0.0075,0.0032)	(-0.0092,0.0017)	(-0.0195,0.0065)	(-0.0221,0.0040)	(-0.0186,0.0065)	(-0.0212,0.0043)
HE	Mean	-9.094e-04	-0.0035	0.0034	1.155e-04	0.0032	7.649e-04
	SD	(0.0058)	(0.0060)	(0.0101)	(0.0102)	(0.0103)	(0.0099)
	CI	(-0.0120,0.0105)	(-0.0150,0.0079)	(-0.0147,0.0238)	(-0.0205,0.0211)	(-0.0171,0.0236)	(-0.0178,0.0210)
Pay	Mean	-0.0604	-0.0496	-0.0547	-0.0439	-0.0546	-0.0453
	SD	(0.0087)	(0.0091)	(0.0182)	(0.0181)	(0.0195)	(0.0193)
	CI	(-0.0774,-0.0438)	(-0.0675,-0.0326)	(-0.0911,-0.0196)	(-0.0790,-0.0093)	(-0.0904,-0.0147)	(-0.0835,-0.0085)
ρ	Mean	-	0.2797	-	0.2818	-	0.2758
	SD	-	(0.1132)	-	(0.1972)	-	(0.1932)
	CI	-	(0.0546,0.5107)	-	(-0.1010,0.6857)	-	(-0.1224,0.6478)
τ	Mean	-	-	0.0865	0.0844	0.0209	0.0196
	SD	-	-	(0.0471)	(0.0415)	(0.0111)	(0.0101)
	CI	-	-	(0.0255,0.2046)	(0.0250,0.1875)	(0.0056,0.0479)	(0.0050,0.0425)
	DIC	269.3	264.3	228.2	224.1	233.5	231.1
	WAIC	251.1	250.6	204.9	203.4	229.7	230.7

Table 2.15: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for some of the reduced versions of the spatial conditional normal binomial model fitted to the mother's postnatal period screening test in Colombia data set.

		Intercept	**Viol**	**IBN**	**HE**	**Pay**	ρ	τ
DIC = 224.1	Mean	1.7248	-0.0321	-0.0085	1.155e-04	-0.0439	0.2818	0.0844
WAIC = 203.4	SD	(0.6542)	(0.0169)	(0.0067)	(0.0102)	(0.0181)	(0.1972)	(0.0415)
	CI	(0.4011,2.8948)	(-0.0660,5.533e-05)	(-0.0221,0.0040)	(-0.0205,0.0211)	(-0.0790,-0.0093)	(-0.1010,0.6857)	(0.0250,0.1875)
DIC = 222.8	Mean	0.7053	-	-0.0133	-0.0030	-0.0230	0.2742	0.0992
WAIC = 203.4	SD	(0.4150)	-	(0.0064)	(0.0102)	(0.0160)	(0.2161)	(0.0491)
	CI	(-0.1183,1.4949)	-	(-0.0264,-7.395e-04)	(-0.0214,0.0174)	(-0.0541,0.0085)	(-0.1274,0.6890)	(0.0334,0.2191)
DIC = 227.1	Mean	1.7818	-0.0407	-	0.0100	-0.0608	0.2002	0.0812
WAIC = 206.1	SD	(0.6678)	(0.0161)	-	(0.0073)	(0.0142)	(0.1840)	(0.0418)
	CI	(0.4833,3.0527)	(-0.0730,-0.0092)	-	(-0.0035,0.0249)	(-0.0871,-0.0340)	(-0.1514,0.5658)	(0.0239,0.1789)
DIC = 224.5	Mean	1.7497	-0.0325	-0.0085	-	-0.0444	0.2789	0.0796
WAIC = 203.6	SD	(0.6515)	(0.0161)	(0.0046)	-	(0.0166)	(0.1988)	(0.0399)
	CI	(0.4700,3.0422)	(-0.0632,-1.691e-04)	(-0.0182,-2.344e-04)	-	(-0.0762,-0.0121)	(-0.0990,0.6595)	(0.0247,0.1743)
DIC = 218.9	Mean	0.8216	-0.0073	-0.0188	-0.0103	-	0.4715	0.1191
WAIC = 202	SD	(0.5957)	(0.0144)	(0.0058)	(0.0098)	-	(0.2089)	(0.0551)
	CI	(-0.3290,2.0118)	(-0.0362,0.0208)	(-0.0312,-0.0084)	(-0.0298,0.0084)	-	(0.0589,0.8763)	(0.0473,0.2483)
DIC = 215.7	Mean	0.4952	-0.0050	-0.0158	-	-	0.4629	0.1250
WAIC = 200.5	SD	(0.5031)	(0.0149)	(0.0048)	-	-	(0.2083)	(0.0488)
	CI	(-0.4708,1.5224)	(-0.0345,0.0238)	(-0.0264,-0.0069)	-	-	(0.0680,0.8681)	(0.0530,0.2469)
DIC = 234.0	Mean	1.9690	-0.0425	-	-	-0.0586	0.1964	0.0788
WAIC = 208.3	SD	(0.6362)	(0.0157)	-	-	(0.0143)	(0.1799)	(0.0438)
	CI	(0.7487,3.3148)	(-0.0751,-0.0134)	-	-	(-0.0875,-0.0310)	(-0.1498,0.5387)	(0.0188,0.1761)
DIC = 220.3	Mean	0.6352	-	-0.0124	-	-0.0240	0.2708	0.0934
WAIC = 203.1	SD	(0.3468)	-	(0.0044)	-	(0.0133)	(0.1980)	(0.0435)
	CI	(-0.0425,1.3085)	-	(-0.0217,-0.0038)	-	(-0.0488,0.0024)	(-0.1163,0.6472)	(0.0319,0.1963)
DIC = 225.6	Mean	0.0882	-0.0103	-	-	-	0.4285	0.1839
WAIC = 204.6	SD	(0.5424)	(0.0166)	-	-	-	(0.2338)	(0.0750)
	CI	(-0.9864,1.1713)	(-0.0426,0.0228)	-	-	-	(-0.0088,0.8817)	(0.0759,0.3797)
DIC = 216.3	Mean	0.3404	-	-0.0154	-	-	0.4389	0.1169
WAIC = 200.9	SD	(0.3207)	-	(0.0046)	-	-	(0.1879)	(0.0493)
	CI	(-0.2835,0.9592)	-	(-0.0247,-0.0063)	-	-	(0.0882,0.8050)	(0.0494,0.2457)
DIC = 230.9	Mean	0.4649	-	-	-	-0.0365	0.1064	0.1194
WAIC = 207.6	SD	(0.3616)	-	-	-	(0.0133)	(0.2026)	(0.0557)
	CI	(-0.2507,1.1960)	-	-	-	(-0.0622,-0.0109)	(-0.2870,0.5090)	(0.0412,0.2495)

The model including the two explanatory variables Viol and IBN and the model including only the variable IBN could be considered as the best fitting models, taking into account that their information criteria values are the smallest ones. First, we will examine with more detail the spatial conditional normal binomial model with regression structure for π_i:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \rho A_i + \nu_i,\ \nu_i \sim N(0, \tau),\ \tau > 0 \tag{2.36}$$

This model produced information criteria values of DIC = 215.7 and WAIC = 200.5. The estimated coefficient for the variable Viol was $\hat{\beta}_1 = -0.0050$ with a standard deviation of 0.0149. This variable is not statistically significant, as the value zero is contained in its 95% credible interval. For the variable IBN the estimated coefficient was $\hat{\beta}_2 = -0.0158$, with a standard deviation of 0.0048. The 95% credible interval for this variable does not contain the value zero, hence IBN is statistically significant in this model. In addition, the estimated coefficient of the spatial parameter was $\hat{\rho} = 0.4629$, with a standard deviation of 0.2083 and it is statistically significant, as zero is not contained in its 95% credible interval. This fact suggests the existence of positive spatial correlation in the data, which is being captured by the spatial term employed.

We will also further investigate the spatial conditional normal binomial model with the following regression structure for the probability of success:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \text{IBN}_i + \rho A_i + \nu_i,\ \nu_i \sim N(0, \tau),\ \tau > 0 \tag{2.37}$$

For the model where the probability of success follows the regression structure in equation (2.37), the estimated coefficient for the variable IBN was $\hat{\beta}_1 = -0.0154$, with a standard deviation of 0.0046. The 95% credible interval for this variable does not contain the value zero, hence IBN is statistically significant. The estimated spatial parameter was $\hat{\rho} = 0.4389$, with a standard deviation of 0.1879 and statistically significant according to its 95% credible interval, again suggesting that the spatial dependence in the variable is being explained by this term.

Finally, we have also considered the following model for a more comprehensive examination:

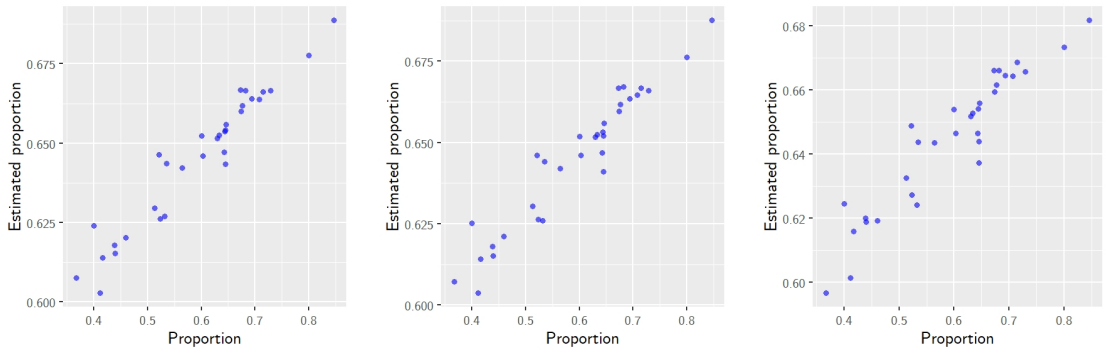
$$\text{logit}(\pi_\text{i}) = \beta_0 + \beta_1 \text{IBN}_\text{i} + \beta_2 \text{Pay}_\text{i} + \rho \text{A}_\text{i} + \nu_\text{i},\ \nu_\text{i} \sim \text{N}(0, \tau),\ \tau > 0 \tag{2.38}$$

Here, the estimated coefficient for the variable IBN was $\hat{\beta}_1 = -0.0124$ with a standard deviation of 0.0044, and, for the variable Pay was $\hat{\beta}_2 = -0.0240$, with a standard deviation of 0.0133. According to their 95% credible interval, in this case only the variable IBN is statistically significant.

These three models suggest that the probability of a woman that gives birth goes through a postnatal period screening test tends to be smaller in regions where the index of basic needs not satisfactorily attended have larger values, since there is evidence of the existence of a negative statistically significant relation among these two variables. Moreover, models where the probability of success π_i followed the regression structures in equation (2.36) and equation (2.37) indicate the existence of positive spatial correlation

in the data, which is being captured by the spatial term employed. In these models closely examined, we have not found evidence of any other statistically significant relationship between the response variable Nscreen and the other covariates considered.

In Figure 2.10 we can see the scatterplots for the observed versus the predicted proportions obtained from some of the fitted models to the Colombia mother's postnatal period screening test data set. The proportion of successes can be considered as a rough approximation to the probability of successes π being modelled. Therefore, in these plots, we can examine the predictive accuracy of the fitted models. For the spatial conditional normal binomial models where the probability of success follows equations (2.36), (2.37) and (2.38), we can see that the predictive accuracy behaves in a similar way for each of the three cases. The observed and predicted proportions are distributed fairly close to the line which would indicate that they have similar values.



(a) Spatial conditional normal binomial model in equation (2.36) fitted in JAGS.

(b) Spatial conditional normal binomial model in equation (2.37) fitted in JAGS.

(c) Spatial conditional normal binomial model in equation (2.38) fitted in JAGS.

Figure 2.10: Scatterplots for the observed versus the predicted proportions obtained from some of the fitted models to the Colombia mother's postnatal period screening test data set, fitted in JAGS

We can assess the size of the effect of the explanatory variables considered in some of the fitted models by examining their marginal effects at the means, with their results included in Table 2.16. For the variable IBN, in the model in equation (2.36), this effect was -0.0170, indicating that an increment of 1 percentage point in the variable IBN would decrease the proportion of mothers that goes through a postnatal period screening test by a 1.7%. Taking into account that the 95% credible interval of this effect is $(-0.0286, -0.0074)$, it can be considered as significant. Moreover, significant effects according to their credible intervals, were also obtained for this variable in the models in equation (2.37) and equation (2.38).

Table 2.16: Marginal effects at the means for the spatial conditional normal binomial models in equations (2.36), (2.37) and (2.38) fitted to the mother's postnatal period screening test in Colombia data set.

		Mean	SD	95% CI
Model in equation (2.36)	**Viol**	-0.0053	(0.0160)	(-0.0377,0.0258)
	IBN	-0.0170	(0.0053)	(-0.0286,-0.0074)
Model in equation (2.37)	**IBN**	-0.0166	(0.0050)	(-0.0266,-0.0068)
Model in equation (2.38)	**IBN**	-0.0133	(0.0048)	(-0.0237,-0.0040)
	Pay	-0.0257	(0.0142)	(-0.0519,0.0026)

In order to assess the convergence of the MCMC chains for the models we have implemented, we have computed some convergence performance diagnostics such as the the effective sample size (i.e., N_{eff}) and the potential scale reducing factor (i.e., $\hat{R}$), which have already been described in Section 2.6.1. Here, we have also simulated 3 Markov chains, with 10000 iterations and a burn in period of 5000 each, where a total of 15000 samples remained. The results of the diagnostics applied, shown in Table 2.17, suggest that there were enough independent samples for each parameter, since all values of N_{eff} are larger than 400. Note that, as we already mentioned in Section 2.6.1, it is generally considered for the N_{eff} that a value of 100 is enough to perform reasonable posterior inference. Regarding the values obtained for $\hat{R}$, we can see that they are all very close to one, which indicates the convergence of the chains to the target distributions.

Table 2.17: Convergence diagnostics for the spatial conditional normal binomial models in equations (2.36), (2.37) and (2.37) fitted to the mother's postnatal period screening test in Colombia data set.

		Intercept	**Viol**	**IBN**	**Pay**	ρ	τ
Model in equation (2.36)	N_{eff}	3000	3000	3000	-	3000	3000
	$\hat{R}$	1.0005	1.0007	1.0009	-	1.0012	1.0006
Model in equation (2.37)	N_{eff}	2300	-	1600	-	420	420
	$\hat{R}$	1.0014	-	1.0022	-	1.0053	1.0053
Model in equation (2.38)	N_{eff}	930	-	1500	3000	840	770
	$\hat{R}$	1.0027	-	1.0021	1.0006	1.0029	1.0034

Finally, it is worth mentioning that the results obtained when fitting the models with probability of success given by equations (2.36), (2.37) and (2.38), are consistent with the results found by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), when analysing a similar data set. This corroborates the fact that the spatial conditional overdispersion models are suitable and practical when modelling real data sets by being able to explain spatial dependence and overdispersion that can be present in this data.

2.7.2 Sensitivity analysis for the precision of the prior distributions

As it was explained in Section 2.6.2, when the prior distribution for the precision parameter of the random effects $\psi = 1/\tau$, the inverse of the variance, that is $G(\alpha, \alpha)$, is specified, the choice of the values α could considerably affect the posterior inference process. Therefore, in the same way as we proceeded in the infant mortality study in Section 2.6.2, we have performed a sensitivity analysis on the model in equation (2.36), by considering different possible values of α, from $\alpha = 0.1$ to $\alpha =$ 1e-08. This process is performed in order to select a value for α which ensures that this effect is controlled for.

The results included in Table 2.18 indicate that for values of $\alpha =$ 1e-04, as well as for smaller values, the differences in the estimates are insignificant, where only changes in the third decimal place are observed for a few cases. In addition, no changes are observed in the posterior marginal densities for the estimated precision parameter ψ, shown in the right panel of Figure 2.11, when the value of α is set from $\alpha =$ 1e-04 up to $\alpha =$ 1e-08. Taking these facts into account, considering the value $\alpha =$ 1e-04 for the prior distribution $G(\alpha, \alpha)$ would be justified since it would ensure that there is no undesirable influence on the inferential process.

Table 2.18: Posterior means for parameter estimates together with standard deviations, DIC and WAIC values, for the spatial conditional normal binomial model in the analysis of the mother's postnatal period screening test in Colombia data set with different prior distributions for the precision parameter of the random effects.

	Intercept	**Viol**	**IBN**	ρ	τ	DIC	WAIC
$\alpha = 0.1$	0.5123	-0.0061	-0.0156	0.4712	0.1398	214.0	199.3
	(0.5554)	(0.0161)	(0.0049)	(0.2172)	(0.0564)		
$\alpha = 0.01$	0.4955	-0.0050	-0.0160	0.4677	0.1256	215.0	200.4
	(0.5122)	(0.0149)	(0.0048)	(0.1988)	(0.0505)		
$\alpha = 0.001$	0.4760	-0.0047	-0.0156	0.4622	0.1260	214.8	200.5
	(0.5119)	(0.0151)	(0.0046)	(0.2053)	(0.0529)		
$\alpha =$ 1e-04	0.4952	-0.0050	-0.0158	0.4629	0.1250	215.7	200.5
	(0.5031)	(0.0149)	(0.0048)	(0.2083)	(0.0488)		
$\alpha =$ 1e-05	0.4953	-0.0050	-0.0158	0.4629	0.1250	215.7	200.5
	(0.5031)	(0.0149)	(0.0048)	(0.2083)	(0.0488)		
$\alpha =$ 1e-06	0.4953	-0.0050	-0.0158	0.4629	0.1250	215.7	200.5
	(0.5031)	(0.0149)	(0.0048)	(0.2083)	(0.0488)		
$\alpha =$ 1e-07	0.4953	-0.0050	-0.0158	0.4629	0.1250	215.7	200.5
	(0.5031)	(0.0149)	(0.0048)	(0.2083)	(0.0488)		
$\alpha =$ 1e-08	0.4953	-0.0050	-0.0158	0.4629	0.1250	215.7	200.5
	(0.5031)	(0.0149)	(0.0048)	(0.2083)	(0.0488)		

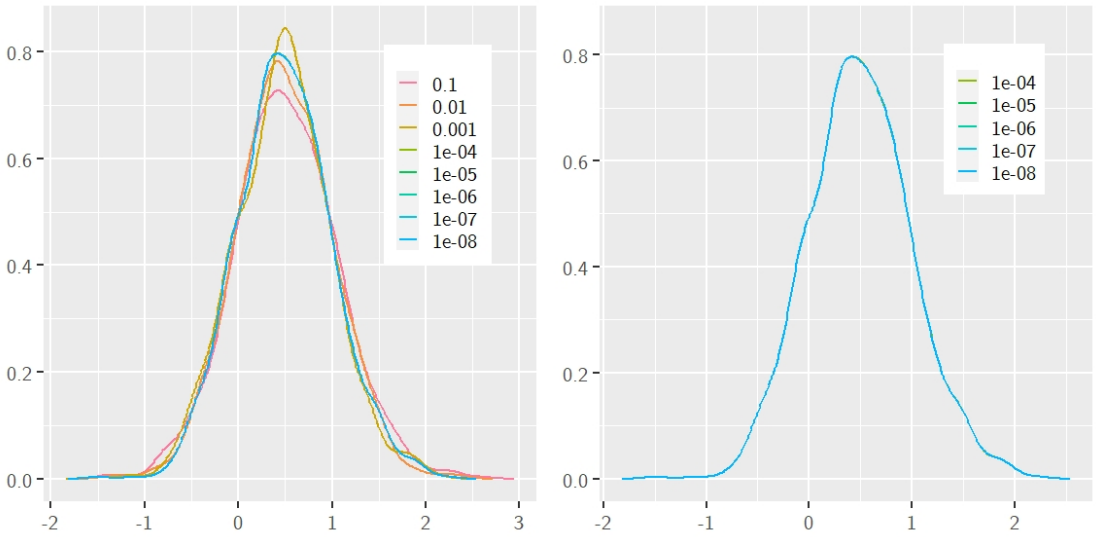


Figure 2.11: Posterior marginal distributions for the precision parameter $\psi = 1/\tau$, the inverse of the variance parameter τ for the random effects, for different values of α, where $\psi \sim G(\alpha, \alpha)$.

2.7.3 Fitting of the generalized spatial conditional normal binomial model

In the previous sections, for the study of mother's postnatal period screening test, it has been assumed that the dispersion parameters in the fitted models are constant. Here, in order to investigate the possibility that these parameters might depend on some covariates or spatial terms, we have considered to fit the generalized spatial conditional overdispersion models proposed by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018). Therefore, we have fitted different combinations of the generalized spatial conditional normal binomial model. One of these is given by the model with probability of success π_i given by the regression structure:

$$\begin{array}{rcl} \text{logit}(\pi_i) & = & \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{Pay}_i + \rho A_i + \nu_i, \quad \nu_i \sim N(0, \tau_i),\ \tau_i > 0, \\ \log(\tau_i) & = & \gamma_0 + \gamma_1 \text{IBN}_i \end{array} \tag{2.39}$$

Another model further examined is the one where π_i follows the regression model:

$$\begin{array}{rcl} \text{logit}(\pi_i) & = & \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{Pay}_i + \nu_i, \quad \nu_i \sim N(0, \tau_i),\ \tau_i > 0, \\ \log(\tau_i) & = & \gamma_0 + \gamma_1 \text{IBN}_i + \rho A_i \end{array} \tag{2.40}$$

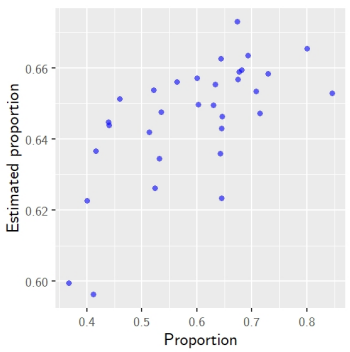
Table 2.19 includes the corresponding parameter estimates for the fitting of the generalized spatial conditional normal binomial models with regression structures given in equations (2.39) and (2.40), together with its standard deviations and 95% credible intervals in parenthesis.

Table 2.19: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the generalized spatial conditional normal binomial models fitted to the mother's postnatal period screening test in Colombia data set.

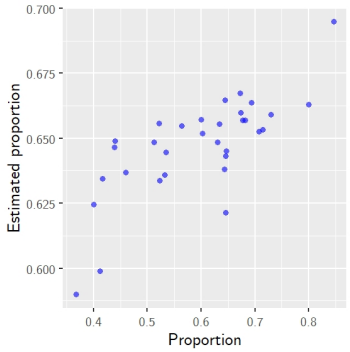
		Model in equation (2.39)	Model in equation (2.40)
π **Intercept** - β_0	Mean	1.7570	2.0161
	SD	(0.3413)	(0.2929)
	CI	(1.0827,2.3994)	(1.4596,2.5957)
Viol	Mean	-0.0372	-0.0337
	SD	(0.0073)	(0.0071)
	CI	(-0.0518,-0.0228)	(-0.0477,-0.0201)
Pay	Mean	-0.0557	-0.0624
	SD	(0.0083)	(0.0074)
	CI	(-0.0716,-0.0400)	(-0.0759,-0.0480)
ρ	Mean	0.2206	-39.0248
	SD	(0.1082)	(14.8624)
	CI	(0.0125,0.4191)	(-65.7017,-14.0839)
τ **Intercept** - γ_0	Mean	14.5718	-30.5057
	SD	(20.9224)	(15.6823)
	CI	(-33.9406,43.3918)	(-62.8097,-4.2467)
IBN	Mean	-3.5596	-3.0872
	SD	(0.6047)	(1.5349)
	CI	(-4.7779,-2.5861)	(-6.5984,-0.7618)
		DIC = 238.7 WAIC = 245.3	DIC = 229.7 WAIC = 230.1

The information criteria values obtained with the fitting of the generalized spatial conditional normal binomial models were DIC = 238.7 and WAIC = 245.3 for the model where π_i follows equation (2.39) and DIC = 229.7 and WAIC = 230.1 for the model where π_i follows equation (2.40). These values do not show improvements, compared to the information criteria values obtained with the fitting of the spatial conditional normal binomial models with results shown in Table 2.15.

Furthermore, after comparing the scatterplots in Figures 2.12(a) and 2.12(b), with the ones from Figures 2.10(a), 2.10(b) and 2.10(c), we could say that the accuracy of the predictions obtained with the generalized models is not as good as the accuracy obtained when fitting the spatial conditional normal binomial models. In fact, the predicted proportions obtained appear to be considerably deviated from the observed ones. In addition, the maps of the predicted proportions, shown in Figures 2.15(e) and 2.15(f) show some discrepancies with the observed proportions in Figure 2.15(a) for a substantial number of regions.



(a) Generalized spatial conditional normal binomial model in equation (2.39) fitted in JAGS.



(b) Generalized spatial conditional normal binomial model in equation (2.40) fitted in JAGS.

Figure 2.12: Scatterplots for the observed versus the predicted proportions obtained from the generalized spatial conditional normal binomial models fitted to the Colombia mother's postnatal period screening test data set, in JAGS.

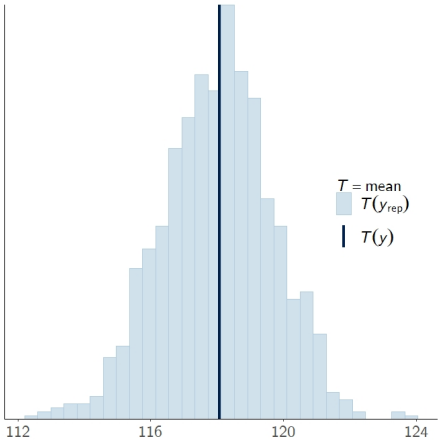
Even though it has been shown that these models do not seem to provide a better fit to the data than the spatial conditional normal binomial models, we believe it is convenient to highlight some fact about them. For instance, the coefficient of the variable IBN was $\hat{\gamma_1} = -3.5596(0.6047)$ for the model in equation (2.39) and $\hat{\gamma_1} = -3.0872(1.5349)$ for the model in equation (2.40). This parameter is statistically significant in the regression models for the dispersion in both cases, according to the 95% credible intervals, which could be indicating that the dispersion is not constant, as was assumed before, but depends on this variable in these cases. Moreover, the fact that the parameter is negative suggests that regions that present larger values of this variable also tend to have smaller dispersion values.

Regarding the spatial term, the value for its coefficient for the model in equation (2.39) was $\hat{\rho} = 0.2206(0.1082)$, resulting statistically significant given that its 95% credible intervals does not contain the value zero. The value obtained for the spatial parameter in the model in equation (2.40) was $\hat{\rho} = -39.0248(14.8624)$, resulting statistically significant according to its 95% credible interval. Hence, the spatial autocorrelation that seems to be present in the data is being properly captured in both cases.

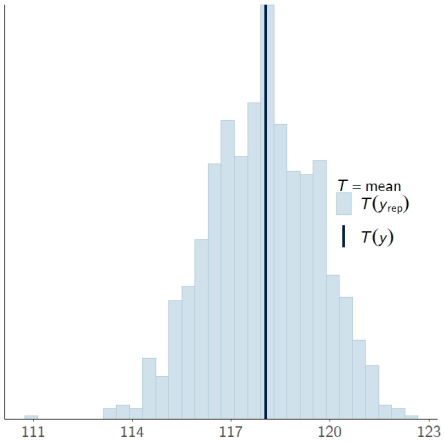
On the one hand, taking all of the above into consideration, we could say that the generalized models might not be the most suitable choice to model this data, when compared to the spatial conditional normal binomial models fitted above. On the other hand, we have found evidence that the dispersion is accommodated properly with these models and could vary according to some explanatory variables or even the spatial term, whereas with the spatial conditionals it is considered as a constant parameter. It is in the hands of the researchers to evaluate which model provides a better and more appealing approach in their specific case study.

Finally, Figure 2.13 includes histograms of the posterior predictive distributions for

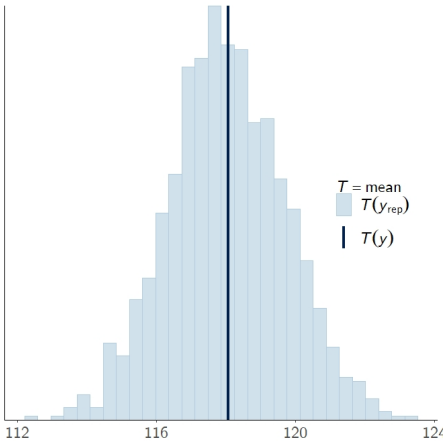
the means over replicated simulations of the proportions, estimated from the fitting of the spatial conditional normal binomial models in equation (2.36) (see Figure 2.13(a)), equation (2.37) (see Figure 2.13(b)) and equation (2.38) (see Figure 2.13(c)), and by the generalized spatial conditional normal binomial models in equation (2.39) (see Figure 2.13(d)). Note that the dark blue vertical line represents the value of the mean for the observed data. As we can see from these figures, they show how the means of the replicated data vary with respect to the actual mean of the observed values, a fact suggesting that the considered models offer an adequate fit to the data set under study.



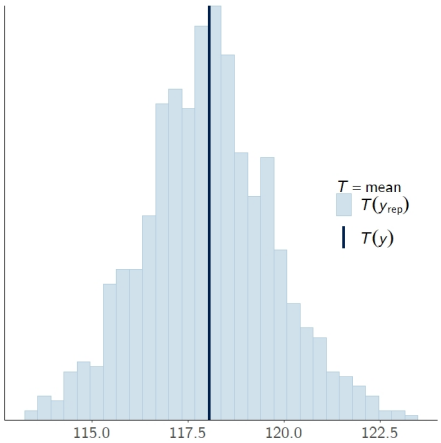
(a) Spatial conditional normal binomial model in equation (2.36).



(b) Spatial conditional normal binomial model in equation (2.37).



(c) Spatial conditional normal binomial model in equation (2.38).



(d) Generalized spatial conditional normal binomial model in equation (2.39).

Figure 2.13: Posterior predictive checks for some of the fitted models to the mother's postnatal period screening test in Colombia data set.

2.7.4 Comparisons to the BYM model for binomial responses

In this section, we will compare the performance between the spatial conditional overdispersion models fitted to the mother's postnatal period screening test with the BYM model for binomial responses. In the same way we proceeded for the infant mortality study in Section 2.6.4, we have selected the best fitting model, which was the spatial conditional normal binomial model as a point of comparison with the BYM model. For this purpose, we have fitted the binomial BYM model with the following structure for the probability of success:

$$\text{logit}(\pi_\text{i}) = \beta_0 + \beta_1 \text{Viol}_\text{i} + \beta_2 \text{IBN}_\text{i} + \beta_3 \text{HE}_\text{i} + \beta_4 \text{Pay}_\text{i} + \nu_\text{i} + \eta_\text{i}, \tag{2.41}$$

where ν_i and η_i, $i = 1, \ldots, n$, are the normally and intrinsic CAR distributed sets of random effects, with variance parameters $\tau > 0$ and $\tau_\eta > 0$, respectively. In the same way as the previous models, we assume noninformative normal priors for the regression parameters; that is, $N(0, 1 \times 10^{-5})$ and, noninformative gamma $G(1 \times 10^{-4}, 1 \times 10^{-4})$ prior distributions for the precision parameters $\psi = 1/\tau$ and $\psi_\eta = 1/\tau_\eta$. In addition, the same first order neighbourhood structure is also assumed.

The implementation of the BYM model in JAGS is not straightforward as it would be in R-INLA or even in OpenBUGS, which has a function for specifically specifying CAR distributed random effects (Lawson, 2008). Nevertheless, it has been shown that the computational times achieved when fitting models with structured random effects are much lower with the R-INLA package (Vranckx, Neyens and Faes, 2019). Hence, we have decided to fit these models by using the INLA approach and, in order to be able to compare the results obtained to the estimations from the fitting of the spatial conditional normal binomial models, we would also need to fit these models in this software package. This is required due to the fact that information criteria values such as the DIC and the WAIC obtained from different software such as OpenBUGS and R-INLA cannot be directly compared (Vranckx, Neyens and Faes, 2019).

The results of the fitting of the binomial BYM model in equation (2.41) and some of its reduced versions are included in Table 2.20. In addition, Table 2.21 includes the results of some of the reduced versions of the spatial conditional normal binomial model from equation (2.35) fitted in INLA, the same combinations of variables that were fitted in JAGS and shown in Section 2.7.1, Table 2.15.

Table 2.20: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for some of the reduced versions of the binomial BYM model fitted in INLA to the mother's postnatal period screening test in Colombia data set.

		Intercept	**Viol**	**IBN**	**HE**	**Pay**	τ	τ_η
DIC = 211.5	Mean	1.9684	-0.0295	-0.0063	0.0044	-0.0523	0.0831	0.0290
WAIC = 210.9	SD	(0.6270)	(0.0175)	(0.0062)	(0.0105)	(0.0178)	(0.0352)	(0.0352)
	CI	(0.7443,3.2248)	(-0.0644,0.0048)	(-0.0188,0.0058)	(-0.0160,0.0253)	(-0.0872,-0.0167)	(0.0342,0.1704)	(0.0018,0.1195)
DIC = 210.7	Mean	1.0882	-	-0.0118	4.391e-04	-0.0344	0.0802	0.0464
WAIC = 208.2	SD	(0.3604)	-	(0.0063)	(0.0115)	(0.0167)	(0.0152)	(0.0379)
	CI	(0.3878,1.8147)	-	(-0.0245,3.775e-04)	(-0.0220,0.0234)	(-0.0685,-0.0016)	(0.0538,0.1131)	(0.0092,0.1489)
DIC = 212.1	Mean	1.9552	-0.0373	-	0.0114	-0.0619	0.0761	0.0259
WAIC = 212.4	SD	(0.6148)	(0.0154)	-	(0.0078)	(0.0148)	(0.0305)	(0.0296)
	CI	(0.7534,3.1864)	(-0.0683,-0.0074)	-	(-0.0035,0.0273)	(-0.0910,-0.0326)	(0.0319,0.1499)	(0.0026,0.1022)
DIC = 211.5	Mean	2.0147	-0.0276	-0.0080	-	-0.0497	0.0760	0.0238
WAIC = 211.7	SD	(0.6212)	(0.0172)	(0.0049)	-	(0.0168)	(0.0377)	(0.0349)
	CI	(0.7901,3.2662)	(-0.0618,0.0070)	(-0.0184,0.0010)	-	(-0.0820,-0.0152)	(0.0246,0.1694)	(8.726e-04,0.1079)
DIC = 210.9	Mean	1.1405	-0.0011	-0.0164	-0.0046	-	0.1076	0.1611
WAIC = 207.8	SD	(0.6662)	(0.0184)	(0.0065)	(0.0123)	-	(0.0682)	(0.1358)
	CI	(-0.1593,2.4783)	(-0.0391,0.0341)	(-0.0294,-0.0035)	(-0.0280,0.0206)	-	(0.0326,0.2886)	(0.0222,0.5196)
DIC = 210.1	Mean	1.0328	-0.0018	-0.0148	-	-	0.0893	0.1869
WAIC = 207.5	SD	(0.6142)	(0.0178)	(0.0051)	-	-	(0.0606)	(0.1536)
	CI	(-0.1433,2.2843)	(-0.0380,0.0325)	(-0.0250,-0.0050)	-	-	(0.0226,0.2503)	(0.0318,0.5935)
DIC = 213	Mean	2.1823	-0.0394	-	-	-0.0615	0.0742	0.0309
WAIC = 214.5	SD	(0.6043)	(0.0155)	-	-	(0.0147)	(0.0430)	(0.0996)
	CI	(1.0043,3.4007)	(-0.0707,-0.0092)	-	-	(-0.0908,-0.0323)	(0.0195,0.1832)	(-7.039e-04,0.1814)
DIC = 209.8	Mean	1.0894	-	-0.0116	-	-0.0344	0.0767	0.0328
WAIC = 208.3	SD	(0.2025)	-	(0.0047)	-	(0.0151)	(0.0153)	(0.0280)
	CI	(0.7005,1.4971)	-	(-0.0211,-0.0027)	-	(-0.0650,-0.0045)	(0.0499,0.1096)	(0.0065,0.1080)
DIC = 213	Mean	0.8938	-0.0147	-	-	-	0.1225	0.2501
WAIC = 211.2	SD	(0.7022)	(0.0201)	-	-	-	(0.0846)	(0.2081)
	CI	(-0.4522,2.3193)	(-0.0554,0.0239)	-	-	-	(0.0318,0.3470)	(0.0392,0.7960)
DIC = 209.8	Mean	0.9747	-	-0.0148	-	-	0.0994	0.1396
WAIC = 207.2	SD	(0.2063)	-	(0.0049)	-	-	(0.0661)	(0.1235)
	CI	(0.5773,1.3905)	-	(-0.0248,-0.0054)	-	-	(0.0284,0.2760)	(0.0161,0.4610)
DIC = 213.4	Mean	0.6749	-	-	-	-0.0426	0.1108	0.0329
WAIC = 213.9	SD	(0.1247)	-	-	-	(0.0150)	(0.0534)	(0.0536)
	CI	(0.4370,0.9335)	-	-	-	(-0.0742,-0.0140)	(0.0397,0.2449)	(4.247e-04,0.1639)

Table 2.21: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for some of the reduced versions of the spatial conditional normal binomial model fitted in INLA to the mother's postnatal period screening test in Colombia data set.

		Intercept	**Viol**	**IBN**	**HE**	**Pay**	ρ	τ
DIC = 210.6	Mean	1.6970	-0.0313	-0.0087	2.860e-04	-0.0436	0.2854	0.0851
WAIC = 209.5	SD	(0.6625)	(0.0170)	(0.0066)	(0.0101)	(0.0186)	(0.1970)	(0.0448)
	CI	(0.3851,3.0113)	(-0.0654,0.0021)	(-0.0221,0.0040)	(-0.0195,0.0205)	(-0.0799,-0.0062)	(-0.0967,0.6830)	(0.0237,0.1962)
DIC = 210.9	Mean	0.7087	-	-0.0137	-0.0032	-0.0216	0.2773	0.0988
WAIC = 208.9	SD	(0.4152)	-	(0.0064)	(0.0104)	(0.0153)	(0.2064)	(0.0476)
	CI	(-0.1181,1.5270)	-	(-0.0267,-0.0014)	(-0.0236,0.0175)	(-0.0514,0.0092)	(-0.1245,0.6919)	(0.0330,0.2170)
DIC = 211.5	Mean	1.7517	-0.0402	-	0.0098	-0.0595	0.2038	0.0842
WAIC = 211.5	SD	(0.6592)	(0.0157)	-	(0.0073)	(0.0142)	(0.1870)	(0.0449)
	CI	(0.4455,3.0600)	(-0.0719,-0.0097)	-	(-0.0040,0.0246)	(-0.0877,-0.0314)	(-0.1584,0.5825)	(0.0222,0.1954)
DIC = 209.9	Mean	1.6998	-0.0312	-0.0087	-	-0.0435	0.2841	0.0792
WAIC = 209	SD	(0.6416)	(0.0163)	(0.0046)	-	(0.0160)	(0.1897)	(0.0418)
	CI	(0.4318,2.9746)	(-0.0638,8.850e-04)	(-0.0183,2.368e-05)	-	(-0.0744,-0.0111)	(-0.0826,0.6680)	(0.0215,0.1829)
DIC = 210.2	Mean	0.8318	-0.0073	-0.0189	-0.0104	-	0.4718	0.1172
WAIC = 207.2	SD	(0.6059)	(0.0148)	(0.0056)	(0.0098)	-	(0.2012)	(0.0519)
	CI	(-0.3618,2.0352)	(-0.0368,0.0217)	(-0.0303,-0.0080)	(-0.0294,0.0093)	-	(0.0756,0.8715)	(0.0447,0.2454)
DIC = 209.1	Mean	0.4707	-0.0052	-0.0155	-	-	0.4736	0.1228
WAIC = 205.7	SD	(0.5098)	(0.0149)	(0.0046)	-	-	(0.2045)	(0.0512)
	CI	(-0.5181,1.4980)	(-0.0351,0.0239)	(-0.0248,-0.0066)	-	-	(0.0716,0.8800)	(0.0507,0.2490)
DIC = 213	Mean	1.9791	-0.0427	-	-	-0.0583	0.1928	0.0799
WAIC = 214.8	SD	(0.6252)	(0.0152)	-	-	(0.0139)	(0.1829)	(0.0453)
	CI	(0.7568,3.2403)	(-0.0738,-0.0133)	-	-	(-0.0860,-0.0309)	(-0.1627,0.5629)	(0.0175,0.1917)
DIC = 210.1	Mean	0.6474	-	-0.0123	-	-0.0240	0.2635	0.0938
WAIC = 208.4	SD	(0.3509)	-	(0.0045)	-	(0.0134)	(0.1993)	(0.0449)
	CI	(-0.0502,1.3373)	-	(-0.0216,-0.0037)	-	(-0.0497,0.0032)	(-0.1235,0.6645)	(0.0314,0.2050)
DIC = 213.2	Mean	0.0919	-0.0101	-	-	-	0.4217	0.1820
WAIC = 210.8	SD	(0.5712)	(0.0171)	-	-	-	(0.2346)	(0.0725)
	CI	(-1.0362,1.2242)	(-0.0441,0.0235)	-	-	-	(-0.0385,0.8899)	(0.0781,0.3594)
DIC = 208.7	Mean	0.3353	-	-0.0156	-	-	0.4460	0.1157
WAIC = 205.6	SD	(0.3253)	-	(0.0045)	-	-	(0.1843)	(0.0481)
	CI	(-0.3012,0.9845)	-	(-0.0247,-0.0069)	-	-	(0.0815,0.8102)	(0.0477,0.2341)
DIC = 213.5	Mean	0.4621	-	-	-	-0.0368	0.1098	0.1204
WAIC = 213.8	SD	(0.3728)	-	-	-	(0.0134)	(0.2057)	(0.0567)
	CI	(-0.2846,1.1924)	-	-	-	(-0.0632,-0.0101)	(-0.2909,0.5244)	(0.0403,0.2596)

We can compare the results obtained by fitting the binomial BYM models from Table 2.20 with the ones obtained after fitting the spatial conditional normal binomial models, included in Table 2.21. First of all, we can observe that the information criteria values are very similar, not favouring any of the models in Table 2.20 over the ones in Table 2.21.

As for the estimated coefficients, although they maintain their signs, there are some changes in their values. For example, for the binomial BYM model only including the variables Viol and IBN, the coefficient of Viol is $\hat{\beta}_1 = -0.0018(0.0178)$ and the coefficient of IBN is $\hat{\beta}_2 = -0.0148(0.0051)$, whereas, in the case of the spatial conditional normal binomial model including the same two explanatory variables, these values were $\hat{\beta}_1 = -0.0052(0.0149)$ and $\hat{\beta}_2 = -0.0155(0.0046)$, respectively. Here, only the variable IBN is statistically significant, according to its 95% credible intervals, in both cases. In the binomial BYM model including the variables IBN and Pay, the coefficient of IBN is $\hat{\beta}_1 = -0.0116(0.0047)$ and the coefficient of Pay is $\hat{\beta}_2 = -0.0344(0.0151)$, whereas, for the spatial conditional normal binomial model containing these two explanatory variables, the values were $\hat{\beta}_1 = -0.0123(0.0045)$ and $\hat{\beta}_2 = -0.0240(0.0134)$, respectively. Here, the variable IBN is statistically significant in both cases, but the variable Pay is only statistically significant in the binomial BYM model.

In the cases discussed above, the estimated coefficient of the covariate IBN seems to be slightly larger for the spatial conditional models and the variable Pay is statistically significant only for the BYM model. In general, the effect of the covariates seems to be rather different, as it is in the inference that can be performed on the coefficients of the explanatory variables. This is something that we can further investigate by examining the marginal effects at the means for the explanatory variables in some of the fitted binomial BYM models, which are included in Table 2.22. Here, we can confirm that the estimated size of the effects obtained from the fitting of the BYM models considered is smaller than the effects obtained from the fitting of the spatial conditional normal binomial models.

Moreover, the absolute values of these effects reported in INLA are slightly smaller than the absolute values for the marginal effects at the means obtained from the fitting of the spatial conditional normal binomial models in JAGS reported in Table 2.16 in Section 2.7.1. We also notice that the credible intervals obtained in JAGS for these marginal effects are wider than the ones obtained with INLA. This could be due to the different estimation methods that are used within the MCMC and the INLA approaches. In any case, the effects obtained in both software packages are rather small, but significant for the variable IBN.

Table 2.22: Marginal effects at the means for some of the models fitted in INLA to the mother's postnatal period screening test in Colombia data set.

		Mean	SD	95% CI
Spatial conditional normal binomial model in equation (2.36)	**Viol**	-0.0038	(0.0086)	(-0.0230,0.0121)
	IBN	-0.0094	(0.0027)	(-0.0146,-0.0042)
Spatial conditional normal binomial model in equation (2.37)	**IBN**	-0.0094	(0.0028)	(-0.0142,-0.0046)
Spatial conditional normal binomial model in equation (2.38)	**IBN**	-0.0076	(0.0027)	(-0.0145,-0.0026)
	Pay	-0.0140	(0.0076)	(-0.0269,2.734e-04)
Binomial BYM model including the variables Viol and IBN	**Viol**	-0.0017	(0.0101)	(-0.0229,0.0161)
	IBN	-0.0087	(0.0033)	(-0.0162,-0.0025)
Binomial BYM model including the variable IBN	**IBN**	-0.0089	(0.0029)	(-0.0142,-0.0041)
Binomial BYM model including the variables IBN and Pay	**IBN**	-0.0071	(0.0027)	(-0.0124,-0.0025)
	Pay	-0.0197	(0.0093)	(-0.0373,-0.0013)

Regarding the parameter τ_η, which is the variance of the spatially distributed random effect, large standard deviations were obtained for its estimates, a fact that makes it difficult to perform inference on this term. For example, for the models where the probability of success follows the regression in equation (2.41), the value was $\hat{\tau_\eta} = 0.0290(0.0352)$ and, for the model including the explanatory variables Viol and IBN, this value was $\hat{\tau_\eta} = 0.1869(0.1536)$.

For some of these models, the value of the parameter τ_η was larger than the variance of the unstructured random effect τ, giving more importance to the variability explained by the spatial structure than to variability explained by the unstructured effect. This is the case for the models containing the variables Viol, IBN and HE, the model including Viol and IBN, the model containing only the covariate Viol and the model which only includes the variable IBN. For the rest of the models, the values of τ_η are smaller than the values for τ, giving thus more importance to the unexplained variability.

We can assess the predictive accuracy of the BYM models fitted by examining the scatterplots of the observed versus the predicted proportions in Figures 2.14(d), 2.14(e) and 2.14(f). We can see that these plots are also very similar to the ones obtained for the spatial conditional normal binomial models shown in Figures 2.14(a), 2.14(b) and 2.14(c).

Finally, it could also be convenient to compare the estimations obtained for the spatial conditional normal binomial models implemented in JAGS, in Table 2.15 and the results obtained when fitting this model in R-INLA, shown in Table 2.21. We notice that the values of the coefficients are very similar, which is not the case for the values obtained for the information criteria, as these considerably differ from one software to

the other. As we had already mentioned, the information criteria values obtained with different software are not necessarily always equivalent and thus, cannot be directly compared.

2.7.5 Comparisons to the BYM2 model for binomial responses

Finally, we have also fitted the binomial BYM2 model to the data set considered, which solves the identifiability issue of the BYM model. Here, we assume that the probabilities of success are given by equation:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{HE}_i + \beta_4 \text{Pay}_i + \frac{1}{\sqrt{\tau_s}} \left(\sqrt{1-\phi_s}\nu_i + \sqrt{\phi_s}\eta_i \right), \quad (2.42)$$

where $\nu_i,\ i = 1, \ldots, n$, is a set of normally distributed random effects and $\eta_i,\ i = 1, \ldots, n$ is a set of spatially structured random effects following an intrinsic CAR prior distribution, each one also with scaled variance approximately equal to one. The unknown precision parameter τ_s captures the variance contribution from the sum of the two random effects, and the mixing parameter ϕ_s controls for the variance contribution of the spatially structured effect $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$, whereas the variance contribution of the unstructured random effect $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_m)^\top$ is explained by $1 - \phi_s$.

As in the Poisson case in Section 2.6.5, here we will also assign complexity prior for the precision and mixing parameters. Under the probability statement $\text{Prob}(1/\sqrt{\tau_s} > U) = \alpha$, for the parameter τ_s, we will assume an upper bound for the marginal standard deviation of 0.2, and set $U = 0.2/31$ and $\alpha = 0.01$. Regarding the mixing parameter ϕ_s, we will assume that $\text{Prob}(\phi_s < 0.5) = 2/3$, which would represent the initial assumption that the proportion of the variability captured by the unstructured random effect $\boldsymbol{\nu}$ is larger than the one explained by the spatially structured effect $\boldsymbol{\eta}$. In addition, we will assume noninformative normal priors for the regression parameters and the same first order neighbourhood structure for the spatial weights matrix that we have been using so far.

Table 2.23 includes the results for the estimation of the binomial BYM2 model in equation (2.42) and some of its reduced versions. The estimations obtained for the coefficients are very similar to the ones obtained for the BYM models in Table 2.20, which also occurs for the information criteria values for each of the models. The mixing parameter ϕ_s, explaining the amount of variability captured by the spatial structure, is capturing approximately more than 20% of the variability in all the fitted models. As a way to compare these results with previous ones, for example, for the model in equation (2.42), this parameter's estimate was $\hat{\phi}_s = 0.2499(0.2274)$, and, for the model including the variable Viol, it was $\hat{\phi}_s = 0.3472(0.2339)$, explaining around 35% of the variability in the model.

Table 2.23: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for some of the reduced versions of the binomial BYM2 model fitted in INLA to the mother's postnatal period screening test in Colombia data set.

		Intercept	**Viol**	**IBN**	**HE**	**Pay**	τ_s	ϕ_s
DIC = 211.8	Mean	1.9730	-0.0297	-0.0060	0.0041	-0.0528	17.0301	0.2499
WAIC = 212.1	SD	(0.6034)	(0.0168)	(0.0060)	(0.0101)	(0.0172)	(9.5273)	(0.2274)
	95% CI	(0.7881,3.1777)	(-0.0631,0.0034)	(-0.0181,0.0056)	(-0.0154,0.0244)	(-0.0862,-0.0183)	(5.7421,41.6317)	(0.0088,0.8175)
DIC = 211.7	Mean	1.0562	-	-0.0109	2.050e-04	-0.0338	14.1501	0.2767
WAIC = 211.4	SD	(0.3199)	-	(0.0056)	(0.0102)	(0.0143)	(7.0771)	(0.2446)
	95% CI	(0.4304,1.6959)	-	(-0.0223,-5.217e-05)	(-0.0198,0.0205)	(-0.0620,-0.0053)	(5.2504,32.2471)	(0.0093,0.8601)
DIC = 212.3	Mean	1.9571	-0.0370	-	0.0107	-0.0621	18.3813	0.2365
WAIC = 213.5	SD	(0.5915)	(0.0148)	-	(0.0075)	(0.0141)	(10.7347)	(0.2173)
	95% CI	(0.7962,3.1396)	(-0.0669,-0.0083)	-	(-0.0035,0.0262)	(-0.0899,-0.0339)	(6.0476,46.4173)	(0.0087,0.7902)
DIC = 211.3	Mean	2.0213	-0.0282	-0.0074	-	-0.0505	18.3485	0.2363
WAIC = 212.1	SD	(0.5787)	(0.0159)	(0.0044)	-	(0.0155)	(10.5183)	(0.2191)
	95% CI	(0.8866,3.1793)	(-0.0598,0.0034)	(-0.0166,9.140e-04)	-	(-0.0804,-0.0189)	(6.1210,45.7448)	(0.0084,0.7957)
DIC = 211.2	Mean	1.1202	-1.155e-05	-0.0163	-0.0057	-	9.2955	0.3133
WAIC = 209	SD	(0.6226)	(0.0165)	(0.0061)	(0.0113)	-	(3.7846)	(0.2422)
	95% CI	(-0.0961,2.3643)	(-0.0335,0.0316)	(-0.0284,-0.0044)	(-0.0273,0.0172)	-	(4.0453,18.6043)	(0.0161,0.8568)
DIC = 210.4	Mean	0.9749	-4.331e-04	-0.0144	-	-	9.4230	0.3311
WAIC = 208.4	SD	(0.5722)	(0.0164)	(0.0048)	-	-	(3.7397)	(0.2396)
	95% CI	(-0.1205,2.1385)	(-0.0337,0.0311)	(-0.0240,-0.0051)	-	-	(4.1614,18.5833)	(0.0220,0.8557)
DIC = 213.9	Mean	2.1684	-0.0390	-	-	-0.0615	20.2426	0.1986
WAIC = 216.7	SD	(0.5635)	(0.0144)	-	-	(0.0137)	(13.0036)	(0.1929)
	95% CI	(1.0742,3.3090)	(-0.0682,-0.0111)	-	-	(-0.0887,-0.0344)	(6.1679,54.4461)	(0.0073,0.7157)
DIC = 211	Mean	1.0584	-	-0.0109	-	-0.0338	15.0208	0.2713
WAIC = 211.1	SD	(0.1862)	-	(0.0043)	-	(0.0133)	(7.5551)	(0.2423)
	95% CI	(0.6997,1.4341)	-	(-0.0196,-0.0028)	-	(-0.0602,-0.0071)	(5.5538,34.3513)	(0.0091,0.8549)
DIC = 213.7	Mean	0.7795	-0.0114	-	-	-	7.2307	0.3472
WAIC = 212.8	SD	(0.6430)	(0.0183)	-	-	-	(2.8046)	(0.2339)
	95% CI	(-0.4538,2.0847)	(-0.0487,0.0238)	-	-	-	(3.2804,14.1006)	(0.0302,0.8501)
DIC = 210	Mean	0.9590	-	-0.0143	-	-	9.9140	0.3181
WAIC = 208.2	SD	(0.1978)	-	(0.0046)	-	-	(3.9162)	(0.2329)
	95% CI	(0.5770,1.3566)	-	(-0.0237,-0.0054)	-	-	(4.3864,19.5017)	(0.0223,0.8401)
DIC = 214	Mean	0.6673	-	-	-	-0.0420	12.4944	0.2328
WAIC = 215.7	SD	(0.1146)	-	-	-	(0.0135)	(6.1876)	(0.2160)
	95% CI	(0.4460,0.9006)	-	-	-	(-0.0696,-0.0157)	(4.6761,28.2902)	(0.0082,0.7848)

In addition, the marginal effects at the means computed for some of the fitted binomial BYM2 models are included in Table 2.24. These effects are quite similar to those obtained for the binomial BYM models shown in Table 2.22.

Table 2.24: Marginal effects at the means for some of the binomial BYM2 models fitted in INLA to the mother's postnatal period screening test in Colombia data set.

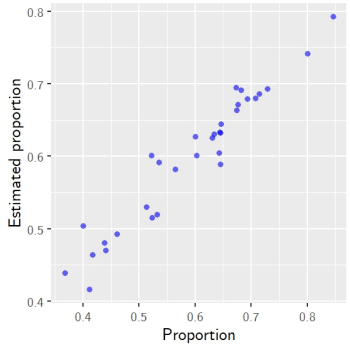
		Mean	SD	95% CI
Binomial BYM2 model including the variables Viol and IBN	**Viol**	-0.0013	(0.0094)	(-0.0211,0.0151)
	IBN	-0.0085	(0.0030)	(-0.0147,-0.0026)
Binomial BYM2 model including the variable IBN	**IBN**	-0.0086	(0.0027)	(-0.0133,-0.0039)
Binomial BYM2 model including the variables IBN and Pay	**IBN**	-0.0066	(0.0024)	(-0.0122,-0.0029)
	Pay	-0.0199	(0.0079)	(-0.0359,-0.0049)

The scatterplots of the observed versus the predicted proportions, obtained from the fitting of some of the binomial BYM2 models, are included in Figures 2.14(g), 2.14(h) and 2.14(i). These scatterplots are almost identical to the ones obtained from the fitting of the binomial BYM models, showing no improvements in terms of predictive accuracy. This confirms the fact that, although the BYM2 model does not represent a better fit than the BYM model in terms of information criteria or, in terms of predictive accuracy, it does offer a considerable advantage, given by the fact that it allows to quantify the amount of variability explained by the spatial structure.

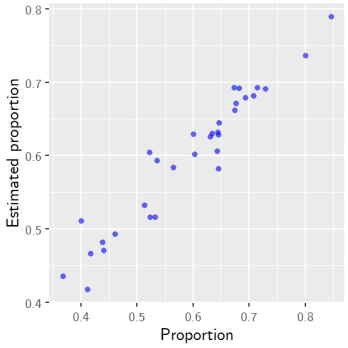
The predictive accuracy of the models can also be assessed by examining the maps for the observed and the predicted proportions obtained from some of the fitted models to the Colombia mother's postnatal period screening test data set. In Figure 2.15 we can compare the map of the observed proportions from Figure 2.15(a) with the maps of the estimated proportions in Figures 2.15(b), 2.15(c) and 2.15(d) obtained after fitting the spatial conditional normal binomial models where the probability of success follows equations (2.36), (2.37) and (2.38), respectively. In general, we could say that the predictions are accurate, as the models are able to generate estimated proportions which are similar to the observed ones for most of the regions.

The maps of the predicted proportions obtained by fitting some of the binomial BYM models are shown in Figures 2.15(j), 2.15(k) and 2.15(l), and for some of the binomial BYM2 models, in Figures 2.15(m), 2.15(n) and 2.15(o). These maps are also very similar, to the spatial conditional normal binomial models shown in Figures 2.15(g), 2.15(h) and 2.15(i), with the exception of a few regions that differ in their predictions.

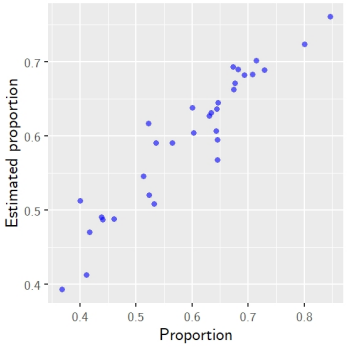
In any case, taking into account that the information criteria values obtained for the BYM and BYM2 models were almost identical to those of the spatial conditional normal binomial models, and also, that there were no improvements in the predictive accuracy, we believe that these models do not necessarily offer a better fit than that of the spatial conditional models.



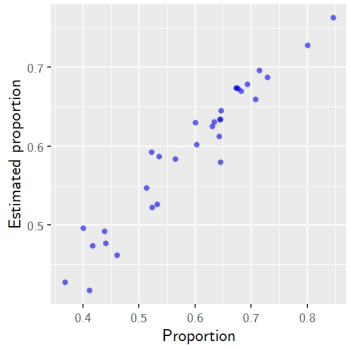
(a) Spatial conditional normal binomial model in equation (2.36) fitted in INLA.



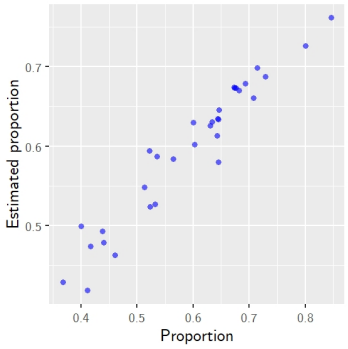
(b) Spatial conditional normal binomial model in equation (2.37) fitted in INLA.



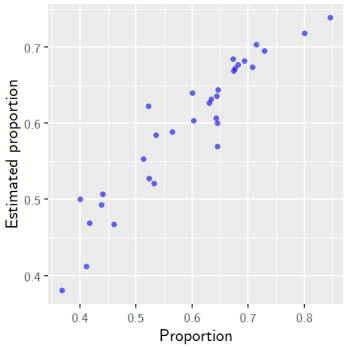
(c) Spatial conditional normal binomial model in equation (2.38) fitted in INLA.



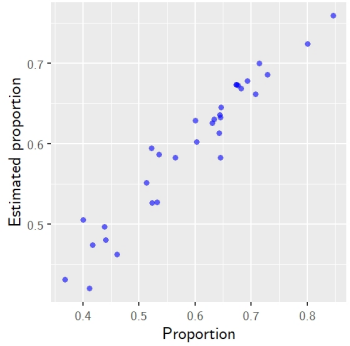
(d) Binomial BYM model including the variables Viol and IBN fitted in INLA.



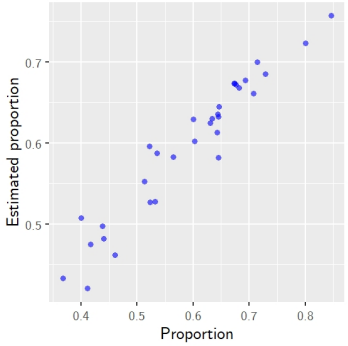
(e) Binomial BYM model including the variable IBN fitted in INLA.



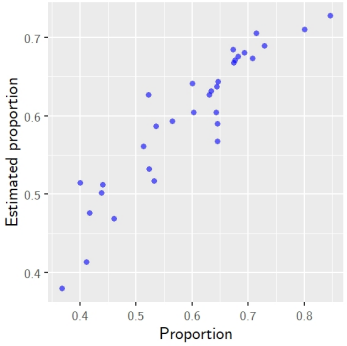
(f) Binomial BYM model including the variables IBN and Pay fitted in INLA.



(g) Binomial BYM2 model including the variables Viol and IBN fitted in INLA.

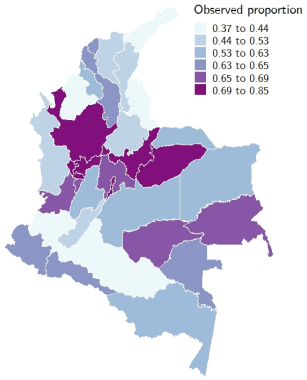


(h) Binomial BYM2 model including the variable IBN fitted in INLA.

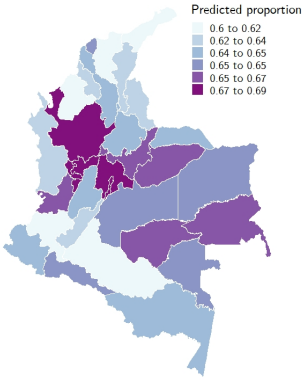


(i) Binomial BYM2 model including the variables IBN and Pay fitted in INLA.

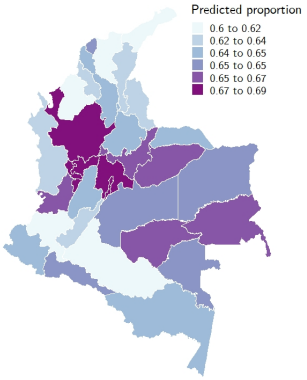
Figure 2.14: Scatterplots for the observed versus the predicted proportions obtained from some of the fitted models to the Colombia mother's postnatal period screening test data set, in INLA.



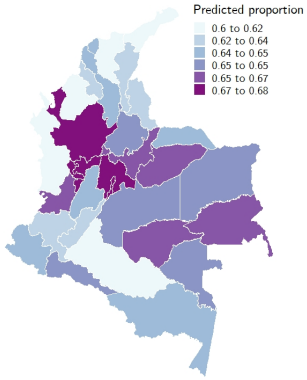
(a) Observed proportions.



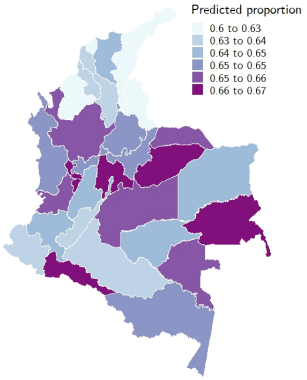
(b) Spatial conditional normal binomial model in equation (2.36) fitted in JAGS.



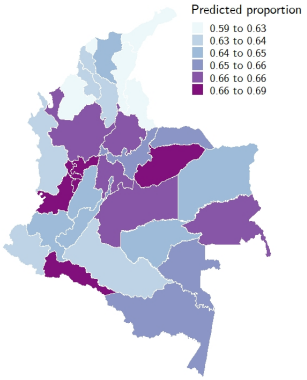
(c) Spatial conditional normal binomial model in equation (2.37) fitted in JAGS.



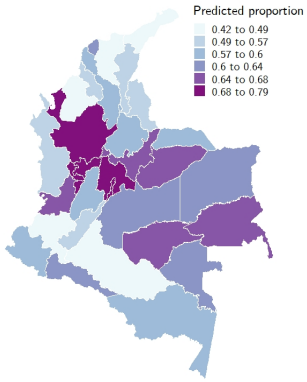
(d) Spatial conditional normal binomial model in equation (2.38) fitted in JAGS.



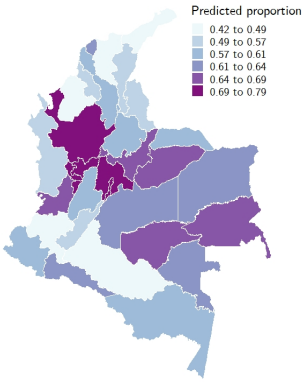
(e) Generalized spatial conditional normal binomial model in equation (2.39) fitted in JAGS.



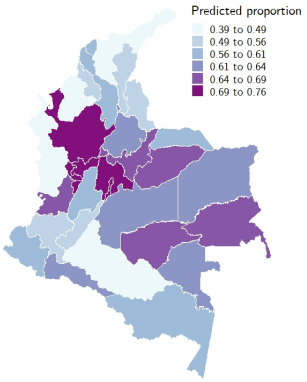
(f) Generalized spatial conditional normal binomial model in equation (2.40) fitted in JAGS.



(g) Spatial conditional normal binomial model in equation (2.36) fitted in INLA.

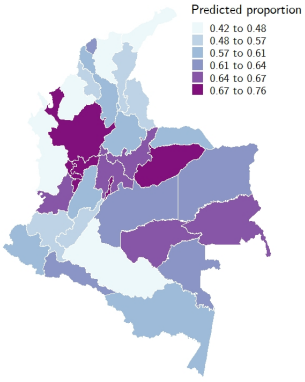


(h) Spatial conditional normal binomial model in equation (2.37) fitted in INLA.

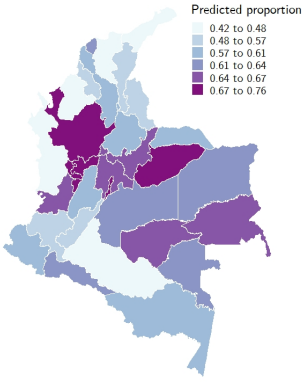


(i) Spatial conditional normal binomial model in equation (2.38) fitted in INLA.

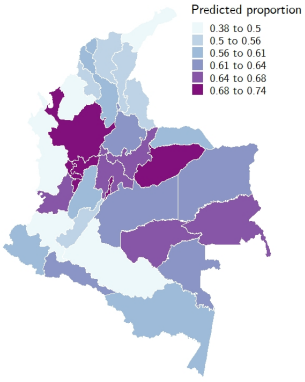
Figure 2.15: Maps for the observed and the predicted proportions obtained from some of the fitted models to the Colombia mother's postnatal period screening test data set.



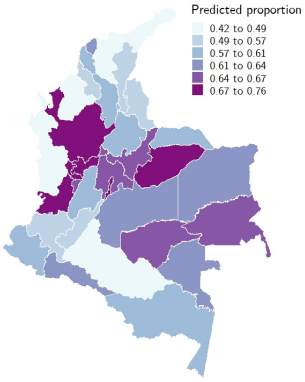
(j) Binomial BYM model including the variables Viol and IBN fitted in INLA.



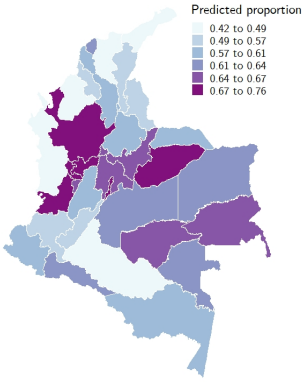
(k) Binomial BYM model including the variable IBN fitted in INLA.



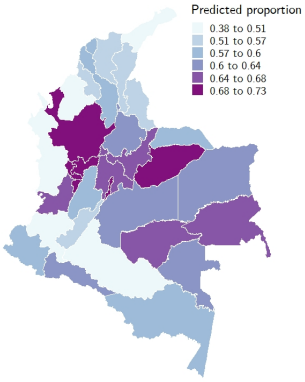
(l) Binomial BYM model including the variables IBN and Pay fitted in INLA.



(m) Binomial BYM2 model including the variables Viol and IBN fitted in INLA.



(n) Binomial BYM2 model including the variable IBN fitted in INLA.



(o) Binomial BYM2 model including the variables IBN and Pay fitted in INLA.

Figure 2.15: Maps for the observed and the predicted proportions obtained from some of the fitted models to the Colombia mother's postnatal period screening test data set (Continued).

2.8 Discussion

In this chapter we have performed a revision of Bayesian spatial conditional overdispersion models (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018) for area count data and also provided a comparison with the well known Besag-York-Mollié (BYM) models (Besag, York and Mollié, 1991), widely used in the literature for spatial count data modelling. In this context, we would like to emphasize the importance of taking into account the overdispersion, as well as the dependence that can arise from the correlation among the values of the response variable in neighbouring locations. We believe that the spatial

conditional models offer a good, flexible, reasonable, and worth considering alternatives to the BYM model. Moreover, we have shown their usefulness and appropriateness by fitting them to infant mortality and to mother's postnatal period screening test data from Colombia.

We have fitted these models to data sets similar to the ones analysed by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) and we have performed a thorough analysis of the results obtained from these fittings. We have also provided posterior predictive checks, such as the scatterplots of observed versus fitted values and maps of the predictions obtained. Moreover, we have computed the marginal effects at the means of the covariates included in the models, offering a better understanding of their effect over the responses. Another fact we believe is worth mentioning is the application of the BYM2 models, an innovative approach that allows to overcome the issue of non identifiability of the BYM models. In addition, convergence diagnostics were applied and illustrated for the MCMC chains for some of the models considered.

In the work by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), the models were fitted using the software OpenBUGS. Here, besides using this same program, we have also fitted models in JAGS, which follows the same MCMC approach and, in addition, we have implemented some of the models in INLA, an alternative to the MCMC approach. We have offered comparison between these two different approaches and highlighted their advantages and disadvantages in each case.

The spatial conditional normal Poisson and negative binomial models were fitted for the case of Poisson distributed response variables, showing to be good candidates for fitting the infant mortality rates data from Colombia. Moreover, they were also able to account for overdispersion, as well as to explain the intensity of the spatial autocorrelation that was present in the specific data set under study. More specifically, according to the DIC and WAIC information criteria values, the spatial conditional normal Poisson model was selected as the best fitting model. We were able to fit this model and its reduced versions in OpenBUGS and R-INLA and, by setting the same prior distributions, we obtained very close results in both implementations. Given that these two software packages are based on different methodologies, we believe it was convenient and necessary to confirm the consistency of the estimations in both software packages.

Results obtained from the fitting of the aforementioned models were consistent with the results obtained in previous analyses of a similar data set, reported in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), where the authors modelled mortality rates for children younger than five years old of age in Colombia. Their analysis found positive significant relations with the variable explaining the index of unsatisfied basic needs (i.e., IBN) and negative significant relations with the variable representing the resources provided by the government for academic achievement of the population (i.e., Rec) when assessing their effect on the infant mortality rates under study. In our case, for infant mortality rates for children under one year of age, this significant relationships also hold and, in addition, a positive significant relationship was also identified with the percentage of women who had suffered physical violence from their current partners

(i.e., Viol). As in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), in this application we have also found evidence of positive spatial autocorrelation in the data under study. Moreover, we have also fitted the generalized spatial conditional normal Poisson model, which provided more flexibility by allowing the dispersion to vary as a function of some explanatory variables.

In order to be able to compare the performance of the spatial conditional models, we have also fitted the BYM model to the data under study. There were some difficulties when fitting these models in OpenBUGS, which could be due to a number of problems, such as, for example, a possible existing conflict between the data and the assumed prior distributions (Spiegelhalter et al., 2002). In any case, these difficulties did not represent a problem for fitting these model in R-INLA, an issue that was confirmed with the models fitted in previous sections for this particular data set. The estimated parameters obtained by fitting the BYM model showed quite large standard deviations, especially for the variance parameters for both the spatially correlated and the uncorrelated random effects. However, the model seemed to favour the spatial structure over the extra-variability, although, because of the way these effects are specified in the BYM model, no further information about the specific spatial dependence could be obtained from it.

Furthermore, in order to be able to provide more information than the one obtained from the fitting of the standard BYM model, we have also fitted the BYM2 model, which allowed us to identify the spatially structured and the unstructured effects separately. As can be seen in the reported results, this model provided useful findings about the amount of variability in the data that could be explained by the assumed spatial structure in the fitted models. In addition, we were also able to conclude that no significant improvements were suggested by the information criteria values, or in terms of posterior predictive accuracy, when the BYM2 model was compared to the previously fitted spatial conditional and to the BYM models.

We have also performed posterior predictive checks on the fitted models, concluding that they can provide a reasonable accuracy in the predictions of the mortality rates for most cases, especially for the spatial conditional normal Poisson and the generalized spatial conditional normal Poisson models. In our view, and based on the previously reported results, the performance of the spatial conditional normal Poisson model was considerably better, keeping in mind that the information criteria slightly favoured it over the BYM model, and also taking into account the aforementioned issues when fitting the latter. The spatial conditional normal Poisson model allowed for the overdispersion to be taken into account and, unlike the BYM model, it provided information on the type, and also the strength, of the spatial association which was present in the data. Besides, with the results obtained from the fitting of this model, appropriate and well justified inference could be made about the regression parameters in the model.

For the case of binomially distributed response variables, the spatial conditional normal binomial and the beta binomial models were applied to the Colombian mother's postnatal period screening test data set where, according to the information criteria values and the posterior predictive checks performed, the spatial conditional normal binomial was selected as the best fitting model. This model and its reduced versions

provided a good fit and we were able to properly accommodate the overdispersion and spatial dependence in the data. In this case, the software packages used to implement the models were JAGS and R-INLA, obtaining similar results for the same models with the same assumed prior distributions in both cases.

Reported results suggest that, in regions with large values of the variable measuring the percentage of the population that has basic needs not satisfactorily attended (i.e., variable IBN), the probability that a mother goes through a postnatal period screening test increases. This fact is consistent with the results found by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) when analysing a similar data set. In addition, although it did not provide a better fit than the previously fitted models, by fitting the generalized spatial conditional normal binomial models, we found a statistically significant association of the dispersion with the variable IBN, and also with the spatial term.

As was done for the infant mortality data, for the mother's postnatal period screening test data, we have also fitted the BYM and BYM2 models and compared them with the best fitting models we had so far, which were the spatial conditional normal binomial models. Here, the binomial BYM models offered no improvements neither in terms of information criteria nor in terms of predictive accuracy. In addition, there were models that favoured the unexplained over the spatially structured variability, whereas for other models it was the opposite way, a fact that made it difficult to obtain a clear information about the spatial association that was present in the data.

Finally, with regard to model implementation, we can mention that the models' fitting in the software package OpenBUGS was quite flexible, mainly because it allows the researcher to specify any kind of Bayesian models in a very simple and intuitive way. The same holds for JAGS, as both software are based in the BUGS programming language, with the drawback that fitting spatial models in JAGS, especially when specifying CAR distributions for random effects, is not as simple as it would be in OpenBUGS. Furthermore, model implementation is straightforward in R-INLA, since most models are already specified in this package. However, it becomes more complex when the researcher wishes to employ a particular model, different from the ones already available therein, an issue that also occurs when a different prior specification is required.

More specifically, the implementation of new prior distributions for Bayesian analysis in R-INLA is one of the current main challenges for the developers of the package (Rue et al., 2017). This particular fact will be illustrated in Chapter 6, where we present a class of models that cannot be directly implemented in R-INLA and, therefore, we propose a method that allows to fit these models in this software package. Nevertheless, a point in favour of R-INLA is the much shorter computation time that it requires for the fitting of a model, when compared to OpenBUGS or JAGS.

Chapter 3

Semiparametric extensions of the generalized spatial conditional overdispersion models

3.1 Introduction

In the previous chapter, we have described and applied the generalized spatial conditional overdispersion models in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) (See Sections 2.2.2 and 2.3.2 in Chapter 2). Throughout this thesis, we have shown the great advantages these models offer, since they have been shown to be able to account for overdispersion in count data, to take into account the possible existing spatial correlation, and they are also flexible enough as to allow for the dispersion to vary according to covariates and/or spatial terms.

However, when including covariates in the regression structures for these models, we are assuming that the possible existing relationship between each covariate and the predictor is linear, which may not necessarily be the case, since they could be given by another maybe non linear pattern (Ruppert, Wand and Carroll, 2003). In this sense, smoothing methods should be considered in generalized linear models (GLM), so that the linearity hypothesis can be relaxed. Therefore, in this chapter we propose a semiparametric extension of the generalized spatial conditional overdispersion models that will allow us to capture such non linear relations.

In Section 3.2, we provide a short review of the semiparametric modelling approach. In Section 3.3, we propose an extension of the generalized spatial conditional models in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), so that non linear relations can be properly captured. Sections 3.4 and 3.5 include the application of this model proposal to the infant mortality rates and to the mother's postnatal period screening test in Colombia data sets, respectively. Finally, a discussion on the reported results is provided in Section 3.6.

3.2 Semiparametric models overview

Let the variables Y_i , for $i = 1, \ldots, n$, follow a count data distribution with $\mathrm{E}(Y_i) = \mu_i$ and let v_i be a variable observed for the same units, $i = 1, \ldots, n$. If we specify a generalized linear model (GLM) for the response Y_i and the covariate v_i, we could, for example, define the linear predictor so that:

$$g(\mu_i) = \beta_0 + \beta_1 v_i, \tag{3.1}$$

where $g(\cdot)$ is a given link function and β_0 and β_1 are the unknown intercept and the coefficient for the covariate, respectively.

Let us now assume that the relationship between v_i and the link function for the mean is not necessarily linear, but it is given by an unknown smooth function $f(\cdot)$, so that the regression structure for the mean can be specified as:

$$g(\mu_i) = f(v_i) \tag{3.2}$$

Here, the smooth function $f(\cdot)$ could be modelled in many different ways that include, for example, local linear or polynomial regression, kernel smoothing and regression splines, among others (Green and Silverman, 1994). In this work, we will focus on regression splines. A spline is a piecewise function (Smith, 1979), defined for a variable v_i in a series of K fixed knots $\kappa_1, \kappa_2, \ldots, \kappa_K$, which divide the range of v_i into regions. These knots are normally chosen so that $\kappa_1 < \kappa_2 < \cdots < \kappa_K$.

There are numerous spline basis such as, for example, natural cubic splines, truncated polynomials, B-splines and penalized splines, among others (Ruppert, Wand and Carroll, 2003). For example, we can mention the truncated power basis of degree p, that is:

$$1, v_i, \ldots, v_i^p, (v_i - \kappa_1)_+^p, \ldots, (v_i - \kappa_K)_+^p, \tag{3.3}$$

where $(v_i - \kappa_k)_+^p$ is a truncated power function with $(v_i - \kappa_k)_+^p = (v_i - \kappa_k)^p$ if $v_i - \kappa_k > 0$, and $(v_i - \kappa_k)_+^p = 0$ otherwise, for $k = 1, \ldots, K$.

If the smooth function $f(v_i)$ is specified as a spline with this basis, it would be defined in the following way:

$$f(v_i, \boldsymbol{\theta}) = \beta_0 + \beta_1 v_i + \cdots + \beta_p {v_i}^p + \sum_{k=1}^{K} b_k (v_i - \kappa_k)_+^p, \tag{3.4}$$

with $\boldsymbol{\theta}^\top = (\beta_0, \beta_1, \ldots, \beta_p, b_1, \ldots, b_K)$ being a vector of unknown coefficients that need to be estimated. Note that, if defined in this way, on each interval between two consecutive knots, $f(v_i)$ is a p-th degree polynomial having $(p-1)$ derivatives (Ruppert, 2002).

Basis splines or B-splines (De Boor, 1978) are widely used due to the fact that they have certain properties that offer more numerical stability, when compared to the truncated power basis (Ruppert, Wand and Carroll, 2003). Then, if $\mathbf{B}$ is the $n \times (K + p)$ matrix of p-th degree B-splines for K knots, with elements $B_{i,j} = B_j(v_i)$, for $i = 1, \ldots, n$

and $j = 1, \ldots, K + p$, then the smooth function $f(v_i)$ could be written in the following way:

$$f(v_i, \mathbf{b}) = \sum_{j=1}^{K+p} b_j B_j(v_i), \tag{3.5}$$

with $\mathbf{b}$ being the vector of unknown coefficients, that is $\mathbf{b}^\top = (b_1, \ldots, b_{K+p})$.

B-splines are completely defined by the selection and location of the knots (Eilers and Marx, 2021). In this way, having too many knots could result in overfitting, whereas a small number of knots may not be flexible enough to represent the data. In this sense, penalized splines or P-splines for GLMs were proposed by Eilers and Marx (1996). The relevance of the number of knots is reduced by imposing a penalty that controls for the smoothing of the data. Moreover, the dimensionality of the problem is also reduced to the number of chosen knots instead of the number of observations (Ruppert, 2002).

P-splines are specified by considering a spline basis and modifying the likelihood function by imposing a penalization based on differences between adjacent coefficients. For example, if we consider a p-th degree B-spline basis, the smooth function $f(\cdot)$ would be given by equation (3.5). Then, we could impose a penalization on the coefficients $\mathbf{b}$ (i.e., $\lambda \mathbf{b}^\top \mathbf{D}^\top \mathbf{D}\mathbf{b}$), so that the penalized likelihood would be:

$$l_p(\mathbf{y}|\boldsymbol{\theta}) = l(\mathbf{y}|\boldsymbol{\theta}) - \frac{\lambda}{2} \mathbf{b}^\top \mathbf{D}^\top \mathbf{D}\mathbf{b}, \tag{3.6}$$

where $\mathbf{y}$ is the vector of observations of the response variable, $l(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood of the data given the parameters and λ is the penalization parameter, which is considered to be fixed. In addition, $\mathbf{D}$ is a $(K+1) \times (K+p)$ matrix that computes the second order differences among adjacent coefficients of the basis.

Currie and Durbán (2002) proposed a formulation that allows to write a P-spline regression model as a mixed model. Although the authors here only referred to Gaussian models, the extension of their proposed methodology is also possible for GLMs, leading to the specification of a generalized linear mixed model (Ruppert, Wand and Carroll, 2003).

In particular, if we start from the model in equation (3.2), where the smooth function $f(\cdot)$ is defined with a B-spline basis, such as the one in equation (3.5), we could impose the penalization over $\mathbf{b}$ and the model could be written in the following way:

$$g(\mu_i) = f(v_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i \boldsymbol{\alpha}, \tag{3.7}$$

where $\mathbf{Z}_i$ is the i-th row of the $n \times (K+1)$ matrix $\mathbf{Z}$, with K being the number of chosen knots, so that $\mathbf{Z} = \mathbf{B}\mathbf{D}^\top(\mathbf{D}\mathbf{D}^\top)^{-\mathbf{1}}$, $\mathbf{B}$ is the B-splines basis matrix and $\mathbf{D}$ as before. In addition, $\boldsymbol{\alpha}^\top = (\alpha_1, \ldots, \alpha_{K+1})$ is a vector of independent random coefficients with unknown variance τ_α, so that $\alpha_k \sim N(0, \tau_\alpha)$, $\tau_\alpha > 0$, for $k = 1, \ldots, K+1$. Note that β_0, β_1 and v_i are as before. Here, the number of knots is usually defined between $K = 5$ and $K = 40$, depending on the size of the data set under study (Ruppert, 2002). Additional details on this specification can be found in Currie and Durbán (2002).

This specification is particularly useful, since it allows to employ the existing methodology for the estimation and inference of mixed models, which is much more appealing than the one for nonparametric models and P-splines. In addition, it offers great flexibility, as it is possible to include these semiparametric terms in any kind of model. Moreover, the implementation of these models is quite simple when employing Bayesian estimation approaches such as MCMC (Crainiceanu, Ruppert and Wand, 2005). A detailed discussion about these advantages is included in Eilers, Marx and Durbán (2015).

Semiparametric modelling with P-splines is often found in the literature. For example, Crainiceanu, Ruppert and Wand (2005) illustrated how to fit P-spline regression models in the WinBUGS software package, by using the mixed models representation. Among the examples presented therein, they included a Bernoulli model applied to a data set concerning the wages and union membership of 534 workers in the USA for the year 1985. A smooth function of the wage per hour was included in the regression for the logistic function of the probability of success in the way of a mixed model specification.

Kazembe (2009) studied the determinants of fertility for women in Malawi for the year 2000. The author fitted a Poisson regression model for the number of children ever born per woman, where he included smooth functions of some variables specified by using P-splines. He found evidence of the existence of non linear relations between the response and the year of marriage, and also the age at which women got married.

Ugarte et al. (2010) studied the mortality risk of breast cancer in 50 Spanish provinces between 1975 and 2005. For three age groups, the authors fitted a spatio-temporal model including a smooth function of longitude, latitude and time, modelled as a P-spline. Their results suggested a different behaviour of the mortality risks in the provinces for each group but, in general, they observed a gradual decline, which was slower for the oldest age group.

3.3 Semiparametric extensions of the generalized spatial conditional overdispersion models

We propose to extend the models in Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón (2012) and Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) so that, in addition to being able to capture the overdispersion and the spatial correlation in the data, they are also able to account for possible non linear relationships between the covariates and the predictor. For brevity of exposition and, as in the previous Sections, we will only refer to the situation when there is only one covariate under analysis for this type of non linear relationships.

Let us consider the case where we have Poisson distributed count variables Y_i, observed in $i = 1, \ldots, n$, areas. The spatial conditional normal Poisson model has been shown to be an adequate candidate for fitting this type of data (see Chapter 2). In this model, we assume that $(Y_i \mid Y_{\sim i}, \nu_i) \sim \text{Poi}(\mu_i)$, for $i = 1, \ldots, n$, where $Y_{\sim i}$ is the set of all the neighbours of the i-th area, excluding the i-th area itself and ν_i is a Gaussian random effect, so that $\nu_i \sim N(0, \tau)$, with $\tau > 0$ being the unknown dispersion parameter.

If we assume that the relationship between a variable v_i and $\log(\mu_i)$, for $i = 1, \ldots, n$,

is not necessarily linear, but that is given by an unknown smooth function $f(\cdot)$, then we could fit the following model for the conditional mean:

$$\log(\mu_i) = f(v_i) + \rho \mathbf{W}_i \mathbf{y} + \nu_i, \tag{3.8}$$

where $\mathbf{W}_i$ is the i-th row of the spatial weights matrix $\mathbf{W}$ and $\mathbf{y}$ is the vector of observations of the response variable. Therefore, $\mathbf{W}_i\mathbf{y}$ is the spatial lag of the response variable. In addition, ρ is the spatial parameter.

We propose to use the P-splines mixed model representation of the smooth function $f(\cdot)$ (Currie and Durbán, 2002), described in equation (3.7), where $f(v_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha}$, so that the linear predictor in equation (3.8) can be rewritten as:

$$\log(\mu_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha} + \rho \mathbf{W}_i \mathbf{y} + \nu_i, \tag{3.9}$$

where β_0, β_1, v_i, $\mathbf{Z}_i$, $\boldsymbol{\alpha}$, ρ, $\mathbf{W}_i\mathbf{y}$ and ν_i are as before.

Here, we could also consider the inclusion of other covariates, assuming that their possible relationship with the predictor is linear. Therefore, let $\mathbf{x}_i$ be the $q \times 1$ vector of q new covariates, corresponding to the i-th observation. If this term was included in equation (3.9), the linear predictor would be given by:

$$\log(\mu_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha} + \mathbf{x}_i^\top \boldsymbol{\delta} + \rho \mathbf{W}_i \mathbf{y} + \nu_i, \tag{3.10}$$

where $\boldsymbol{\delta}^\top = (\delta_1, \ldots, \delta_q)$ is the vector of unknown coefficients for the new covariates and the remainder terms are as before.

In the case of binomially distributed response variables, we could assume, for example, the spatial conditional normal binomial model for the variables Y_i, observed in $i = 1, \ldots, n$, areas. That is $(Y_i \mid Y_{\sim i}, \nu_i) \sim \text{Bin}(n_i, \pi_i)$, for $i = 1, \ldots, n$, with n_i and π_i the number of trials and the probability of success on the i-th region. Assuming that the relationship between the variable v_i and the logistic function of π_i, for $i = 1, \ldots, n$, is given by a smooth function $f(\cdot)$ and, considering its P-splines mixed model representation (i.e., $f(v_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha}$), the regression model would be defined as follows:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha} + \rho A_i + \nu_i, \tag{3.11}$$

with A_i being the i-th element of the spatial term $\mathbf{A}$ (See Section 2.3.1 in Chapter 2 for the definition of this term). In addition, β_0, β_1, v_i, $\mathbf{Z}_i$, $\boldsymbol{\alpha}$, ρ and ν_i are as before. Note that here, we could also include some covariates assuming that their possible relationship with the predictor is linear, as was done in equation (3.10) for the Poisson case.

We also propose the semiparametric extension of the generalized spatial conditional normal Poisson model. For example, we could include the smooth function for the variable v_i (i.e., $f(v_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha}$) in the regression structure for the mean, then the model would be given by:

$$\begin{array}{rcl} \log(\mu_i) & = & \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha} + \mathbf{x}_i^{(\boldsymbol{\mu})\top}\boldsymbol{\delta} + \rho_1 \mathbf{W}_i \mathbf{y} + \nu_i,\ \nu_i \sim N(0, \tau_i),\ \tau_i > 0, \\ \log(\tau_i) & = & \mathbf{x}_i^{(\boldsymbol{\tau})\top}\boldsymbol{\gamma} + \rho_2 \mathbf{W}_i \mathbf{y}, \end{array} \tag{3.12}$$

where $\mathbf{x}_i^{(\boldsymbol{\mu})}$, is the vector of covariates corresponding to the i-th area for the mean structure, where we assume that their possible relationship with the predictor is linear, and $\mathbf{x}_i^{(\boldsymbol{\tau})}$ is the vector of covariates for the dispersion structure. Here, $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ are the vectors of unknown coefficients for the variables in $\mathbf{x}_i^{(\boldsymbol{\mu})}$ and $\mathbf{x}_i^{(\boldsymbol{\tau})}$, respectively. In addition, ρ_1 and ρ_2 are the spatial parameters for the mean and dispersion structures, respectively. Note that β_0, β_1, v_i, $\mathbf{Z}_i$, $\boldsymbol{\alpha}$ and $\mathbf{W}_i\mathbf{y}$ are as before.

In the same way, for the generalized spatial conditional normal binomial model, the model can be specified as:

$$\begin{aligned} \text{logit}(\pi_i) &= \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha} + \mathbf{x}_i^{(\boldsymbol{\mu})\top}\boldsymbol{\delta} + \rho_1 A_i + \nu_i, \quad \nu_i \sim N(0,\tau_i),\ \tau_i > 0, \\ \log(\tau_i) &= \mathbf{x}_i^{(\boldsymbol{\tau})\top}\boldsymbol{\gamma} + \rho_2 A_i, \end{aligned} \tag{3.13}$$

where $\mathbf{x}_i^{(\boldsymbol{\mu})}$, $\boldsymbol{\delta}$, $\mathbf{x}_i^{(\boldsymbol{\tau})}$, $\boldsymbol{\gamma}$, β_0, β_1, v_i, $\mathbf{Z}_i$, $\boldsymbol{\alpha}$, ρ_1, ρ_2 and A_i are as before.

At this point, it would be convenient to mention that all the models proposed in this section can be estimated in the same usual way we have used throughout this thesis, employing Bayesian estimation approaches such as MCMC or INLA. This is possible thanks to the fact that we are considering the P-splines representation of the smooth function as a mixed model, since we only need to include the terms corresponding to the matrix $\mathbf{Z}$ in the regression equation, together with their associated random effects $\boldsymbol{\alpha}$. Therefore, it is really only necessary to estimate one additional parameter, given by the precision of these random effects.

3.4 Application to infant mortality in Colombia: Poisson models

In this section, we will fit the proposed semiparametric models to the infant mortality data in Colombia, which was presented in Section 2.6, Chapter 2. Let us recall that this data set consists of multiple variables observed for $n = 32$ departments in this country, which included the number of children under one year of age who died in 2005 (i.e., variable ND), the total number of births in 2005 (i.e., variable NB), an index representing the percentage of the population not having their basic services satisfactorily attended for the same year (i.e., variable IBN), the amount of resources (in thousands of dollars) for academic achievement or education and integral attention for young children provided by the government per household in the year 2005 (i.e., variable Rec), the percentage of women over the age of 18 who had suffered physical violence from their current partners (i.e., variable Viol), the percentage of young people (i.e., between 18 and 24 years) who were able to opt for a higher educational level (i.e., variable HE), and the percentage of children under one year of age who received the third dose of the polio vaccine in the year 2004 (i.e., variable Vac).

In addition, the infant mortality rates can be approximated as the number of children under one year of age who died in the year 2005 per 1000 born alive in each of the departments in Colombia, so that the variable Rates can be obtained as $\text{Rates}_i = \text{ND}_i/\text{NB}_i \times 1000$, for $i = 1, \ldots, n$. Here, it would be useful to mention that the

spatial structure of the data is represented by the spatial weights matrix $\mathbf{W}$ following the contiguity of order one criterion.

In this study, we wish to investigate whether there are non linear relationships between each covariate available in this data set and the mean of the response variable, which is the number of children under one year of age who died in 2005 (i.e., variable ND). Therefore, we have fitted different alternative models, where some of them have included the smooth term associated with each variable and others have excluded it. Then, we have assessed the statistical significance of the included terms, the behaviour of the curve and the information criteria values obtained for each of the fitted models. For brevity of exposition here, we will only illustrate the results obtained for four of these models, where we assume v_i to be any of the variables under study, observed for the $i = 1, \ldots, n$, regions.

The first model considered (i.e., Model 1) is a spatial conditional normal Poisson model where we assume that $(\text{ND}_i \mid \text{ND}_{\sim i}, \nu_i) \sim \text{Poi}(\mu_i)$, for $i = 1, \ldots, n$. Here, we include the explanatory variable v_i, for $i = 1, \ldots, n$, the spatial lag of the rates (i.e., $\mathbf{W}_i\mathbf{Rates}$) and the random effect (i.,e., ν_i) in the regression structure for the conditional mean, so that it can be written in the following way:

$$\log(\mu_i) = \log(\text{NB}_i) + \beta_0 + \beta_1 v_i + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \qquad \nu_i \sim N(0, \tau),\ \tau > 0 \tag{3.14}$$

Note that the logarithm of the number of births (NB_i) has been included as an offset in order to be able to model the mortality rates (see Section 2.6.1 in Chapter 2).

The second alternative (i.e., Model 2) is also a spatial conditional normal Poisson model, but where we consider the relationship between the covariate and the mean of the response to be given by a smooth function $f(\cdot)$. Here, we will use the mixed model representation of $f(\cdot)$ (i.e., $f(v_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha}$), specifying the model proposed in equation (3.9), so that the regression structure is as follows:

$$\begin{aligned} \log(\mu_i) =& \log(\text{NB}_i) + \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha} + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \\ & \nu_i \sim N(0, \tau),\ \tau > 0,\ \alpha_k \sim N(0, \tau_\alpha),\ \tau_\alpha > 0 \end{aligned} \tag{3.15}$$

The third model (i.e., Model 3) is a generalized spatial conditional normal Poisson model, where we include the covariate as a fixed effect in the regression structure for the conditional mean and the spatial lag in the structure for the variance parameter. In particular, the model is as follows:

$$\begin{aligned} \log(\mu_i) =& \log(\text{NB}_i) + \beta_0 + \beta_1 v_i + \nu_i,\ \nu_i \sim N(0, \tau_i),\ \tau_i > 0 \\ \log(\tau_i) =& \gamma + \rho \mathbf{W}_i \mathbf{Rates} \end{aligned} \tag{3.16}$$

Finally, the last considered model (i.e., Model 4) is a generalized spatial conditional normal Poisson model, including the smooth term for the variable v_i, that is $f(v_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha}$, in the regression structure for the conditional mean. Note that this is the model proposed in equation (3.12) and, for this particular case, it is specified as:

$$\begin{aligned} \log(\mu_i) =& \log(\text{NB}_i) + \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha} + \nu_i, \\ & \nu_i \sim N(0, \tau_i),\ \tau_i > 0,\ \alpha_k \sim N(0, \tau_\alpha),\ \tau_\alpha > 0 \\ \log(\tau_i) =& \gamma + \rho \mathbf{W}_i \mathbf{Rates} \end{aligned} \tag{3.17}$$

Note that the term $\mathbf{Z}_i\boldsymbol{\alpha}$ will be computed in the way shown in equation (3.7), where the number of knots considered will be $K = 5$, so that $\mathbf{Z}$ is a 32×6 matrix and we have $k = 1, \ldots, 6$ random coefficients α_k. This quantity has been chosen due to the limited sample size of the data set under study (i.e., $n = 32$) and also, taking into account the fact that, according to Ruppert (2002), this number of knots is considered to be enough as to ensure flexibility for most applications.

Regarding the estimation of the models, they will be implemented in the JAGS software package, using the MCMC approach, where we will specify the same noninformative prior distributions we have assumed in previous chapters. In particular, normal priors for the fixed effects (i.e., $N(0, 1\text{e-}05)$) and gamma priors for the precision parameters of the random effects (i.e., $G(1\text{e-}04, 1\text{e-}04)$). In addition, model comparison will be carried out by using the Deviance Information Criterion (DIC) and the Watanabe-Akaike Information Criterion (WAIC).

3.4.1 Fitting of the semiparametric generalized spatial conditional over-dispersion models for Poisson responses

Tables 3.1 to 3.5 include the mean, standard deviation and credible interval for the estimated parameters, and DIC and WAIC values obtained after fitting the four models described in the previous Section, considering each of the variables Viol, IBN, Rec, HE and Vac, respectively. Note that in equations (3.14) to (3.17) the term v_i represents the variable under study corresponding to the i-th region.

In order to assess the relationship obtained between the variable and the predictor, Figures 3.1 to 3.6 include the estimated curves obtained for the two models considering the smooth terms for each one of the variables included in the study (i.e., Models 2 and 4). Let us recall that the smooth function $f(\cdot)$ is being represented by $f(v_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha}$. Therefore, the estimated curve can be obtained as $\hat{f}(v_i) = \hat{\beta}_0 + \hat{\beta}_1 v_i + \mathbf{Z}_i\hat{\boldsymbol{\alpha}}$. See Eilers and Marx (2021) for a detailed explanation on how to visualize these curves.

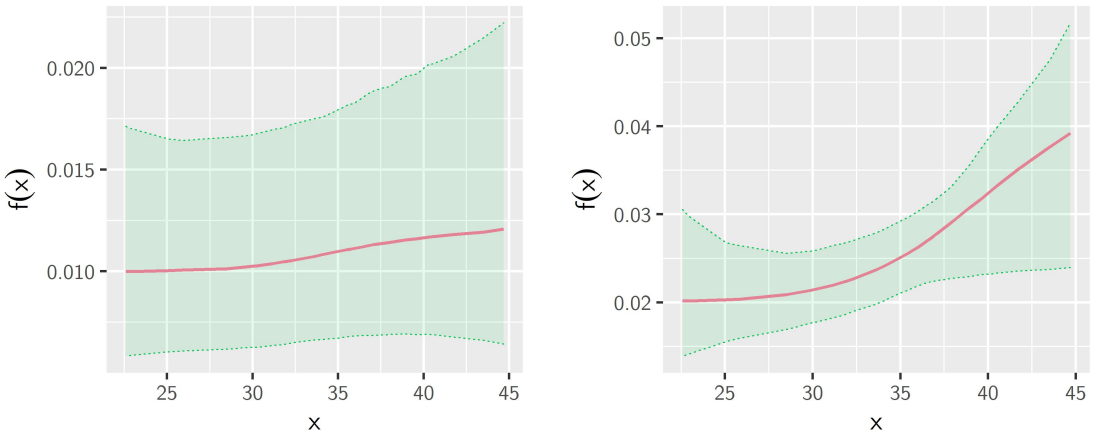
In these figures, the red curve represents the estimated mean obtained for $\hat{f}(v_i)$ and the green bands, its 95% credible interval. We will assume that there is evidence of the existence of a non linear relationship between the covariate under study and the predictor in the cases where it is not possible to draw a straight line that falls within the band corresponding to the credible interval. In other words, if such line can be fitted within the band, we will consider that, if there exists a relationship between the variable and the predictor, there is no reason to believe it may not be linear.

The first variable under study is Viol, with results reported in Table 3.1. Here, we can compare the four models according to their DIC and WAIC values. The smallest ones are given for Model 2, which is the spatial conditional normal Poisson model that includes the smooth function of the variable Viol in the regression equation for the conditional mean. However, the fixed effect in this model is not significant, according to its 95% credible interval and, moreover, the curve obtained from this model shown in Figure 3.1(a) does not suggest any relationship different than the linear one between Viol and the conditional mean of the mortality rates.

In addition, the largest information criteria values were obtained for Model 4, that is the generalized spatial conditional normal Poisson model which also includes the smooth function of Viol. In this case, although the credible interval band is wider than the one from Model 2, the curve shown in Figure 3.1(b) does not suggest any evidence of the existence of any relationship other than the linear one either.

Table 3.1: Results obtained from the fitting of the four models considered for the variable Viol.

		Model 1	Model 2	Model 3	Model 4
Intercept	Mean	-4.8764	-4.7905	-4.7055	-4.5109
	SD	(0.3999)	(0.4842)	(0.4496)	(0.5441)
	95% CI	(-5.6581,-4.0843)	(-5.6841,-3.7910)	(-5.5405,-3.7803)	(-5.4625,-3.3438)
Viol	Mean	0.0103	0.0081	0.0300	0.0265
	SD	(0.0107)	(0.0136)	(0.0137)	(0.0155)
	95% CI	(-0.0113,0.0315)	(-0.0214,0.0326)	(0.0013,0.0541)	(-0.0067,0.0528)
ρ	Mean	0.0314	0.0312	-0.1239	-0.1840
	SD	(0.0101)	(0.0099)	(0.1108)	(0.1327)
	95% CI	(0.0110,0.0514)	(0.0118,0.0500)	(-0.3446,0.0565)	(-0.4538,0.0424)
τ	Mean	0.1028	0.1019	-	-
	SD	(0.0296)	(0.0297)	-	-
	95% CI	(0.0601,0.1752)	(0.0592,0.1710)	-	-
τ_α	Mean	-	0.0740	-	0.1112
	SD	-	(0.3476)	-	(0.3681)
	95% CI	-	(8.302e-05,0.5596)	-	(1.143e-04,0.7699)
γ	Mean	-	-	0.8453	2.1268
	SD	-	-	(2.6006)	(2.9927)
	95% CI	-	-	(-3.4355,5.9506)	(-3.0422,8.1330)
	DIC	309.1	308.8	314.3	317.8
	WAIC	293.3	293.2	297.5	298.3



(a) Curve obtained for Model 2

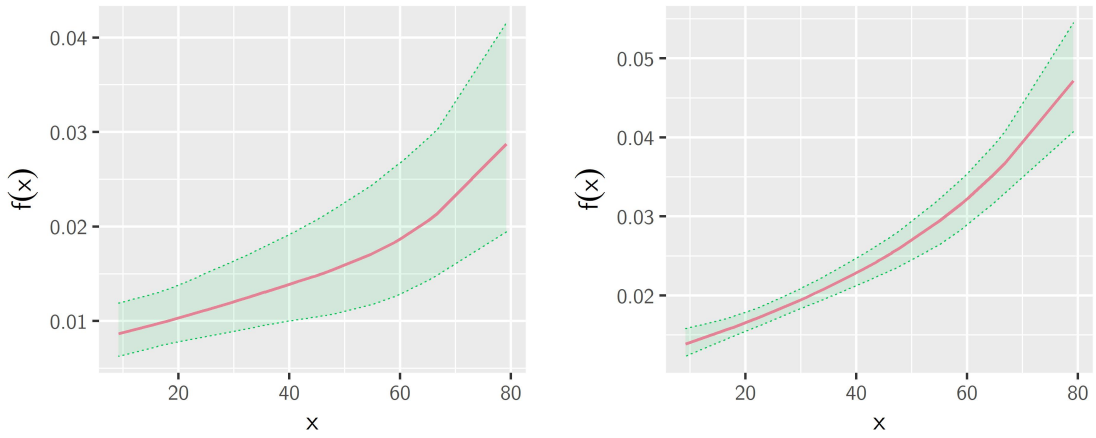
(b) Curve obtained for Model 4

Figure 3.1: Curves obtained for the semiparametric models considering the variable Viol.

For the variable IBN, with results included in Table 3.2, Model 3 has the lowest information criteria values (i.e., DIC = 312 and WAIC = 295.7). In this model, the smooth function of the variable is not included and hence, its relationship with the conditional mean is assumed to be linear. In addition, the values for Model 4, which does consider the smooth term for the variable, were DIC = 312.5 and WAIC = 296, very close to that of the best fitting model. The curves shown in Figure 3.2 might give a slight sense that the relationship between the variable and the conditional mean resembles an exponential function, especially the one obtained for Model 4 (i.e., Figure 3.2(b)), which has a narrower credible interval band. However, given that it would still be possible to fit a straight line that falls within this interval, we conclude that there is no evidence of the existence of a relationship other than the linear one here either.

Table 3.2: Results obtained from the fitting of the four models considered for the variable IBN.

		Model 1	Model 2	Model 3	Model 4
Intercept	Mean	-4.8525	-4.9358	-4.4500	-4.4510
	SD	(0.1465)	(0.2006)	(0.0582)	(0.0790)
	95% CI	(-5.1544,-4.5692)	(-5.3767,-4.5787)	(-4.5626,-4.3337)	(-4.6169,-4.2905)
IBN	Mean	0.0158	0.0174	0.0172	0.0176
	SD	(0.0019)	(0.0032)	(0.0014)	(0.0017)
	95% CI	(0.0121,0.0196)	(0.0120,0.0252)	(0.0144,0.0200)	(0.0143,0.0211)
ρ	Mean	0.0207	0.0227	0.1863	0.2007
	SD	(0.0059)	(0.0064)	(0.0584)	(0.0640)
	95% CI	(0.0093,0.0324)	(0.0101,0.0353)	(0.0775,0.3055)	(0.0745,0.3307)
τ	Mean	0.0272	0.0255	-	-
	SD	(0.0091)	(0.0089)	-	-
	95% CI	(0.0144,0.0489)	(0.0126,0.0474)	-	-
τ_α	Mean	-	0.0622	-	0.0145
	SD	-	(0.2119)	-	(0.0475)
	95% CI	-	(9.068e-05,0.4354)	-	(7.632e-05,0.1037)
γ	Mean	-	-	-8.2272	-8.6307
	SD	-	-	(1.5231)	(1.6897)
	95% CI	-	-	(-11.2720,-5.4117)	(-12.0734,-5.3016)
	DIC	312.8	315.6	312	312.5
	WAIC	295.9	296.6	295.7	296



(a) Curve obtained for Model 2 **(b)** Curve obtained for Model 4

Figure 3.2: Curves obtained for the semiparametric models considering the variable IBN.

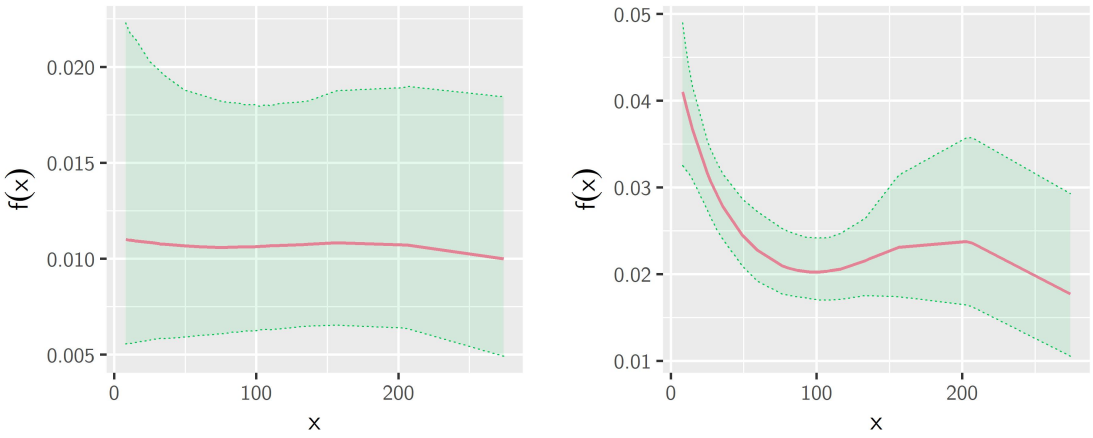
The results obtained for the variable Rec, shown in Table 3.3, indicate that the best

fitting model, according to the information criteria values, is Model 2, which includes the smooth term for this variable. In addition, Model 1 also provides a similar fit in terms of DIC and WAIC. However, the variable Rec is not statistically significant in any of these two models. Moreover, from the curve shown in Figure 3.3(a), it is not possible to infer any evidence that indicates the existence of a relationship other than the linear one between the variable and the conditional mean.

Small information criteria values, close to the ones from Model 2, were also obtained for Model 4, which is the generalized spatial conditional normal Poisson model where the smooth term is included in the regression structure for the mean and the spatial lag in the variance structure. In this case, the curve obtained from this model, shown in Figure 3.3(b), does seem to show a non linear behaviour, as it changes according to the values of the variable Rec. In particular, it suggests that for regions where the amount of resources provided by the government for academic achievement or education is less than 100 thousand dollars, when the value of this variable increases, the mortality rates decrease. For regions with values of Rec between 100 and 200 thousand dollars, the mortality rates increase when the resources increase, and for values larger than 200, the mortality rates decrease again when the resources increase. These results will be discussed in more detail in Section 3.4.3.

Table 3.3: Results obtained from the fitting of the four models considered for the variable Rec.

		Model 1	Model 2	Model 3	Model 4
Intercept	Mean	-4.5407	-4.4801	-3.4357	-3.0185
	SD	(0.3294)	(0.3711)	(0.1380)	(0.3061)
	95% CI	(-5.1827,-3.8776)	(-5.1597,-3.7142)	(-3.7589,-3.2119)	(-3.4602,-2.2699)
Rec	Mean	-1.527e-04	-5.127e-04	-0.0029	-0.0043
	SD	(0.0010)	(0.0014)	(0.0013)	(0.0020)
	95% CI	(-0.0022,0.0018)	(-0.0036,0.0018)	(-0.0051,-8.830e-05)	(-0.0091,-0.0011)
ρ	Mean	0.0328	0.0322	-0.1176	-0.2210
	SD	(0.0111)	(0.0113)	(0.0817)	(0.0736)
	95% CI	(0.0111,0.0549)	(0.0103,0.0541)	(-0.2712,0.0412)	(-0.3675,-0.0750)
τ	Mean	0.1066	0.1045	-	-
	SD	(0.0321)	(0.0303)	-	-
	95% CI	(0.0625,0.1841)	(0.0607,0.1765)	-	-
τ_α	Mean	-	0.0909	-	1.2826
	SD	-	(0.5946)	-	(4.2924)
	95% CI	-	(8.347e-05,0.7141)	-	(0.0159,8.2577)
γ	Mean	-	-	0.7308	2.6717
	SD	-	-	(1.9564)	(1.6612)
	95% CI	-	-	(-3.0569,4.4696)	(-0.5804,6.0385)
	DIC	309.1	308.9	314.2	309.6
	WAIC	293.3	293.5	296	293.1



(a) Curve obtained for Model 2

(b) Curve obtained for Model 4

Figure 3.3: Curves obtained for the semiparametric models considering the variable Rec.

Here, it could be useful to mention that, for Figure 3.3(b), the credible interval band becomes considerably wider from the value 150 onwards. Therefore, we should examine the variable Rec in more detail, for what we can rely on the information offered by the boxplot for this variable, shown in Figure 3.4. We can see that there might be less information available for this variable, given that the third quartile of the distribution is located between 100 and 150. Moreover, we can observe that there is an outlier given by the point located beyond the value of 250, which corresponds to the region of Antioquía. Let us recall that the presence of this outlier was already discussed in Section 2.6 in Chapter 2. Due to this fact, we should be particularly careful when performing inference regarding this last part of the curve, since it might not be as reliable as the one before the value of 150.

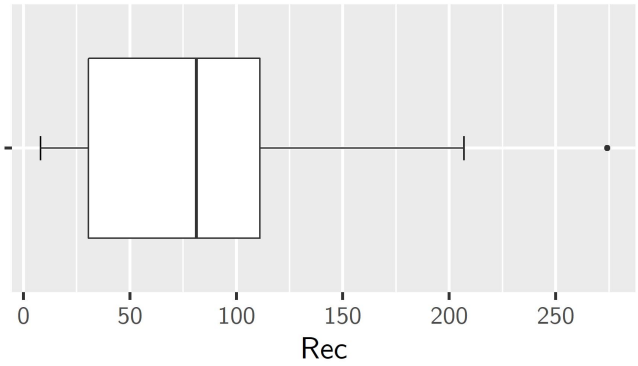


Figure 3.4: Boxplot of the variable Rec.

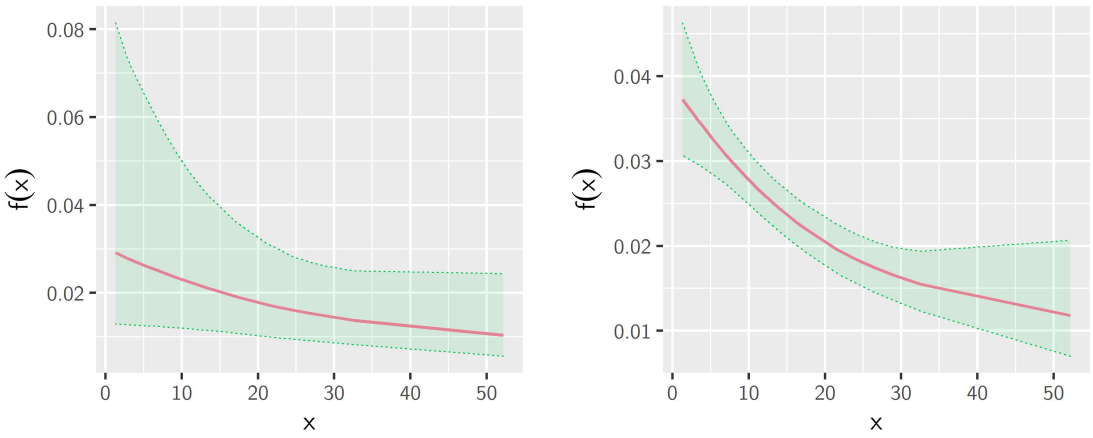
For the variable HE, results are reported in Table 3.4. We can see that the best

fitting model is Model 2, the spatial conditional normal Poisson model which includes the smooth term for the variable. Here, the fixed effect for the variable is statistically significant but the spatial lag is not. In any case, the curves in Figure 3.5 do not suggest any relationship other than a linear one.

Finally, for the last variable under study, which is Vac, results are reported in Table 3.5. Here, we can see that the generalized spatial conditional normal Poisson models (i.e., Models 3 and 4) have the largest DIC and WAIC values among the four models. In addition, in Models 1 and 2, the fixed effect for Vac is not significant. In any case, the curves shown in Figure 3.6 do not indicate any evidence of the existence of non linear relationships between the variable and the predictor.

Table 3.4: Results obtained from the fitting of the four models considered for the variable HE.

		Model 1	Model 2	Model 3	Model 4
Intercept	Mean	-3.7109	-3.5142	-3.3161	-3.3017
	SD	(0.3400)	(0.4596)	(0.0848)	(0.1222)
	95% CI	(-4.3788,-3.0498)	(-4.3434,-2.5215)	(-3.4912,-3.1496)	(-3.5397,-3.0560)
HE	Mean	-0.0202	-0.0202	-0.0263	-0.0225
	SD	(0.0061)	(0.0072)	(0.0048)	(0.0062)
	95% CI	(-0.0324,-0.0085)	(-0.0342,-0.0060)	(-0.0359,-0.0166)	(-0.0345,-0.0102)
ρ	Mean	0.0119	0.0061	-0.0749	-0.0779
	SD	(0.0109)	(0.0133)	(0.0536)	(0.0504)
	95% CI	(-0.0092,0.0335)	(-0.0217,0.0310)	(-0.1829,0.0279)	(-0.1828,0.0157)
τ	Mean	0.0736	0.0712	-	-
	SD	(0.0216)	(0.0209)	-	-
	95% CI	(0.0424,0.1241)	(0.0402,0.1222)	-	-
τ_α	Mean	-	0.1153	-	0.0992
	SD	-	(0.5689)	-	(0.5977)
	95% CI	-	(9.274e-05,0.8647)	-	(1.416e-04,0.6843)
γ	Mean	-	-	-0.9012	-0.9304
	SD	-	-	(1.2974)	(1.2079)
	95% CI	-	-	(-3.3154,1.8178)	(-3.0842,1.6519)
	DIC	310.2	308.3	313.6	312.4
	WAIC	293.2	293.5	295.6	295.4

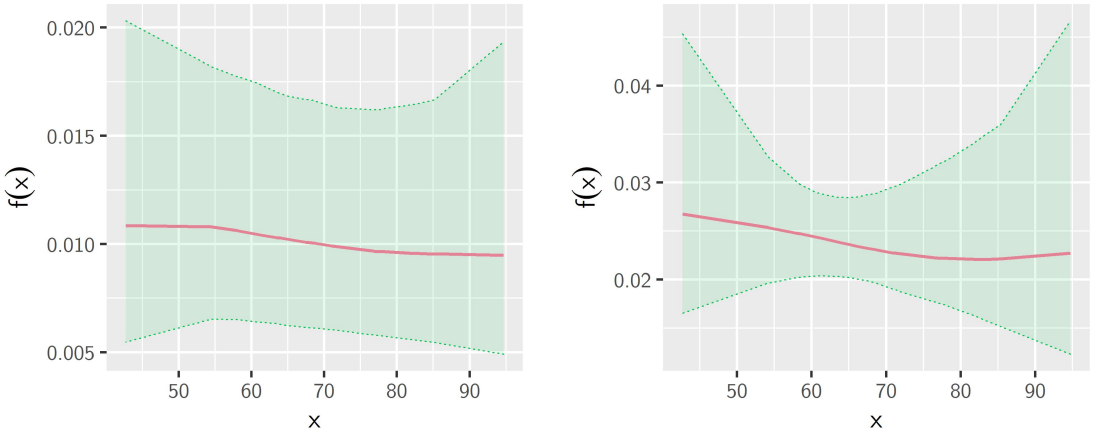


(a) Curve obtained for Model 2 **(b)** Curve obtained for Model 4

Figure 3.5: Curves obtained for the semiparametric models considering the variable HE.

Table 3.5: Results obtained from the fitting of the four models considered for the variable Vac.

		Model 1	Model 2	Model 3	Model 4
Intercept	Mean	-4.3322	-4.4556	-3.4822	-3.5351
	SD	(0.4216)	(0.5778)	(0.5890)	(0.6678)
	95% CI	(-5.1826,-3.5302)	(-5.7295,-3.4455)	(-4.6033,-2.3583)	(-4.8126,-2.2538)
Vac	Mean	-0.0038	-0.0020	-0.0039	-0.0025
	SD	(0.0055)	(0.0075)	(0.0092)	(0.0103)
	95% CI	(-0.0145,0.0073)	(-0.0147,0.0141)	(-0.0212,0.0140)	(-0.0219,0.0172)
ρ	Mean	0.0344	0.0345	0.0189	0.0151
	SD	(0.0098)	(0.0100)	(0.0737)	(0.0747)
	95% CI	(0.0154,0.0535)	(0.0145,0.0544)	(-0.1752,0.1244)	(-0.1865,0.1266)
τ	Mean	0.1031	0.1029	-	-
	SD	(0.0306)	(0.0301)	-	-
	95% CI	(0.0601,0.1759)	(0.0593,0.1748)	-	-
τ_α	Mean	-	0.0865	-	0.0696
	SD	-	(0.5922)	-	(0.2958)
	95% CI	-	(8.740e-05,0.6770)	-	(8.653e-05,0.5498)
γ	Mean	-	-	-2.4486	-2.3575
	SD	-	-	(1.8513)	(1.8602)
	95% CI	-	-	(-5.0482,2.4843)	(-5.0600,2.6644)
	DIC	309	309.5	313.6	314.7
	WAIC	293.8	293.8	296	296



(a) Curve obtained for Model 2 **(b)** Curve obtained for Model 4

Figure 3.6: Curves obtained for the semiparametric models considering the variable Vac.

3.4.2 Comparison with the best fitting model

Let us recall that, in Section 2.6.1, Chapter 2, the best fitting model obtained for the mortality rates in Colombia was the spatial conditional normal Poisson model which included the variables IBN, Rec, as well as the spatial lag of the rates. Since, for one of the models fitted in the previous section, we found evidence of a non linear relationship between the variable Rec and the mortality rates, we believe it may be of interest to investigate whether this non linear relationship is also present in the aforementioned model from Chapter 2. In particular, for this model, the regression structure for the mean was specified so that:

$$\begin{aligned} \log(\mu_i) = &\log(\text{NB}_i) + \beta_0 + \beta_1 \text{Rec}_i + \beta_2 \text{IBN}_i + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \\ &\nu_i \sim N(0, \tau), \ \tau > 0 \end{aligned} \tag{3.18}$$

In order to evaluate if a non linear relationship between the predictor and the variable Rec is present in this model, we consider the inclusion of a smooth term for this variable, given by the P-splines mixed model representation of the smooth function of Rec (i.e., $f(\text{Rec}_i) = \beta_0 + \beta_1 \text{Rec}_i + \mathbf{Z}_i \boldsymbol{\alpha}$), so that the regression structure is as follows:

$$\begin{aligned} \log(\mu_i) = &\log(\text{NB}_i) + \beta_0 + \beta_1 \text{Rec}_i + \mathbf{Z}_i \boldsymbol{\alpha} + \beta_2 \text{IBN}_i + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \\ &\nu_i \sim N(0, \tau), \ \tau > 0, \ \alpha_k \sim N(0, \tau_\alpha), \ \tau_\alpha > 0 \end{aligned} \tag{3.19}$$

Note that, in this chapter, we have used the software package JAGS for model fitting, whereas in Section 2.6.1, Chapter 2, the models were fitted in OpenBUGS. As a direct comparison among models fitted in different software packages cannot be performed (see Morales-Otero and Núñez-Antón, 2021), here we will employ JAGS to fit the two considered models, so that we are able to compare them.

The results obtained from the fitting of the models in equations (3.18) and (3.19) are included in Table 3.6. Here, the model where the smooth term for the variable Rec was included (i.e., model in equation (3.19)) has slightly smaller DIC and WAIC values than those obtained for the model that considers the variable Rec only as a fixed effect (i.e., model in equation (3.18)). However, for the model including the smooth term for Rec, the fixed effect for this variable and the spatial term are not statistically significant. Moreover, from the curve shown in Figure 3.7, we can see that, in this model, no evidence of any non linear relationship between Rec and the mortality rates can be inferred. Therefore, in this specific case, we conclude that the variable Rec is properly specified as a fixed effect.

Table 3.6: Results obtained from the fitting of the models in equations (3.18) and (3.19) to the infant mortality rates in Colombia.

		Model in equation (3.18)	Model in equation (3.19)
Intercept	Mean	-4.6528	-4.5839
	SD	(0.1636)	(0.1911)
	95% CI	(-4.9763,-4.3374)	(-4.9552,-4.1972)
Rec	Mean	-0.0011	-0.0011
	SD	(5.110e-04)	(5.871e-04)
	95% CI	(-0.0021,-9.039e-05)	(-0.0023,8.299e-05)
IBN	Mean	0.0166	0.0168
	SD	(0.0018)	(0.0018)
	95% CI	(0.0130,0.0203)	(0.0131,0.0203)
ρ	Mean	0.0154	0.0132
	SD	(0.0060)	(0.0067)
	95% CI	(0.0036,0.0271)	(-2.390e-04,0.0265)
τ	Mean	0.0234	0.0230
	SD	(0.0081)	(0.0083)
	95% CI	(0.0120,0.0428)	(0.0114,0.0434)
τ_α	Mean	-	0.0150
	SD	-	(0.0507)
	95% CI	-	(7.974e-05,0.1031)
	DIC	313.4	311.5
	WAIC	295.8	295.7

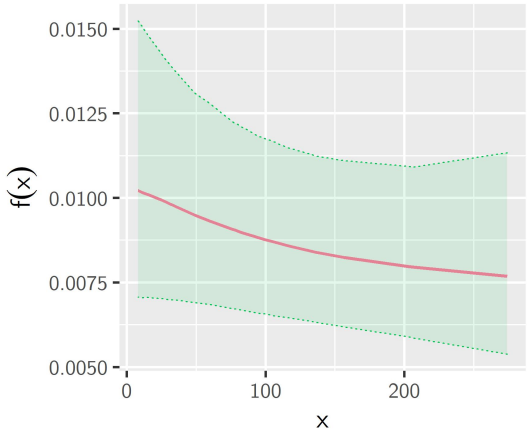


Figure 3.7: Curve obtained for the semiparametric model in equation (3.19).

3.4.3 Summary of the results obtained

We have investigated the presence of non linear relationships between the conditional mean of the mortality rates and the variables included in the study, and in one of the models fitted, we have found evidence of such relation for the variable Rec. This variable represents the amount of resources provided by the government for academic achievement or education per household, in thousands of dollars.

In this case, a smooth relationship with the predictor was observed, where the interpretation was not the same for three different intervals of the values of the variable Rec. In particular, the values of the mortality rates increased for larger values of Rec when the resources were between 100 and 200 thousand dollars. For the rest of the values, that is for less than 100 and more than 200 thousand dollars, the relationship was a decreasing one.

Therefore, if the variable Rec was included in a model as a fixed effect, it is possible that not all of its variability would be captured, so it is be advisable that a smoothing term for this variable would be considered. For the rest of the variables, since no evidence that indicates otherwise was found, if they were to be included in the regression structure, they could be specified as fixed effects.

In particular, the generalized spatial conditional normal Poisson model (i.e., Model 4) that was fitted, considering the smooth function for the variable Rec, was given by:

$$\begin{aligned} \log(\mu_i) =& \log(\text{NB}_i) + \beta_0 + \beta_1 \text{Rec}_i + \mathbf{Z}_i\boldsymbol{\alpha} + \nu_i, \\ & \nu_i \sim N(0, \tau_i),\ \tau_i > 0,\ \alpha_k \sim N(0, \tau_\alpha),\ \tau_\alpha > 0 \\ \log(\tau_i) =& \gamma + \rho \mathbf{W}_i \mathbf{Rates}, \end{aligned} \tag{3.20}$$

where Rec_i is the variable corresponding to the i-th area and the rest of the terms are as before. Note that the results obtained from the fitting of this model have been included in Table 3.3.

We can now examine some of the posterior predictive checks for this model. For example, the scatterplot of the observed versus the predicted rates is shown in Figure 3.8, where we can see that the predictions are quite accurate. In addition, the map of the resulting predictions is included in Figure 3.9(b). If we compare this map with the map of the observed rates in Figure 3.9(a), we can see that the predictions are almost identical to their corresponding observed values.

Finally, as was the case for the model in equation (3.20), evidence of a non linear relationship between the mortality rates and the variable Rec was found. We have further investigated whether there was also evidence of such non linear relationship for the best fitting model from Section 2.6.1, in the previous chapter (i.e., Chapter 2), since this model also included the variable Rec. The conclusion was that no evidence of such relationship was found. Hence, the evidence suggests that this model was properly specified by including the variable Rec as a fixed effect.

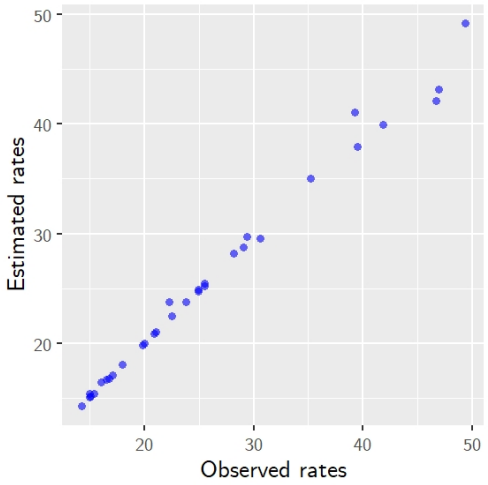
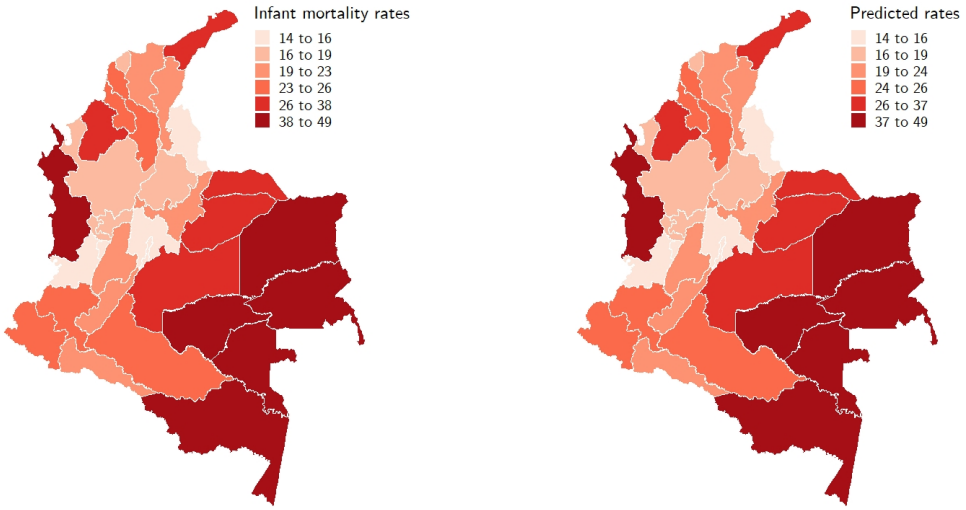


Figure 3.8: Scatterplot of the observed versus the predicted rates obtained after fitting the model in equation (3.20).



(a) Map of the observed mortality rates.

(b) Map of the predicted mortality rates obtained after fitting the model in equation (3.20).

Figure 3.9: Maps of the observed and the predicted mortality rates.

3.5 Application to mother's postnatal period screening test in Colombia

In this section, we will fit the semiparametric models proposed in Section 3.3, for binomially distributed responses, to the data set that was analysed in Section 2.7, Chapter 2, concerning the mother's postnatal period screening test in Colombia. The variables available were the number of women who gave birth to their last child between the years 1999 and 2005 and went through a postnatal period screening test (i.e., variable Nscreen), the number of women who has their last child after 1999 (i.e., variable NMothers), the percentage of women who had to pay for their postnatal check-up (i.e., variable Pay), the percentage of women over 18 years old who declared to have suffered physical violence from their current partners (i.e., variable Viol), the percentage of young people (between 18 and 24 years) who had access to a higher educational level (i.e., variable HE) and the percentage of the population that had basic services not being satisfactorily attended to for the year 2005 (i.e., variable IBN). See Section 2.7 for a more detailed description of this data set.

Here, in order to investigate the possibility of the existence of a non linear behaviour, the relationship of each covariate with the logistic function of the probability of success will be assessed. Therefore, we will fit multiple models such as the ones proposed in Section 3.3, for binomially distributed count variables. As in the previous Section, here we will show four model alternatives that we believe are the ones that best illustrate the

analysis performed.

The first model considered (i.e., Model 1) is a spatial conditional normal binomial model including the explanatory variable v_i, for $i = 1, \ldots, n$, where we assume that $\text{Nscreen}_i \sim \text{Bin}(\text{NMothers}_i, \pi_i)$, with the probability of success, π_i, following the regression structure:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 v_i + \rho A_i + \nu_i,\ \nu_i \sim N(0, \tau),\ \tau > 0 \tag{3.21}$$

Note that here β_0 and β_1 are the intercept and the unknown coefficient, respectively, A_i is the spatial term for the i-th region and ρ is the spatial parameter.

The second model (i.e., Model 2) is also a spatial conditional normal binomial model as Model 1, but where we consider the relationship between the covariate and the predictor to be given by the smooth function $f(\cdot)$. For this function, we will specify the mixed model representation from equation (3.7) (i.e., $f(v_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha}$), so that the regression structure for π_i is:

$$\begin{aligned} \text{logit}(\pi_i) =& \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha} + \rho A_i + \nu_i, \\ & \nu_i \sim N(0, \tau),\ \tau > 0,\ \alpha_k \sim N(0, \tau_\alpha),\ \tau_\alpha > 0 \end{aligned} \tag{3.22}$$

The third model (i.e., Model 3) is a generalized spatial conditional normal binomial model, where we include the covariate in the regression structure for the probability of success and the spatial term in the structure for the variance parameter. In particular, the model for π_i is the following:

$$\begin{aligned} \text{logit}(\pi_i) =& \beta_0 + \beta_1 v_i + \nu_i, \quad \nu_i \sim N(0, \tau_i),\ \tau_i > 0 \\ \log(\tau_i) =& \gamma + \rho \text{A}_i \end{aligned} \tag{3.23}$$

Finally, the last model considered (i.e., Model 4) is the same generalized spatial conditional normal binomial model as the one before, but including the smooth term for the variable v_i (i.e., $f(v_i) = \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha}$) in the regression structure for π_i, so that:

$$\begin{aligned} \text{logit}(\pi_i) =& \beta_0 + \beta_1 v_i + \mathbf{Z}_i\boldsymbol{\alpha} + \nu_i, \\ & \nu_i \sim N(0, \tau_i),\ \tau_i > 0,\ \alpha_k \sim N(0, \tau_\alpha),\ \tau_\alpha > 0, \\ \log(\tau_i) =& \gamma + \rho \text{A}_i \end{aligned} \tag{3.24}$$

As in the previous Section, here we will also specify $K = 5$ knots, so $\mathbf{Z}$ is a 32×6 matrix with $k = 1, \ldots, 6$, random coefficients α_k. Here, the models will also be implemented in JAGS and noninformative priors will be assumed for the parameters. That is, normal priors for the fixed effects (i.e., $N(0, 1\text{e-}05)$) and gamma priors for the precision parameters of the random effects (i.e., $G(1\text{e-}04, 1\text{e-}04)$). In addition, models will be compared with the use of the Deviance and the Watanabe-Akaike information criteria (i.e., DIC and WAIC).

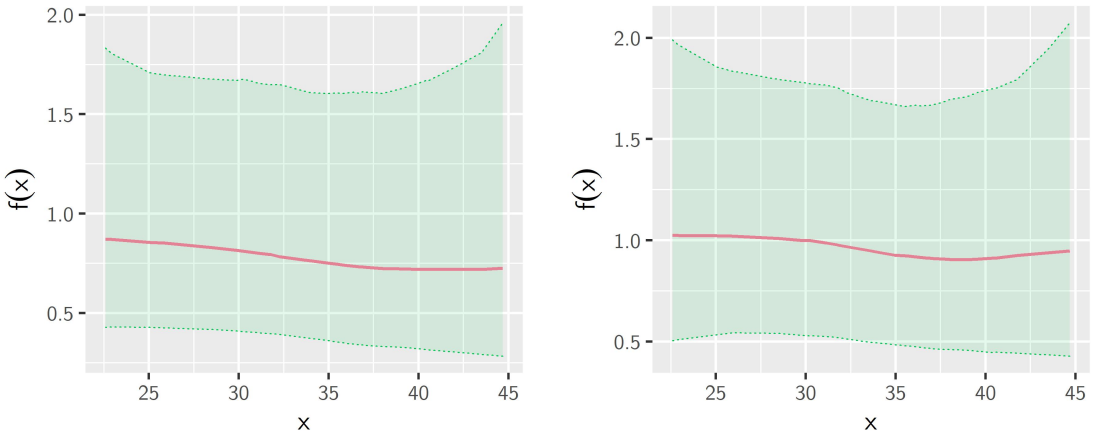
3.5.1 Fitting of the semiparametric generalized spatial conditional over-dispersion models for binomial responses

The results obtained after fitting the four models described above are included in Tables 3.7 to 3.10, when considering each of the variables Viol, IBN, HE and Pay, respectively. In addition, the curves obtained for the smooth functions from Models 2 and 4 are also shown in Figures 3.10 to 3.13, for each of the aforementioned variables.

For the first variable under study, that is Viol, we can see in Table 3.7 that the best fitting models, according to their DIC and WAIC values, are Models 1 and 2. However, in these models the fixed effect for the variable is not significant and neither is the spatial term. Moreover, the curves obtained for the smooth function in Models 2 and 4, shown in Figure 3.10, do not seem to indicate any evidence of the existence of a relationship different than the linear one between the variable and the predictor.

Table 3.7: Results obtained from the fitting of the four models considered for the variable Viol.

		Model 1	Model 2	Model 3	Model 4
Intercept	Mean	0.1070	0.0200	0.3286	0.0715
	SD	(0.5773)	(0.6809)	(0.5615)	(0.7442)
	95% CI	(-1.0238,1.2368)	(-1.3759,1.3331)	(-0.7636,1.4554)	(-1.4692,1.4377)
Viol	Mean	-0.0105	-0.0075	7.414e-04	-0.0028
	SD	(0.0170)	(0.0206)	(0.0160)	(0.0216)
	95% CI	(-0.0447,0.0225)	(-0.0452,0.0360)	(-0.0310,0.0324)	(-0.0424,0.0405)
ρ	Mean	0.4219	0.4273	0.5631	-0.2870
	SD	(0.2311)	(0.2345)	(0.9956)	(0.1949)
	95% CI	(-0.0139,0.8770)	(-0.0354,0.8961)	(-1.3211,2.5493)	(-0.6615,0.0715)
τ	Mean	0.1826	0.1841	-	-
	SD	(0.0699)	(0.0734)	-	-
	95% CI	(0.0793,0.3456)	(0.0800,0.3651)	-	-
τ_α	Mean	-	0.1213	-	0.1841
	SD	-	(0.6073)	-	(0.9195)
	95% CI	-	(8.830e-05,0.9536)	-	(7.573e-05,1.6029)
γ	Mean	-	-	-2.5494	-1.3116
	SD	-	-	(1.5723)	(0.4962)
	95% CI	-	-	(-5.7918,0.3431)	(-2.2551,-0.3020)
	DIC	224.5	224.4	226.2	226.4
	WAIC	204.7	204.3	205.6	205.8



(a) Curve obtained for Model 2

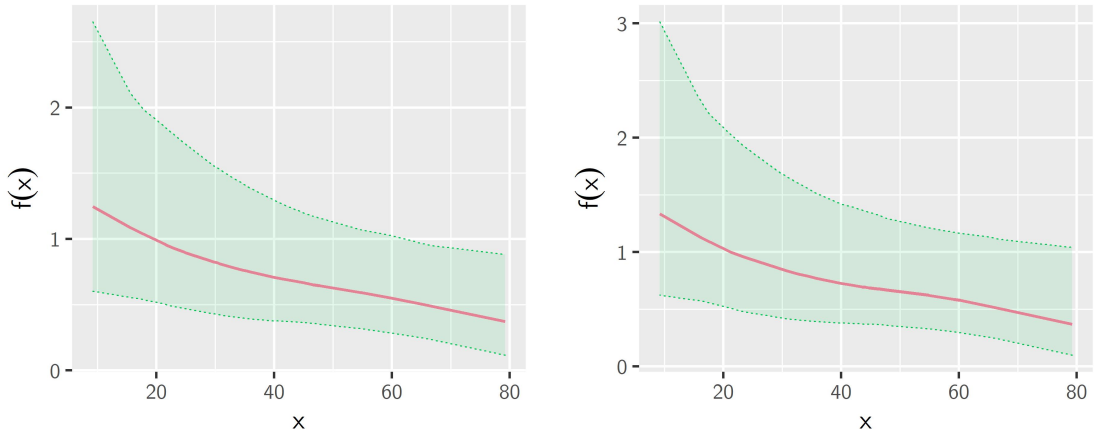
(b) Curve obtained for Model 4

Figure 3.10: Curves obtained for the semiparametric models considering the variable Viol.

For the variable IBN, the results obtained from the fitting of the four models are shown in Table 3.8. Here, the best fit corresponds to Model 3, which is the generalized spatial conditional normal binomial model that does not consider the smooth term for the variable, but that includes it as a fixed effect in the regression structure for the mean, resulting statistically significant. In addition, the spatial term is included in the regression structure for the precision parameter, resulting not statistically significant. In any case, the curves obtained for the models including the smooth term, shown in Figure 3.11, do not suggest any type of non linear relationship between the variable and the predictor.

Table 3.8: Results obtained from the fitting of the four models considered for the variable IBN.

		Model 1	Model 2	Model 3	Model 4
Intercept	Mean	0.3345	0.4201	0.9873	0.5184
	SD	(0.3251)	(0.4007)	(0.2122)	(0.4296)
	95% CI	(-0.3183,0.9692)	(-0.3328,1.2778)	(0.5869,1.4134)	(-0.2668,1.4218)
IBN	Mean	-0.0156	-0.0187	-0.0166	-0.0204
	SD	(0.0046)	(0.0078)	(0.0051)	(0.0091)
	95% CI	(-0.0245,-0.0065)	(-0.0391,-0.0069)	(-0.0268,-0.0069)	(-0.0425,-0.0069)
ρ	Mean	0.4481	0.4754	1.8190	-0.4489
	SD	(0.1821)	(0.1989)	(1.5707)	(0.2180)
	95% CI	(0.0885,0.7998)	(0.0893,0.8639)	(-1.3703,4.9559)	(-0.8433,-0.0186)
τ	Mean	0.1157	0.1106	-	-
	SD	(0.0471)	(0.0492)	-	-
	95% CI	(0.0481,0.2257)	(0.0436,0.2334)	-	-
τ_α	Mean	-	0.2267	-	0.3523
	SD	-	(1.0705)	-	(1.4379)
	95% CI	-	(9.952e-05,1.8031)	-	(1.086e-04,2.6015)
γ	Mean	-	-	-4.9478	-1.6115
	SD	-	-	(2.5979)	(0.5245)
	95% CI	-	-	(-10.4077,0.1639)	(-2.6031,-0.5811)
	DIC	217.7	218.8	215.8	221.9
	WAIC	200.8	201	200.6	201.7



(a) Curve obtained for Model 2 **(b)** Curve obtained for Model 4

Figure 3.11: Curves obtained for the semiparametric models considering the variable IBN.

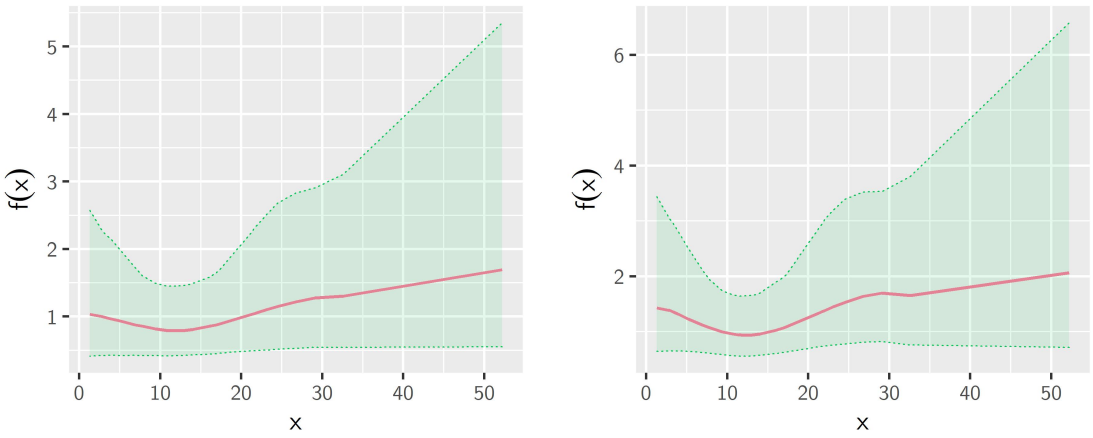
Table 3.9 includes the results obtained after fitting the models considering the variable HE. Here, the fixed effect is not significant in any of the four cases and neither is

the spatial term. Moreover, the curves shown in Figure 3.12 do not suggest any clear non linear relationship between the predictor and HE.

Finally, the results obtained when considering the variable Pay can be seen in Table 3.10. The fixed effect is significant in all the models. However, the spatial term is not. The best fitting model according to the information criteria values is Model 2, the spatial conditional normal binomial model that includes the smooth term for variable Pay. Nevertheless, in Figure 3.13 we can see that the curves for models 2 and 4 do not suggest any non linear relationship between Pay and the predictor.

Table 3.9: Results obtained from the fitting of the four models considered for the variable HE.

		Model 1	Model 2	Model 3	Model 4
Intercept	Mean	-0.3860	-0.1535	0.2011	0.0462
	SD	(0.3735)	(0.7246)	(0.1754)	(0.8684)
	95% CI	(-1.1517,0.3537)	(-1.9610,1.1123)	(-0.1464,0.5389)	(-2.2905,1.4422)
HE	Mean	0.0111	0.0141	0.0103	0.0156
	SD	(0.0088)	(0.0252)	(0.0092)	(0.0313)
	95% CI	(-0.0061,0.0287)	(-0.0204,0.0817)	(-0.0075,0.0283)	(-0.0255,0.1025)
ρ	Mean	0.4011	0.2874	0.6764	-0.1249
	SD	(0.2155)	(0.2184)	(1.0607)	(0.1864)
	95% CI	(-0.0167,0.8323)	(-0.1125,0.7293)	(-1.2625,2.8473)	(-0.4654,0.2799)
τ	Mean	0.1784	0.1300	-	-
	SD	(0.0704)	(0.0710)	-	-
	95% CI	(0.0777,0.3463)	(0.0316,0.2983)	-	-
τ_α	Mean	-	7.2366	-	12.2909
	SD	-	(39.0385)	-	(45.1545)
	95% CI	-	(2.083e-04,64.3950)	-	(5.376e-04,95.8509)
γ	Mean	-	-	-2.7245	-2.1909
	SD	-	-	(1.6651)	(0.8500)
	95% CI	-	-	(-6.2064,0.3112)	(-4.0603,-0.8685)
	DIC	222.2	227.5	222.9	239.5
	WAIC	203.2	206.5	203.8	216.1

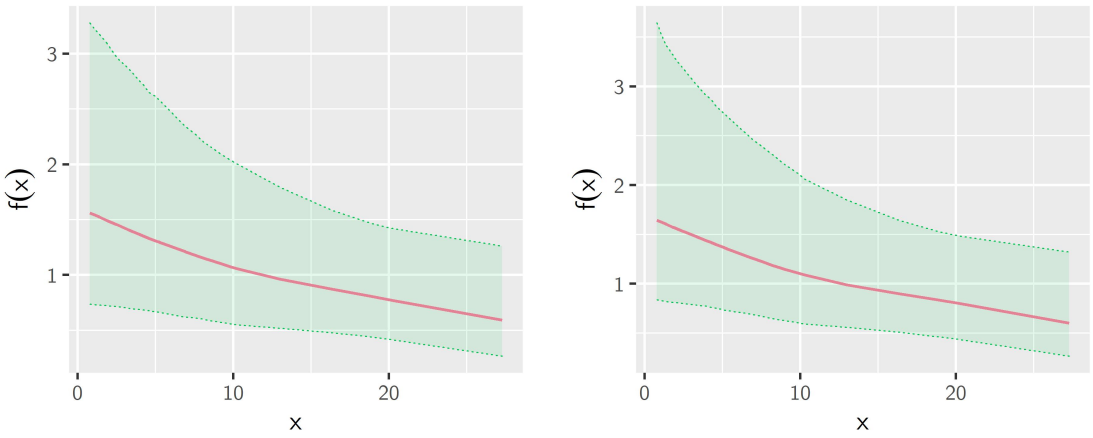


(a) Curve obtained for Model 2 **(b)** Curve obtained for Model 4

Figure 3.12: Curves obtained for the semiparametric models considering the variable HE.

Table 3.10: Results obtained from the fitting of the four models considered for the variable Pay.

		Model 1	Model 2	Model 3	Model 4
Intercept	Mean	0.4629	0.4696	0.6340	0.5327
	SD	(0.3717)	(0.3922)	(0.1051)	(0.3768)
	95% CI	(-0.2714,1.1958)	(-0.2935,1.2316)	(0.4247,0.8399)	(-0.1945,1.3183)
Pay	Mean	-0.0369	-0.0372	-0.0396	-0.0389
	SD	(0.0136)	(0.0157)	(0.0105)	(0.0167)
	95% CI	(-0.0637,-0.0104)	(-0.0686,-0.0068)	(-0.0600,-0.0192)	(-0.0730,-0.0064)
ρ	Mean	0.1100	0.1107	1.3585	-0.0777
	SD	(0.2055)	(0.2104)	(1.6759)	(0.1935)
	95% CI	(-0.2914,0.5199)	(-0.3047,0.5399)	(-1.7287,5.4112)	(-0.4568,0.3677)
τ	Mean	0.1191	0.1216	-	-
	SD	(0.0550)	(0.0571)	-	-
	95% CI	(0.0424,0.2489)	(0.0410,0.2592)	-	-
τ_α	Mean	-	0.1690	-	0.2542
	SD	-	(1.6239)	-	(1.9893)
	95% CI	-	(8.493e-05,1.2968)	-	(8.744e-05,1.7158)
γ	Mean	-	-	-4.6883	-2.0991
	SD	-	-	(2.9807)	(0.5615)
	95% CI	-	-	(-12.4338,0.3167)	(-3.2226,-0.9937)
	DIC	231.2	230.9	248.3	232.5
	WAIC	207.7	207.2	213.7	208.3



(a) Curve obtained for Model 2 **(b)** Curve obtained for Model 4

Figure 3.13: Curves obtained for the semiparametric models considering the variable Pay.

3.5.2 Summary of the results obtained

In this section, we have assessed the behaviour of the relationships between each of the covariates available in the data set under study and the probability of success of women going through a postnatal period screening test. In order to do this, we have included a smooth function of each variable in the models, which was specified by means of a P-splines mixed model representation. In this specific application, we have not found any evidence of any relationship different than the linear one between the logistic model for the probability of success and the independent variables under study. Therefore, we have no reasons to believe that if any of these variables was included in a study, their effect would be inappropriately captured by fixed linear coefficients.

3.6 Discussion

In this chapter, we have proposed an extension of the generalized spatial conditional overdispersion models in Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018 that allows to capture possible existing non linear relationships between the variables under study and the predictor. For the smoothing of such variables, we have considered P-splines in their mixed models representation proposed by Currie and Durbán (2002), which simplifies the fitting of the models and the inferential process in great manner, when compared to the nonparametric methodology. In particular, we have proposed the semiparametric generalized spatial conditional normal Poisson and binomial models, where the smoothing term is included in the regression structure for the conditional mean.

These model proposals have been illustrated with two applications to real data examples, where we have investigated the type of relationship obtained when including

the smoothing term in the models. In addition, we have assessed their behaviour by performing their comparison with other alternatives excluding the smoothing term.

We have fitted the models for Poisson responses to the infant mortality rates in Colombia data set, finding evidence of the existence of a non linear relationship between the mortality rates and the variable Rec, for one of the considered models. The effect of this variable on the mortality rates showed a behaviour that decreased, increased and decreased again according to the value of the variable, hence, clearly showing a non linear effect. For the remaining variables considered, there was no indication of any evidence of a non linear relationship.

Additionally, the proposed semiparametric models for binomially distributed responses were fitted to the data concerning the mother's postnatal period screening test in Colombia. In this case, we did not find any evidence of the existence of non linear relationships between the variables under study and the probability of success.

The applications presented here can be considered as good illustrating examples for the proposed semiparametric models. However, given the limited sample size of the data sets under analysis (i.e., $N = 32$ observations), we specified a number of 5 knots for the definition of the smooth functions of the variables, taking into account that this quantity is usually between 5 and 40 (Ruppert, 2002). Therefore, we did not consider it necessary to perform any additional comparison to settings with more than 5 knots. In this sense, perhaps other data sets could be explored, so that smoothing with a larger number of knots can be studied.

Nevertheless, in these applications, we believe that we have been able to illustrate how non linear relationships between variables and the predictor can be detected so that their effect can be properly expressed in the regression models. Moreover, we believe that the proposed extension of the generalized spatial conditional models offers a great advantage because of the possibility of including semiparametric terms that allow for capturing these possible non linear relationships, adding even more value to these models.

Chapter 4

Spatio-temporal extensions of the spatial conditional overdispersion models

4.1 Introduction

Spatio-temporal data arise in many fields of study, since researchers are often interested in studying a phenomenon observed in several locations and time periods, characterizing its behaviour and perhaps, forecasting it. An example of one of these fields is given by disease mapping (Lawson, 2008; Blangiardo and Cameletti, 2015). This type of data often exhibit correlation among regions and time units (Cressie and Wikle, 2011). Therefore, both issues need to be taken into account when fitting regression models to these data. It is also necessary to investigate the interaction between the spatial and the temporal dimensions and, in addition, to study the dynamic relationship of the spatial processes in time.

In this chapter, we review some of the most frequently used models in the spatio-temporal context. In addition, we propose some extensions of the spatial conditional overdispersion models in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) to allow for the modelling of spatio-temporal data. In Section 4.2, we describe some models often found in the literature, such as the ones proposed by Bernardinelli et al. (1995) and Knorr-Held (2000). The proposed spatio-temporal conditional models and the temporally varying spatial lag coefficient models are presented in Sections 4.3 and 4.3.1, respectively. We assess the performance of these proposals by fitting them to the respiratory hospital admissions in Glasgow and to the low birth weight in Georgia data sets, with their corresponding results included in Sections 4.4 and 4.5, respectively. In these two sections we also include a comparison with the Knorr-Held (2000) model proposals. This chapter ends in Section 4.6 with a discussion of the results for the different models fitted here.

4.2 Spatio-temporal models for count data

Let us assume that the random variables Y_{ij}, for $i = 1, \ldots, n$, and $j = 1, \ldots, J$, represent counts for n regions in J time periods each. In addition, we will also assume that these variables follow a count data distribution, denoted by $f(y_{ij})$ with means μ_{ij}; that is, $\mathrm{E}(Y_{ij}) = \mu_{ij}$, where $f(\cdot)$ could be a Poisson, a binomial, or any other count data distribution. In generalized linear models, a regression structure for the means μ_{ij} is commonly specified as:

$$g(\mu_{ij}) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta}, \tag{4.1}$$

where $g(\cdot)$ is a monotonic and differentiable link function, $\mathbf{x}_{ij}$ is the $k \times 1$ vector of explanatory variables for the i-th area in the j-th time period, and $\boldsymbol{\beta}$ is the $k \times 1$ vector of unknown regression parameters that need to be estimated. The function $g(\cdot)$ is usually assumed to be the logarithmic or the logistic function in the presence of Poisson or binomial distributed response variables, respectively. Note that this specific mean structure for μ_{ij} is commonly known as the linear predictor.

There are several models that have been proposed in the literature to account for spatio-temporal autocorrelation. For example, one model frequently applied is the one proposed by Bernardinelli et al. (1995), which includes a parametric linear time trend in the linear predictor. In particular, under this specification, the means μ_{ij} are modelled as:

$$g(\mu_{ij}) = \beta + \nu_i + \gamma_0 \mathrm{time}_{ij} + \gamma_i \mathrm{time}_{ij}, \tag{4.2}$$

where $g(\cdot)$ is a link function, β is an unknown intercept term to be estimated and ν_i is a normally distributed random effect, that is $\nu_i \sim N(0, \tau_\nu)$, with $\tau_\nu > 0$, for $i = 1, \ldots, n$. In addition, time is a variable representing the time unit (i.e., day, month, year, or any other) and γ_0 is an unknown coefficient, so that the term $\gamma_0 \mathrm{time}_{ij}$ represents a global linear temporal trend. Note that γ_0 is the fixed coefficient that quantifies the slope of the line that describes the evolution of the time series between two time periods. Furthermore, γ_i is an interaction random effect between area and time that would explain the difference between the global and the local time trend for each region (Blangiardo and Cameletti, 2015). For this effect, it is usually assumed that $\gamma_i \sim N(0, \tau_\gamma)$, with $\tau_\gamma > 0$, for $i = 1, \ldots, n$.

It could be useful to mention that Bernardinelli et al. (1995) also described the possibility of specifying a spatially structured prior distribution for the effects γ_i, which can depend on a specific neighbourhood structure. In addition, although the authors did not consider the inclusion of explanatory variables in their original formulation, it would be straightforward to consider this model extension by including them in the linear predictor in equation (4.2).

Another well known, and perhaps the most commonly applied model to spatio-temporal count data, was proposed by Knorr-Held (2000), who considered a dynamic nonparametric formulation. Here, a series of unstructured and structured random effects is included in the linear predictor, so that the means μ_{ij} follow the regression structure:

$$g(\mu_{ij}) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + \nu_i + \eta_i + \delta_j + \phi_j, \tag{4.3}$$

where ν_i and δ_j are unstructured random effects for space and time, respectively, and η_i and ϕ_j are spatially and temporal structured random effects, respectively. In addition, $g(\cdot)$, $\mathbf{x}_{ij}$ and $\boldsymbol{\beta}$ are as before.

In general, a Besag-York-Mollie (BYM) (Besag, York and Mollié, 1991) specification is assumed for the spatial random effects, with ν_i a normally distributed random effect, that is $\nu_i \sim N(0, \tau_\nu)$, with $\tau_\nu > 0$, for $i = 1, \ldots, n$, and η_i following an intrinsic conditionally autoregressive prior distribution (ICAR), so that:

$$(\eta_i | \eta_{\sim i}, \mathbf{W}, \tau_\eta) \sim N\left(\frac{\sum_{l=1}^{n} w_{il}\eta_l}{\sum_{l=1}^{n} w_{il}}, \frac{\tau_\eta}{\sum_{l=1}^{n} w_{il}}\right), \text{ for } i = 1, \ldots, n, \tag{4.4}$$

where $\eta_{\sim i}$ represents the set of values of all neighbours of the i-th region, except for the i-th region itself, $\mathbf{W}$ is the spatial weights matrix and $\tau_\eta > 0$ is an unknown variance parameter that needs to be estimated. Note that $\mathbf{W}$ is a spatial neighbourhood structure that could be given by contiguity, distance or any other criterion.

The term δ_j is a normally distributed random effect for the temporal dimension, so that $\delta_j \sim N(0, \tau_\delta)$, $\tau_\delta > 0$, for $j = 1, \ldots, J$. For the structured temporal random effect ϕ_j, a random walk process of order one (RW1) can be specified. That is, we could assume that the differences $(\phi_j - \phi_{j-1})$ follow a Normal distribution, so that $(\phi_j - \phi_{j-1}) \sim N(0, \tau_\phi)$, for $j = 2, \ldots, J$, and $\phi_1 \sim N(0, \tau_\phi)$, with $\tau_\phi > 0$. Additionally, for this effect, an autoregressive prior (AR1) can also be considered, so that $\phi_j \sim N(\rho_\phi \phi_{j-1}, \tau_\phi)$ for $j = 2, \ldots, J$, and $\phi_1 \sim N(0, \tau_\phi)$, with ρ_ϕ being the autoregressive parameter and $\tau_\phi > 0$. Note that random walk or autoregressive priors of higher order could also be specified.

Moreover, Knorr-Held (2000) also proposed to include a spatio-temporal interaction term ϵ_{ij}, so that the means μ_{ij} are modelled in the following way:

$$g(\mu_{ij}) = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \nu_i + \eta_i + \delta_j + \phi_j + \epsilon_{ij} \tag{4.5}$$

For this term, an unstructured normal prior distribution with zero mean and variance $\tau_\epsilon > 0$ is often assumed, so that $\epsilon_{ij} \sim N(0, \tau_\epsilon)$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. However, other structured prior distributions could also be considered by combining different effects (Lawson, 2008). This term would allow us to capture the variability of the temporal trend of the response variable for the different regions (Blangiardo and Cameletti, 2015).

4.3 Proposed models: Spatio-temporal conditional models

In this section, we propose some extensions of the spatial conditional models in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) that will allow us to model spatio-temporal count data. In these proposals, the regression structure for the mean will include, for each time period, the lag term of the response variable under study. The parameter associated with this term, would represent the strength of the global spatial autocorrelation which can be present in the data. In this sense, positive significant values would suggest positive spatial autocorrelation in the whole time period considered and, negative significant values, negative spatial autocorrelation.

In particular, we assume that the response variables Y_{ij}, conditioned on the values of all the neighbours of the i-th region, but not including the i-th region itself (i.e., $Y_{\sim ij}$), and for the j-th time period, follow a count data distribution $f(\cdot)$, with conditional means $\mathrm{E}(Y_{ij}|Y_{\sim ij}) = \mu_{ij}$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$, following the regression structure:

$$g(\mu_{ij}) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{y}_j, \tag{4.6}$$

where $\mathbf{W}_i$ is the i-th row of the $n \times n$ spatial weights matrix $\mathbf{W}$, $\mathbf{y}_j$ is the $n \times 1$ vector of observations for all n spatial units for time period j, and ρ is the parameter that captures the strength of the spatial association. Additionally, $g(\cdot)$, $\mathbf{x}_{ij}$ and $\boldsymbol{\beta}$ are as before.

In these models, by including the spatial lags $\mathbf{W}\mathbf{y}_j$, together with the spatial coefficient ρ, it is possible to account for the spatial dependence on the whole time period. Note that $\mathbf{W}\mathbf{y}_j$ are $n \times 1$ vectors representing the spatial lags of the response variable for each time period, for which we will assume that the spatial structure does not change over time and, hence, that the spatial matrix $\mathbf{W}$ remains invariant.

Furthermore, in order to be able to take into account the temporal correlation that might be present in the data, as well as its interaction with the possible existing spatial correlation, we propose to include in the model a set of random effects such as the ones that were proposed by Knorr-Held (2000) and were described in the previous section for equation (4.5).

Therefore, we assume that Y_{ij}, conditioned on $Y_{\sim ij}$ and on the random effects ν_i, δ_j, ϕ_j and ϵ_{ij}, follows a count data distribution $f(\cdot)$, with conditional mean $\mathrm{E}(Y_{ij}|Y_{\sim ij}, \nu_i, \delta_j, \phi_j, \epsilon_{ij}) = \mu_{ij}$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$, so that:

$$g(\mu_{ij}) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{y}_j + \nu_i + \delta_j + \phi_j + \epsilon_{ij}, \tag{4.7}$$

where $g(\cdot)$, $\mathbf{x}_{ij}$, $\boldsymbol{\beta}$, $\mathbf{W}_i$, $\mathbf{y}_j$, ρ, ν_i, δ_j, ϕ_j and ϵ_{ij} are as before.

For Poisson distributed responses, we will assume that $(Y_{ij}|Y_{\sim ij}, \nu_i, \delta_j, \phi_j, \epsilon_{ij}) \sim \mathrm{Poi}(\mu_{ij})$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$, where the conditional mean $\mathrm{E}(Y_{ij}|Y_{\sim ij}, \nu_i, \delta_j, \phi_j, \epsilon_{ij}) = \mu_{ij}$ follows the regression structure:

$$\log(\mu_{ij}) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{y}_j + \nu_i + \delta_j + \phi_j + \epsilon_{ij} \tag{4.8}$$

In the case of binomially distributed responses, we consider that the number of trials on each region i and time period j, is denoted by n_{ij}, so that we can assume that $(Y_{ij}|Y_{\sim ij}, \nu_i, \delta_j, \phi_j, \epsilon_{ij}) \sim \mathrm{Bin}(n_{ij}, \pi_{ij})$, where π_{ij} is the probability of success of a trial in region i and time period j. In order to be able take into account the possible existent spatial dependence, Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) proposed to include the spatial term $\mathbf{A}$, already described in Chapter 2, Section 2.3.1, for cross-sectional data.

We propose to adapt this term for the case of spatio-temporal data. Hence, we define $\mathbf{A}$ as an $n \times J$ matrix, where each column $\mathbf{A}_j$ corresponds to the $n \times 1$ vector of spatial terms for the j-th time period. Each of the elements of this vector will be denoted by A_{ij}, so that $A_{ij} = \frac{\hat{\pi}_{\sim ij}}{1-\hat{\pi}_{\sim ij}}$, where $\hat{\pi}_{\sim ij} = \frac{\mathbf{W}_i \mathbf{y}_j}{\mathbf{W}_i \mathbf{n}_j}$ with $\mathbf{W}_i$ being the i-th row of spatial

weights matrix $\mathbf{W}$, corresponding to the i-th region, and $\mathbf{y}_j$ and $\mathbf{n}_j$ being the vector of observations of the response variable and the vector of the number of trials, respectively, for the j-th time period.

Therefore, we specify a regression structure on the probability of success π_{ij} where we will include the spatial terms A_{ij}, described above, together with a spatial parameter ρ, that will explain the spatial dependence in the data for the whole time period. In particular, π_{ij} is modelled in the following way:

$$\text{logit}(\pi_{ij}) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + \rho A_{ij} + \nu_i + \delta_j + \phi_j + \epsilon_{ij}, \tag{4.9}$$

where $\mathbf{x}_{ij}$, $\boldsymbol{\beta}$, $\mathbf{W}_i$, $\mathbf{y}_j$, ρ, ν_i, δ_j, ϕ_j and ϵ_{ij} are as described before.

We believe it is important to mention that, with regard to the possible existence of overdispersion in the data, these models would be able to account for it. For example, the inclusion of the unstructured areal random effect would take into account the extra-variability in the regions that the spatial lag terms might not able to explain. In addition, the temporal random effect can capture the overdispersion across time not explained by the structured temporal effect. Finally, the interaction term would be able to capture the overdispersion that would still be present after taking the other effects into account.

Finally, it would also be useful to highlight that the models proposed in equations (4.8) and (4.9) can be considered as extensions for spatio-temporal data of the spatial conditional normal Poisson and the spatial conditional normal binomial models, respectively, proposed by Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018 within the cross-sectional data context.

4.3.1 Proposed models: Temporally varying spatial lag coefficient models

The spatial conditional models have been shown to be quite flexible alternatives for modelling spatial count data (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018; Morales-Otero and Núñez-Antón, 2021). The fact that the spatial autocorrelation is captured by a fixed effect or parameter allows us to specify different extensions for these models within the spatio-temporal context. One of these possibilities could be the specification of a temporally varying coefficient for the spatial term that is included in the linear predictor.

Varying coefficient models were proposed by Hastie and Tibshirani (1993), where the authors introduced the so called effect modifiers, meaning that the coefficient of a given covariate is allowed to vary smoothly as a function of another variable. This could be any variable, including factors such as time, age, gender or any other variable/factor.

Franco-Villoria, Ventrucci and Rue (2019) provided a thorough revision on varying coefficient models, indicating that they are especially useful when the effect of a covariate might depend on some other covariate, that could represent time or space. The authors describe how, for the temporal case, these coefficients can be defined as unstructured effects, or given by structures dependant of time, such as random walks or autoregressive processes, among others.

In a spatio-temporal context, when applying the proposed spatio-temporal conditional models given by equation (4.7), we could assume that time might influence the effect of the spatial lag on the response variable under study. A random coefficient on the spatial term would allow for this term to have a different effect on the response for each time period.

Therefore, we propose the temporally varying spatial lag coefficient model, where we assume that, in the spatio-temporal conditional model in equation (4.7), the coefficient for the spatial lag is the sum of a fixed coefficient, ρ_0, and a random coefficient, ρ_j, that varies according to the time units $j = 1, \ldots, J$. In particular, we specify the following regression structure for the conditional mean:

$$g(\mu_{ij}) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + (\rho_0 + \rho_j)\mathbf{W}_i\mathbf{y}_j + \nu_i + \delta_j + \phi_j + \epsilon_{ij}, \tag{4.10}$$

where $g(\cdot)$, $\mathbf{x}_{ij}$, $\boldsymbol{\beta}$, $\mathbf{W}_i$, $\mathbf{y}_j$, ρ, ν_i, δ_j, ϕ_j and ϵ_{ij} are as before. These models can be specified for Poisson or binomial responses, by assuming a proper link function, $g(\cdot)$, such as the logarithmic or the logistic function, respectively.

Here, we propose three possible ways in which the temporal varying coefficient could be specified. One possibility could be to consider an unstructured normal distribution, that is $\rho_j \sim N(0, \tau_\rho)$, with $\tau_\rho > 0$, for $j = 1, \ldots, J$. Another way would be given by a random walk process of order one (RW1), so that $(\rho_j - \rho_{j-1}) \sim N(0, \tau_\rho)$, for $j = 2, \ldots, J$ and $\rho_1 \sim N(0, \tau_\rho)$, with $\tau_\rho > 0$. Lastly, an autoregressive prior (AR1) could also be considered, so that $\rho_j \sim N(\omega\rho_{j-1}, \tau_\rho)$ for $j = 2, \ldots, J$, and $\rho_1 \sim N(0, \tau_\rho)$, with ω being the autoregressive parameter to be estimated and $\tau_\rho > 0$.

The estimated value obtained for ρ_0 would represent the strength of the spatial dependence among the regions for the whole time period considered, whereas, the estimated values obtained for ρ_j would indicate whether the spatial association increases or decreases with time. This would allow us to examine the variability of the coefficient of the spatial lag from one time unit to the others.

Taking into account the value obtained for ρ_0, a positive estimated value of ρ_j would suggest that for time period j, the strength of the spatial association is larger than that indicated by ρ_0. On the contrary, a negative estimated value of ρ_j would indicate that, for the j-th time period the spatial autocorrelation is weaker. If $\hat{\rho}_j \approx 0$, this would mean that there are no significant changes in the spatial correlation pattern for the j-th time period, with respect to that of ρ_0.

Finally, it could be useful to mention that the proposed temporally varying coefficient model can relate to the one proposed by Bernardinelli et al. (1995), described in Section 4.2, in the way that they assumed an interaction between space and time given by a spatially varying coefficient for the temporal trend.

4.4 Application to respiratory hospital admissions in Glasgow

In this section, we study a data set concerning the impact of air pollution on the respiratory health of the population living in each of the $n = 271$ regions or statistical sectors

belonging to the Scotland National Health System's board of Greater Glasgow and the Clyde Valley, Scotland, for a time period of $J = 5$ years (i.e., from 2007 to 2011). This data set is presented as an example in Lee, Rushworth and Napier (2018) and can be obtained from the R package CARBayesdata (Lee, Rushworth and Napier, 2018).

The variables available for each region and time period are the observed number of respiratory hospital admissions (i.e., variable Y), the expected number of respiratory hospital admissions (i.e., variable E), the yearly average modelled concentrations of particulate matter less than 10 microns (i.e., variable PM10), the average property price in each region and year, given in hundreds of thousands of pounds (i.e., variable Price) and the proportion of the working age population who are in receipt of job seekers allowance (i.e., variable JSA). Table 4.1 includes some descriptive statistics for each one of the years under study.

Table 4.1: Descriptive statistics for the variables available in the study of respiratory hospital admissions in Glasgow.

	Observed=Y					Expected=E				
	2007	2008	2009	2010	2011	2007	2008	2009	2010	2011
Median	70.0	75.0	73.0	76.0	80.0	84.3	88.9	85.7	91.6	92.4
Mean	75.3	81.0	78.1	78.4	83.2	87.6	92.1	89.3	96.0	96.8
SD	32.8	37.0	34.2	33.0	35.0	22.4	23.7	22.7	24.7	24.9
Min.	15.0	10.0	20.0	23.0	20.0	44.5	47.4	44.7	47.6	49.2
Max.	194.0	208.0	190.0	213.0	189.0	164.3	173.8	164.8	180.5	180.3

	PM10					JSA				
	2007	2008	2009	2010	2011	2007	2008	2009	2010	2011
Median	13.8	12.2	10.7	11.4	12.9	2.5	2.8	4.5	4.6	4.8
Mean	13.9	12.3	10.9	11.5	13.0	2.9	3.2	4.8	5.0	5.1
SD	2.0	1.5	1.5	1.5	1.7	1.8	1.9	2.5	2.7	2.8
Min.	9.9	9.0	7.8	8.5	9.6	0.3	0.4	1.0	0.6	0.7
Max.	19.6	18.4	17.3	17.7	18.8	8.6	8.5	12.7	13.8	13.1

	Price				
	2007	2008	2009	2010	2011
Median	1.2	1.2	1.1	1.1	1.1
Mean	1.3	1.3	1.2	1.2	1.2
SD	0.6	0.5	0.5	0.6	0.6
Min.	0.6	0.5	0.5	0.2	0.3
Max.	4.3	3.7	4.0	3.8	4.0

Note that the expected number of respiratory hospital admissions represents the

number of admissions that would be expected in each region and year if the population in the specific region behaved as the standard or overall population does. In Lee, Rushworth and Napier (2018), this variable was obtained by using indirect standardization and the national age and sex specific admissions rates.

If the population is divided into S subsets (or strata), according to the individuals' age and sex, we can compute the rate of admissions $r_l^{(S)}$ for each of the $l = 1, \ldots, S$ subsets, by dividing the number of observed admissions (i.e., Y_l) by the population (i.e., P_l) in stratum l, so that:

$$r_l^{(S)} = \frac{Y_l}{P_l},$$

where $Y_l = \sum_{i=1}^{n} \sum_{j=1}^{J} Y_{ijl}$ and $P_l = \sum_{i=1}^{n} \sum_{j=1}^{J} P_{ijl}$, with Y_{ijl} and P_{ijl} being the number of cases and the population, respectively, in the stratum l for the i-th area and the j-th time period.

Therefore, the expected number of cases will then be obtained in the following way:

$$E_{ij} = \sum_{l=1}^{S} r_l^{(S)} P_{ij}$$

The standardized incidence ratio (SIR) is often employed in the context of disease mapping. It allows to compare the disease rate of the population in each specific area and the rate of the standard population in the whole area under study (Moraga, 2018). In this case, it can be obtained by dividing the number of respiratory hospital admissions by the expected counts, that is $\text{SIR}_{ij} = Y_{ij}/E_{ij}$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$.

Figure 4.1 shows the spatial distribution of the SIR's for the respiratory hospital admissions in Glasgow for every year under study. A SIR value larger than one, for a specific region and time unit (i.e., $\text{SIR}_{ij} > 1$), indicates that, for this region and year, the incidence of the disease for that population is larger than expected. On the contrary, values smaller than one (i.e., $\text{SIR}_{ij} < 1$) suggest that the incidence in the i-th region for the j-th time period is smaller than it would be expected.

Special attention would need to be given to the population living in regions with SIR values larger than one, since they would have an increased risk of suffering from serious respiratory problems. In general, regions with the highest incidence are concentrated in the central areas to the east of the map, corresponding to the Glasgow City council, a pattern that is maintained over the years under study.

In addition, Figure 4.1 also allows us to observe how the spatial distribution of the incidence changes over time for some of the regions in the periphery. For example, for the region labelled as IZ Sixteen, belonging to the West Dunbartonshire council and located in the north of the area under study, we can see how the SIR considerably increases in the year 2011. This can be better seen in Figure 4.2, where we have included the maps of the SIR's for the last three years, highlighting the region IZ Sixteen by enclosing it in a red box. In order to be able to clearly distinguish the behaviour of the region IZ Sixteen for the last three years under study, we have decided to illustrate this in a separate figure from that included in Figure 4.1.

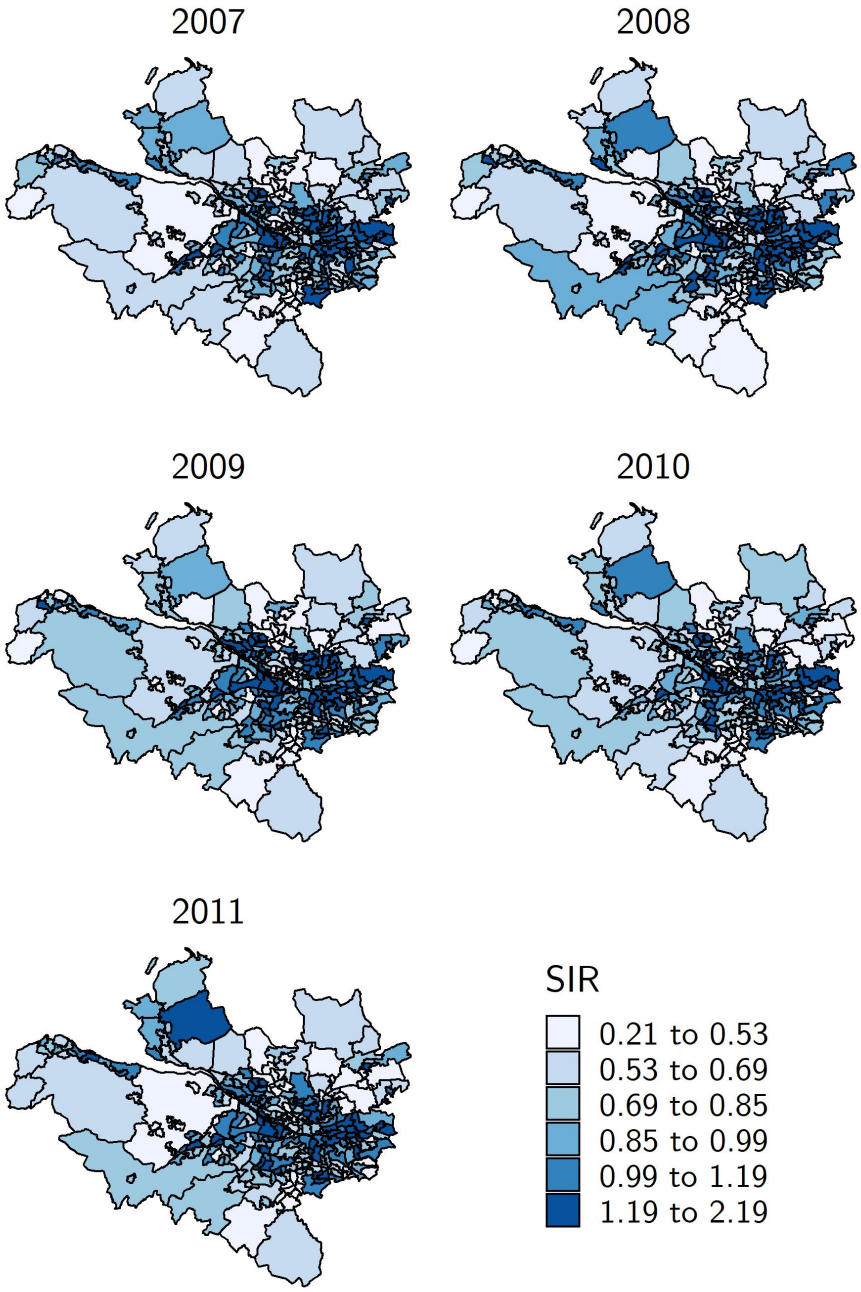


Figure 4.1: Spatial distribution of the SIR's in Glasgow by region and year.

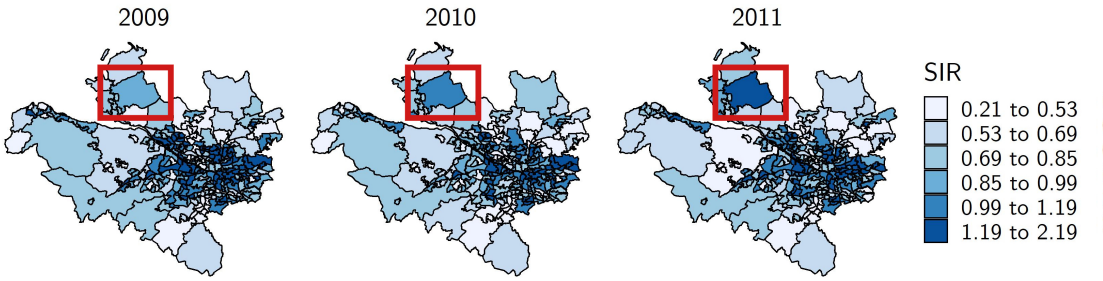


Figure 4.2: Spatial distribution of the SIR's in Glasgow by region for the last three years under study, where the region labelled as IZ Sixteen is indicated with a red box.

Lee, Rushworth and Napier (2018) indicated that the covariates included in this study are suitable for explaining the respiratory hospital admissions in Glasgow for the regions and time period under study, and they found evidence for the statistical significance of the relationship of these variables with the modelled disease risks. In addition, the authors confirmed that there were signs of the presence of spatio-temporal autocorrelation in this data. The applied regression model was able to capture the spatio-temporal autocorrelation by incorporating a random effect with a multivariate autoregressive prior distribution.

In this section, we will mainly focus on evaluating the performance of the models proposed in Section 4.3 by fitting them to the data set presented here. Furthermore, we also wish to compare them with the models proposed by Knorr-Held (2000), previously described in Section 4.2, which we believe are the most commonly used by practitioners in spatio-temporal applications.

In our specific Bayesian framework, we will specify noninformative prior distributions. In particular, normal priors with large variances for all the coefficients β_k, for $k = 0, \ldots, 3$ (i.e., $\beta_k \sim N(0, \text{1e-05})$) and also, for the spatial parameter ρ (i.e., $\rho \sim N(0, \text{1e-05})$). For the precision parameters (i.e., for $\psi_\nu = 1/\tau_\nu$, $\psi_\eta = 1/\tau_\eta$, $\psi_\delta = 1/\tau_\delta$, $\psi_\phi = 1/\tau_\phi$ and $\psi_\epsilon = 1/\tau_\epsilon$), we will assume vague gamma priors, so that $\psi_{(\cdot)} \sim G(\text{1e-04}, \text{1e-04})$.

In order to be able to justify the selected values for these priors, we have carried out a sensitivity analysis and include its results in Section 4.4.3. In addition, all the models will be fitted in R-INLA and we will assume that the spatial structure of the data under study is represented by a spatial weights matrix $\mathbf{W}$, following the contiguity of order one criterion. This criterion was selected after conducting tests where we fitted the models considering other matrices following different criteria, finding that there were no significant differences in the results obtained.

4.4.1 Fitting of the spatio-temporal models

It is known that the SIR's can provide a preliminary indicator about the increased incidence of a disease. However, these quantities do not take into account the information

from neighbouring areas, or the spatial and temporal autocorrelation that may exist in the data. Moreover, they are sensitive to extreme values given by low population in some areas.

Therefore, it is often preferred to estimate the relative risks of a disease by fitting regression models, which can include a series of covariates and terms such as spatial or temporal effects, among others, that can serve as proxies of the characteristics of the areas under study. The relative risk θ_{ij} represents the risk of infection of the population in the i-th region and j-th time period, relative to the risk of the population in general in the whole area under study.

In this specific application, our aim is to model the risk of the population living in each area and time period, for having a respiratory disease that leads the patient to hospitalisation, taking into account the spatio-temporal autocorrelation that might be present in the data. To this end, let us suppose that the observed number of respiratory hospital admissions, Y_{ij}, for each region and year considered, follows a Poisson distribution with mean μ_{ij}, that is $Y_{ij} \sim \text{Poi}(\mu_{ij})$, for $i = 1, ..., n$ and $j = 1, ..., J$. If θ_{ij} represents the relative risk in region i and time period j being modelled, it is assumed that the mean μ_{ij} is a product of the risk and the expected count, that is $\mu_{ij} = \theta_{ij} E_{ij}$. Then, we specify a regression model for μ_{ij}, including the covariates available, which is given by:

$$\log(\mu_{ij}) = \log(E_{ij}) + \beta_0 + \beta_1 \text{JSA}_{ij} + \beta_2 \text{Price}_{ij} + \beta_3 \text{PM10}_{ij}, \tag{4.11}$$

where the expected number of respiratory hospital admissions, E_{ij}, is incorporated as an offset term in the natural logarithmic scale, so that we are able to model the relative risk θ_{ij}.

In order to account for the spatio-temporal autocorrelation that is presumed to be present in the data, we will fit a spatio-temporal model such as the one proposed by Knorr-Held (2000), described in equation (4.5) in Section 4.2. Therefore, we assume that the observed number of respiratory hospital admissions on each region and time unit, given the random effects $\nu_i, \eta_i, \delta_j, \phi_j$ and ϵ_{ij}, follows a Poisson distribution. That is, $(Y_{ij}|\nu_i, \eta_i, \delta_j, \phi_j, \epsilon_{ij}) \sim \text{Poi}(\mu_{ij})$, with means $\mu_{ij} = E_{ij}\theta_{ij}$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. Here, a regression structure for μ_{ij} is specified, so that:

$$\begin{aligned} \log(\mu_{ij}) = &\log(E_{ij}) + \beta_0 + \beta_1 \text{JSA}_{ij} + \beta_2 \text{Price}_{ij} + \beta_3 \text{PM10}_{ij} \\ &+ \nu_i + \eta_i + \delta_j + \phi_j + \epsilon_{ij}, \end{aligned} \tag{4.12}$$

with η_i a spatial ICAR distributed random effect, (i.e., $\eta_i \sim \text{ICAR}(0, \tau_\eta)$, $\tau_\eta > 0$) and ν_i an unstructured spatial random effect (i.e., $\nu_i \sim N(0, \tau_\nu)$, $\tau_\nu > 0$) that will explain the spatial dependence and the extra-variability, respectively, in the areal dimension. The temporal dependence is captured by an effect following a random walk process of order one, that is $\phi_j \sim N(\phi_{j-1}, \tau_\phi)$, for $j = 2, \ldots, J$, and $\phi_1 \sim N(0, \tau_\phi)$, with $\tau_\phi > 0$, and δ_j an unstructured random effect for the temporal dimension (i.e., $\delta_j \sim N(0, \tau_\delta)$, $\tau_\delta > 0$). Finally, the interaction effect ϵ_{ij} will account for the remaining variability that could still be present (i.e., $\epsilon_{ij} \sim N(0, \tau_\epsilon)$, $\tau_\epsilon > 0$).

Note that we have used the random walk process of order one for the structured temporal effect ϕ_j, following the proposals in Knorr-Held (2000), Lawson (2008) and

Blangiardo and Cameletti (2015), which is the simplest temporal structure that can be specified for this term. According to these authors, a random walk process of order two could be considered if the objective of the study was to perform predictions, which is not what we intend in this specific application.

The results obtained after fitting the models in equations (4.11) and (4.12), as well as some of their reduced versions are shown in Table 4.2. For all the fitted models, the covariates included were statistically significant, according to their 95% credible intervals.

The coefficients obtained for the variables PM10 and JSA were always positive, indicating that large values of the concentration of particulate matter less than 10 microns, and also that large values of the proportion of the working age population receiving job seekers allowance, are associated with an increase of the relative risk of respiratory hospital admissions. Note that for the model only including the spatially structured and unstructured random effects η_i and ν_i, respectively, the variable JSA did not result statistically significant. For the variable Price, the coefficient was always negative, suggesting that regions with larger property price values have less risk of respiratory hospital admissions. These results are consistent with the ones obtained by Lee, Rushworth and Napier (2018).

The smallest information criteria values (DIC = 10378 and WAIC = 10336) were obtained for the model where the linear predictor, besides the explanatory variables, includes only the spatially unstructured and structured random effects, ν_i and η_i, respectively, and the spatio-temporal interaction term ϵ_{ij}. That is, the model with linear predictor:

$$\log(\mu_{ij}) = \log(E_{ij}) + \beta_0 + \beta_1 \text{JSA}_{ij} + \beta_2 \text{Price}_{ij} + \beta_3 \text{PM10}_{ij} + \nu_i + \eta_i + \epsilon_{ij} \qquad (4.13)$$

Moreover, when including the temporal effects δ_j and ϕ_j in the linear predictor, the DIC and WAIC values increase. Hence, these terms do not seem to offer improvements in the fitting of the models according to these information criteria. In addition, when including the interaction term ϵ_{ij} these values considerably decrease. Therefore, in this specific case, it appears that the temporal correlation is better explained by the interaction term than with the separate temporal effects. We should indicate that, for this specific model, all of the other temporal effects (δ_j and ϕ_j) are most likely captured by ϵ_{ij}.

Figure 4.3 includes the maps of the estimated relative risks obtained after fitting the model in equation (4.13) to the respiratory hospital admissions data in Glasgow, for each of the years under study. These estimations are very similar to the observed SIR values, shown in Figure 4.1. As was already observed in the spatial distribution of the SIR's, here the regions with the highest relative risk of infection are also located in the central areas of Glasgow. This behaviour is stable for the different years, with small changes that are most visible for the peripheral regions.

Table 4.2: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the spatio-temporal Poisson model and some of its reduced versions, fitted to the respiratory hospital admissions in Glasgow data set.

		Intercept	**PM10**	**JSA**	**Price**	τ_ν	τ_η	τ_δ	τ_ϕ	τ_ϵ
DIC = 14094	Mean	-0.598	0.042	0.060	-0.283	-	-	-	-	-
WAIC = 14110	SD	(0.026)	(0.002)	(0.002)	(0.009)	-	-	-	-	-
	95% CI	(-0.648,-0.547)	(0.039,0.045)	(0.057,0.063)	(-0.300,-0.266)	-	-	-	-	-
DIC = 10727	Mean	-0.150	0.007	-0.001	-0.108	0.025	0.196	-	-	-
WAIC = 10926	SD	(0.054)	(0.003)	(0.003)	(0.022)	(0.007)	(0.037)	-	-	-
	95% CI	(-0.256,-0.043)	(9.44e-04,0.013)	(-0.007,0.004)	(-0.150,-0.065)	(0.013,0.041)	(0.134,0.280)	-	-	-
DIC = 10730	Mean	-0.375	0.017	0.031	-0.136	0.015	0.122	0.006	-	-
WAIC = 10929	SD	(0.090)	(0.006)	(0.005)	(0.021)	(0.005)	(0.030)	(0.005)	-	-
	95% CI	(-0.551,-0.196)	(0.005,0.029)	(0.021,0.041)	(-0.176,-0.094)	(0.007,0.027)	(0.075,0.194)	(0.001,0.020)	-	-
DIC = 10734	Mean	-0.369	0.016	0.032	-0.137	0.015	0.120	0.007	0.006	-
WAIC = 10934	SD	(0.087)	(0.006)	(0.005)	(0.021)	(0.005)	(0.030)	(0.033)	(0.008)	-
	95% CI	(-0.541,-0.198)	(0.004,0.029)	(0.022,0.042)	(-0.177,-0.095)	(0.007,0.027)	(0.074,0.191)	(1.99e-05,0.160)	(2.17e-04,0.027)	-
DIC = 10389	Mean	-0.441	0.019	0.057	-0.187	0.015	0.052	0.003	0.008	0.011
WAIC = 10352	SD	(0.102)	(0.007)	(0.006)	(0.023)	(0.004)	(0.018)	(0.010)	(0.009)	(0.001)
	95% CI	(-0.642,-0.240)	(0.004,0.033)	(0.046,0.068)	(-0.233,-0.141)	(0.008,0.025)	(0.025,0.094)	(-1.48e-04,0.021)	(5.72e-04,0.031)	(0.009,0.014)
DIC = 10387	Mean	-0.446	0.019	0.056	-0.188	0.016	0.050	0.013	-	0.011
WAIC = 10351	SD	(0.111)	(0.008)	(0.006)	(0.024)	(0.004)	(0.018)	(0.013)	-	(0.001)
	95% CI	(-0.663,-0.226)	(0.005,0.034)	(0.045,0.067)	(-0.234,-0.141)	(0.009,0.026)	(0.023,0.092)	(0.002,0.046)	-	(0.009,0.013)
DIC = 10378	Mean	-0.250	0.017	0.013	-0.178	0.017	0.136	-	-	0.011
WAIC = 10336	SD	(0.071)	(0.004)	(0.004)	(0.027)	(0.006)	(0.032)	-	-	(0.001)
	95% CI	(-0.390,-0.110)	(0.009,0.025)	(0.004,0.022)	(-0.230,-0.126)	(0.007,0.030)	(0.087,0.211)	-	-	(0.009,0.013)
DIC = 10406	Mean	-0.575	0.031	0.063	-0.224	0.028	-	0.023	0.010	0.012
WAIC = 10364	SD	(0.086)	(0.006)	(0.005)	(0.021)	(0.003)	-	(0.027)	(0.013)	(0.001)
	95% CI	(-0.744,-0.406)	(0.020,0.043)	(0.052,0.073)	(-0.266,-0.182)	(0.022,0.035)	-	(2.89e-05,0.071)	(4.57e-04,0.042)	(0.010,0.014)
DIC = 10405	Mean	-0.578	0.032	0.062	-0.225	0.028	-	0.011	-	0.012
WAIC = 10364	SD	(0.094)	(0.006)	(0.005)	(0.021)	(0.003)	-	(0.010)	-	(0.001)
	95% CI	(-0.762,-0.392)	(0.020,0.043)	(0.052,0.072)	(-0.267,-0.183)	(0.022,0.035)	-	(0.002,0.039)	-	(0.010,0.014)
DIC = 10410	Mean	-0.280	0.025	0.018	-0.246	0.051	-	-	-	0.013
WAIC = 10357	SD	(0.072)	(0.004)	(0.005)	(0.026)	(0.007)	-	-	-	(0.001)
	95% CI	(-0.421,-0.139)	(0.017,0.033)	(0.009,0.027)	(-0.296,-0.196)	(0.039,0.066)	-	-	-	(0.010,0.015)
DIC = 10662	Mean	-0.612	0.039	0.065	-0.279	-	-	-	-	0.043
WAIC = 10513	SD	(0.054)	(0.003)	(0.003)	(0.017)	-	-	-	-	(0.002)
	95% CI	(-0.718,-0.506)	(0.033,0.046)	(0.059,0.072)	(-0.312,-0.247)	-	-	-	-	(0.039,0.047)

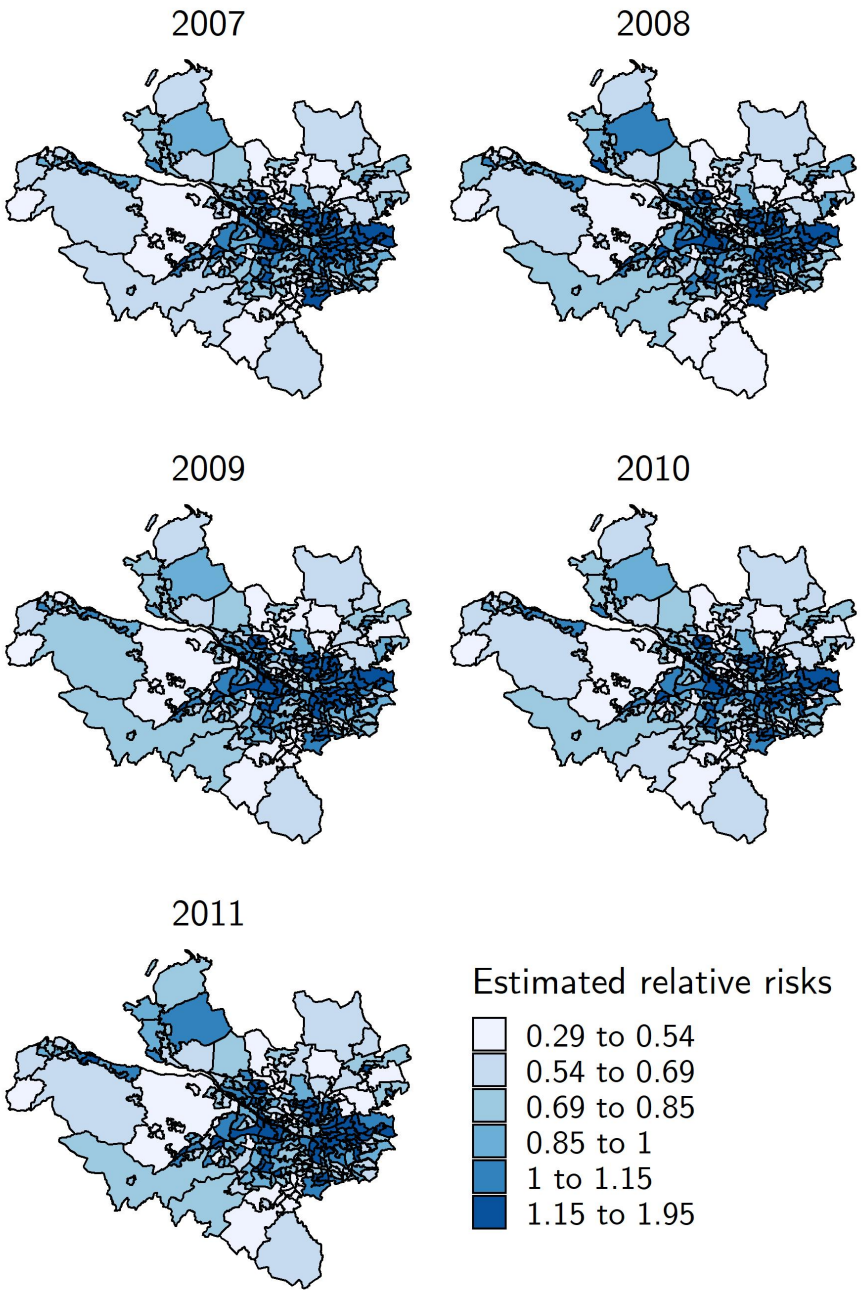


Figure 4.3: Maps of the estimated relative risks, for the different years under study, obtained from the spatio-temporal model in equation (4.13), fitted to the respiratory hospital admissions in Glasgow data set.

4.4.2 Fitting of the spatio-temporal conditional models

In this section, we fit the proposed spatio-temporal conditional models and compare the results obtained with the models fitted in the previous section. Note that, in this case, we are interested in modelling the relative risks and, for this reason, we will include the spatial lags of the SIR's in the regression model. Let us recall that we proceeded in a similar way in Chapter 2, Section 2.6.1, where we fitted the spatial conditional Poisson models for the infant mortality rates in Colombia, including the spatial lag of the observed rates in the regression structure for the conditioned means.

Therefore, we assume that $(Y_{ij}|Y_{\sim ij}, \nu_i, \delta_j, \phi_j, \epsilon_{ij}) \sim \text{Poi}(\mu_{ij})$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. Here, for the means $\mu_{ij} = E_{ij}\theta_{ij}$, we specify the following regression structure:

$$\begin{aligned} \log(\mu_{ij}) = &\log(E_{ij}) + \beta_0 + \beta_1 \text{JSA}_{ij} + \beta_2 \text{Price}_{ij} + \beta_3 \text{PM10}_{ij} + \rho \mathbf{W}_i \mathbf{SIR}_j \\ &+ \nu_i + \delta_j + \phi_j + \epsilon_{ij}, \end{aligned} \tag{4.14}$$

where $\mathbf{W}_i$ is the i-th row of the spatial weights matrix and $\mathbf{SIR}_j$ is the vector of observations of the SIR's for the j-th time period. The elements of this vector are given by the number of respiratory hospital admissions divided by the expected counts, in each of the regions considered. In addition, the random effects ν_i, δ_j, ϕ_j and ϵ_{ij} are as before.

The results obtained from the fitting of this model and some of its reduced versions are included in Table 4.3. Here, we can see that all the covariates were statistically significant, according to their 95% credible intervals. The estimated values for their coefficients are very similar to those obtained when fitting the models from the previous section, and, thus, providing evidence of the same relationships between the covariates included in the study and the respiratory hospital admissions. In addition, the spatial lag coefficient is also statistically significant for all the models considered. Moreover, it has always a positive value, indicating that this term is properly capturing the positive spatial autocorrelation present in the data.

When compared to the spatio-temporal models from Table 4.2, these models have smaller DIC and WAIC values. For example, these values for the spatio-temporal model in equation (4.12) were DIC = 10389 and WAIC = 10352, and for the spatio-temporal conditional model in equation (4.14), were DIC = 10373 and WAIC = 10343, much lower values which indicate that the latter model offers a better fit than the former.

Table 4.3: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the spatio-temporal conditional Poisson model and some of its reduced versions, fitted to the respiratory hospital admissions in Glasgow data set.

		Intercept	**PM10**	**JSA**	**Price**	ρ	τ_ν	τ_δ	τ_ϕ	τ_ϵ
DIC = 13481	Mean	-0.814	0.024	0.049	-0.226	0.468	-	-	-	-
WAIC = 13497	SD	(0.028)	(0.002)	(0.002)	(0.009)	(0.019)	-	-	-	-
	95% CI	(-0.868,-0.760)	(0.021,0.028)	(0.046,0.052)	(-0.243,-0.208)	(0.431,0.505)	-	-	-	-
DIC = 10720	Mean	-0.670	0.011	0.005	-0.147	0.579	0.052	-	-	-
WAIC = 10914	SD	(0.064)	(0.003)	(0.003)	(0.020)	(0.040)	(0.006)	-	-	-
	95% CI	(-0.795,-0.544)	(0.005,0.016)	(-5.40e-04,0.011)	(-0.186,-0.107)	(0.500,0.658)	(0.041,0.064)	-	-	-
DIC = 10735	Mean	-0.764	0.017	0.036	-0.161	0.474	0.033	0.005	-	-
WAIC = 10931	SD	(0.083)	(0.005)	(0.005)	(0.019)	(0.040)	(0.004)	(0.004)	-	-
	95% CI	(-0.926,-0.600)	(0.006,0.027)	(0.027,0.045)	(-0.198,-0.125)	(0.395,0.553)	(0.026,0.042)	(8.49e-04,0.016)	-	-
DIC = 10737	Mean	-0.761	0.016	0.037	-0.161	0.474	0.033	0.002	0.005	-
WAIC = 10933	SD	(0.077)	(0.005)	(0.005)	(0.019)	(0.040)	(0.004)	(0.005)	(0.009)	-
	95% CI	(-0.911,-0.610)	(0.006,0.026)	(0.028,0.046)	(-0.197,-0.125)	(0.395,0.552)	(0.026,0.041)	(-4.63e-05,0.010)	(1.07e-04,0.026)	-
DIC = 10373	Mean	-0.800	0.017	0.052	-0.182	0.463	0.025	0.002	0.007	0.011
WAIC = 10343	SD	(0.083)	(0.006)	(0.005)	(0.021)	(0.048)	(0.003)	(0.009)	(0.009)	(0.001)
	95% CI	(-0.964,-0.636)	(0.006,0.028)	(0.042,0.062)	(-0.222,-0.141)	(0.368,0.557)	(0.020,0.031)	(-7.38e-05,0.015)	(3.38e-04,0.028)	(0.009,0.013)
DIC = 10373	Mean	-0.802	0.017	0.051	-0.182	0.464	0.025	0.007	-	0.011
WAIC = 10344	SD	(0.090)	(0.006)	(0.005)	(0.021)	(0.048)	(0.003)	(0.006)	-	(0.001)
	95% CI	(-0.979,-0.625)	(0.006,0.029)	(0.041,0.061)	(-0.223,-0.141)	(0.370,0.558)	(0.020,0.031)	(0.001,0.024)	-	(0.009,0.013)
DIC = 10377	Mean	-0.781	0.018	0.017	-0.195	0.611	0.036	-	-	0.011
WAIC = 10342	SD	(0.075)	(0.004)	(0.004)	(0.023)	(0.048)	(0.004)	-	-	(0.001)
	95% CI	(-0.927,-0.635)	(0.011,0.025)	(0.009,0.025)	(-0.239,-0.151)	(0.517,0.705)	(0.028,0.045)	-	-	(0.009,0.013)
DIC = 10633	Mean	-0.828	0.024	0.053	-0.224	0.450	-	-	-	0.037
WAIC = 10506	SD	(0.055)	(0.003)	(0.003)	(0.016)	(0.038)	-	-	-	(0.002)
	95% CI	(-0.935,-0.721)	(0.017,0.031)	(0.047,0.060)	(-0.257,-0.192)	(0.376,0.525)	-	-	-	(0.033,0.041)

The model with the smallest information criteria values (i.e., DIC = 10373 and WAIC = 10343) is the one where all the terms have been included. In addition, another model with a very similar fit was the one where the structured temporal random effect, ϕ_j, was excluded from the linear predictor (i.e., DIC = 10373 and WAIC = 10344). Taking into account that the inclusion of this term does not offer any improvement in terms of information criteria, we will select the latter as the best fitting model. In particular, this model follows the regression structure:

$$\begin{aligned}\log(\mu_{ij}) = &\log(E_{ij}) + \beta_0 + \beta_1 \text{JSA}_{ij} + \beta_2 \text{Price}_{ij} + \beta_3 \text{PM10}_{ij} + \rho \mathbf{W}_i \mathbf{SIR}_j \\ &+ \nu_i + \delta_j + \epsilon_{ij}\end{aligned} \tag{4.15}$$

Let us recall that the information criteria values for the best fitting spatio-temporal model obtained in the previous section were DIC = 10378 and WAIC = 10336. Note that the WAIC value is smaller than the one obtained for the model in equation (4.15). However, we are including here the temporal effect δ_j, an extra term that could be increasing the value of the penalization that this criterion receives.

Figure 4.4 shows the maps of the estimated relative risks obtained from the fitting of the model with linear predictor in equation (4.15). These predictions are very similar to the ones obtained from the model in equation (4.13), shown in Figure 4.3, and hence, to the SIR's obtained for this data, shown in Figure 4.1, which suggests that these two models are both able to provide accurate predictions.

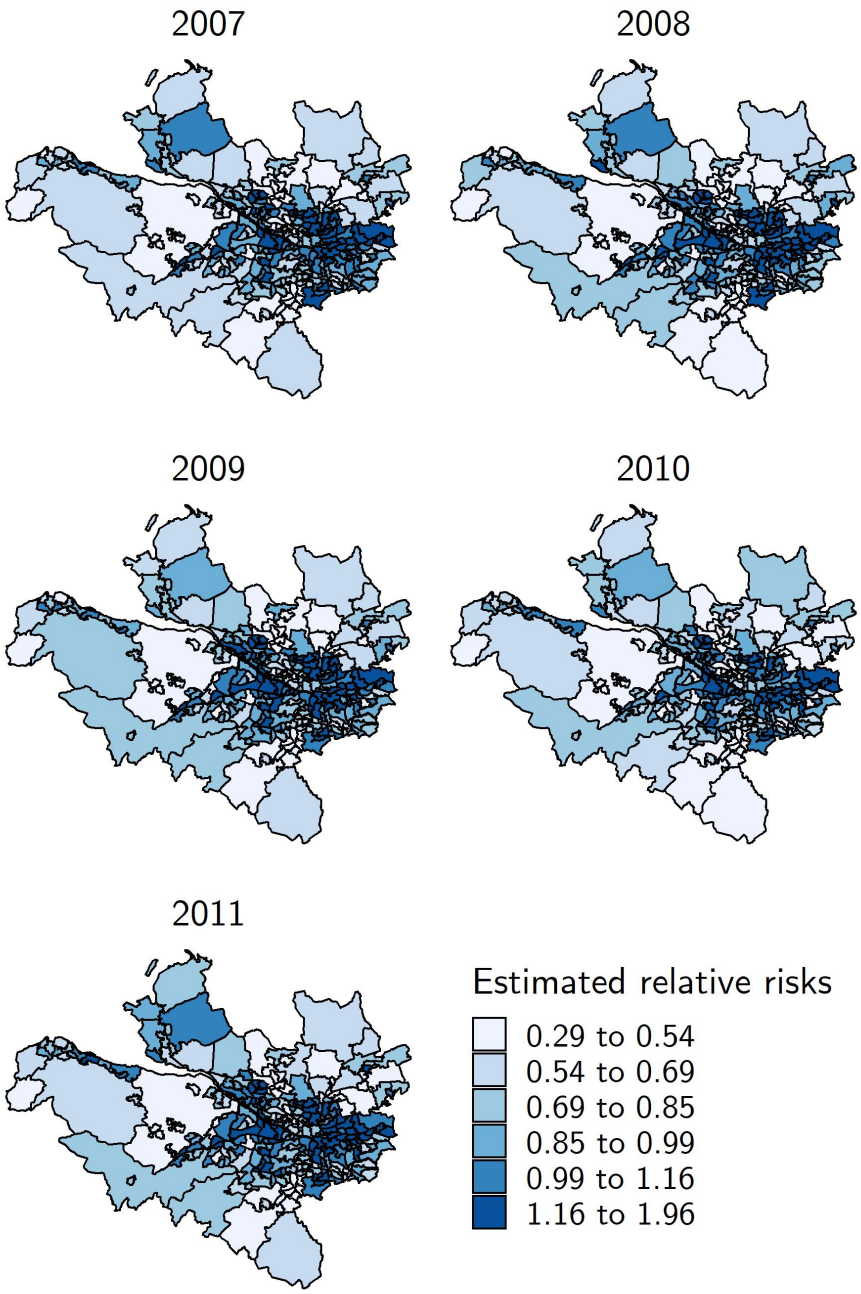


Figure 4.4: Maps of the estimated relative risks, for the different years under study, obtained from the spatio-temporal conditional model in equation (4.15), fitted to the respiratory hospital admissions in Glasgow data set.

4.4.3 Sensitivity analysis for the precision of the prior distributions

As was already discussed for the infant mortality study, in Chapter 2, Section 2.6.2, when including random effects in a model, the choice of the initial values for the prior distributions assigned to the precision parameters could highly influence the posterior inference processes. Therefore, we have performed a sensitivity analysis for these parameters, which we illustrate here for the model with linear predictor in equation (4.15). We have considered different values α, from $\alpha = 0.1$ to $\alpha =$ 1e-08, for the precision parameter in the random effects $\psi_\nu = 1/\tau_\nu$, $\psi_\delta = 1/\tau_\delta$, and $\psi_\epsilon = 1/\tau_\epsilon$, where $\psi_{(.)} \sim G(\alpha, \alpha)$.

Table 4.4 includes the parameters' estimates, DIC and WAIC values obtained after fitting the models for the different values of α in $\psi_\nu \sim G(\alpha, \alpha)$. Here, the value of the prior distribution only changes for the parameter ν, whereas the priors of the other precision parameters remain invariant. The same information is included in Tables 4.5 and 4.6, for the parameters τ_δ and τ_ϵ, respectively. We can see how the values obtained for the estimates are exactly the same, at least up to their fourth decimal place, when the value of α changes from $\alpha =$ 1e-04 up to $\alpha =$ 1e-08.

Figure 4.5 shows the posterior densities obtained after fitting the models for each value of α, where figures 4.5(a), 4.5(b) and 4.5(c) correspond to the precisions of the random effect ν, δ and ϵ, respectively. On the left panels we can see all the posteriors, whereas on the right panels, only the distributions from the value $\alpha =$ 1e-04 up to the value $\alpha =$ 1e-08 are included, where we can see that for these values the distributions do not change. Therefore, we believe that the choice of $\alpha =$ 1e-04 is well justified.

Table 4.4: Posterior means for parameter estimates together with standard deviations, DIC and WAIC values, for the spatio-temporal conditional Poisson model in the analysis of the respiratory hospital admissions in Glasgow data set with different prior distributions for the precision parameter of the random effect τ_ν.

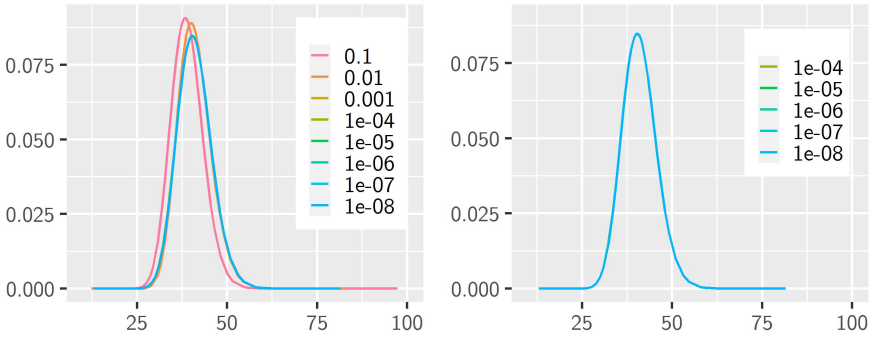
	Intercept	**PM10**	**JSA**	**Price**	ρ	τ_ν	τ_δ	τ_ϵ	DIC	WAIC
$\alpha = 0.1$	-0.8032	0.0174	0.0503	-0.1815	0.4656	0.0261	0.0078	0.0105	10372	10343
	(0.0905)	(0.0058)	(0.0052)	(0.0210)	(0.0485)	(0.0030)	(0.0074)	(0.0010)	10373	10343
$\alpha = 0.01$	-0.8033	0.0174	0.0509	-0.1824	0.4645	0.0247	0.0101	0.0105	10373	10344
	(0.0895)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0027)	(0.0116)	(0.0010)	10373	10344
$\alpha = 0.001$	-0.8025	0.0173	0.0511	-0.1823	0.4638	0.0248	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0061)	(0.0011)	10373	10344
$\alpha = 1e\text{-}04$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344
$\alpha = 1e\text{-}05$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10372	10343
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10343
$\alpha = 1e\text{-}06$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344
$\alpha = 1e\text{-}07$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344
$\alpha = 1e\text{-}08$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344

Table 4.5: Posterior means for parameter estimates together with standard deviations, DIC and WAIC values, for the spatio-temporal conditional Poisson model in the analysis of the respiratory hospital admissions in Glasgow data set with different prior distributions for the precision parameter of the random effect τ_δ.

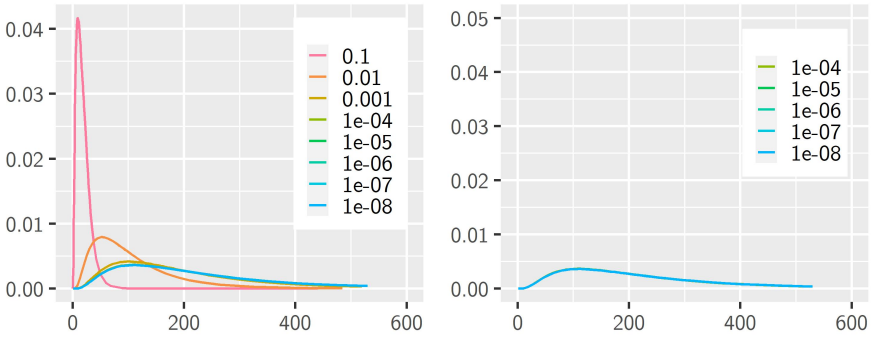
	Intercept	**PM10**	**JSA**	**Price**	ρ	τ_ν	τ_δ	τ_ϵ	DIC	WAIC
$\alpha = 0.1$	-0.7955	0.0164	0.0527	-0.1813	0.4593	0.0245	0.0883	0.0106	10374	10344
	(0.1453)	(0.0059)	(0.0051)	(0.0208)	(0.0480)	(0.0028)	(0.0819)	(0.0010)	10374	10345
$\alpha = 0.01$	-0.7986	0.0168	0.0520	-0.1818	0.4612	0.0246	0.0150	0.0106	10373	10344
	(0.0980)	(0.0058)	(0.0051)	(0.0208)	(0.0480)	(0.0028)	(0.0126)	(0.0010)	10373	10344
$\alpha = 0.001$	-0.8017	0.0172	0.0513	-0.1822	0.4633	0.0247	0.0078	0.0105	10373	10344
	(0.0906)	(0.0058)	(0.0051)	(0.0208)	(0.0481)	(0.0029)	(0.0065)	(0.0011)	10373	10344
$\alpha = 1e\text{-}04$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344
$\alpha = 1e\text{-}05$	-0.8025	0.0173	0.0511	-0.1824	0.4639	0.0248	0.0070	0.0105	10374	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0061)	(0.0010)	10374	10345
$\alpha = 1e\text{-}06$	-0.8026	0.0173	0.0511	-0.1824	0.4639	0.0248	0.0070	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0061)	(0.0010)	10373	10344
$\alpha = 1e\text{-}07$	-0.8026	0.0173	0.0511	-0.1824	0.4639	0.0248	0.0070	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0061)	(0.0010)	10373	10344
$\alpha = 1e\text{-}08$	-0.8026	0.0173	0.0511	-0.1824	0.4639	0.0248	0.0070	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0061)	(0.0010)	10373	10344

Table 4.6: Posterior means for parameter estimates together with standard deviations, DIC and WAIC values, for the spatio-temporal conditional Poisson model in the analysis of the respiratory hospital admissions in Glasgow data set with different prior distributions for the precision parameter of the random effect τ_ϵ.

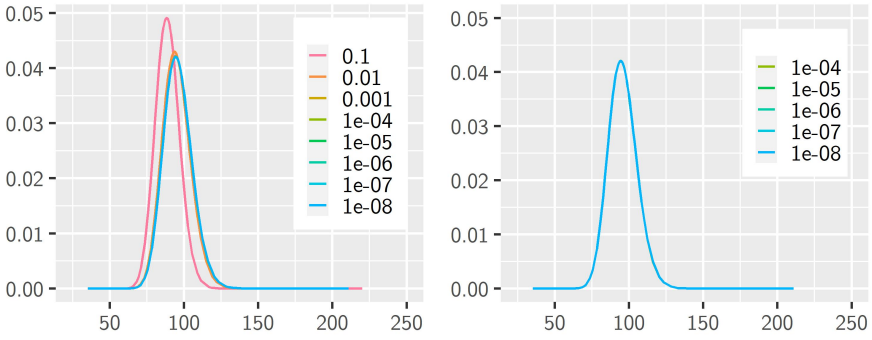
	Intercept	**PM10**	**JSA**	**Price**	ρ	τ_ν	τ_δ	τ_ϵ	DIC	WAIC
$\alpha = 0.1$	-0.8036	0.0173	0.0516	-0.1831	0.4629	0.0245	0.0082	0.0113	10373	10325
	(0.0903)	(0.0058)	(0.0052)	(0.0210)	(0.0486)	(0.0029)	(0.0081)	(0.0010)	10373	10342
$\alpha = 0.01$	-0.8028	0.0173	0.0512	-0.1825	0.4638	0.0247	0.0080	0.0106	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0076)	(0.0010)	10373	10344
$\alpha = 0.001$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0899)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344
$\alpha = 1e\text{-}04$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344
$\alpha = 1e\text{-}05$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10325
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10342
$\alpha = 1e\text{-}06$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344
$\alpha = 1e\text{-}07$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344
$\alpha = 1e\text{-}08$	-0.8025	0.0173	0.0511	-0.1824	0.4638	0.0247	0.0071	0.0105	10373	10344
	(0.0898)	(0.0058)	(0.0051)	(0.0209)	(0.0481)	(0.0029)	(0.0062)	(0.0010)	10373	10344



(a) Posterior marginal distributions for the precision parameter $\psi = 1/\tau_\nu$, the inverse of the variance parameter τ_ν



(b) Posterior marginal distributions for the precision parameter $\psi = 1/\tau_\delta$, the inverse of the variance parameter τ_δ



(c) Posterior marginal distributions for the precision parameter $\psi = 1/\tau_\epsilon$, the inverse of the variance parameter τ_ϵ

Figure 4.5: Posterior marginal distributions for the precision parameters $\psi_\nu = 1/\tau_\nu$, $\psi_\delta = 1/\tau_\delta$ and $\psi_\epsilon = 1/\tau_\epsilon$, for different values of α, where $\psi_{(.)} \sim G(\alpha, \alpha)$.

4.4.4 Fitting of the temporally varying spatial lag coefficient models

In order to illustrate the temporally varying spatial lag coefficient model proposed in Section 4.3.1, we have selected the best fitting spatio-temporal conditional model obtained so far (i.e., model with linear predictor in equation (4.15)), and included the spatial lag with the temporally varying coefficient in this regression structure. Specifically, the model that will be fitted is the following:

$$\begin{aligned}\log(\mu_{ij}) = &\log(E_{ij}) + \beta_0 + \beta_1 \text{JSA}_{ij} + \beta_2 \text{Price}_{ij} + \beta_3 \text{PM10}_{ij} + (\rho_0 + \rho_j)\mathbf{W}_i\mathbf{SIR}_j \\ &+ \nu_i + \delta_j + \epsilon_{ij}\end{aligned} \tag{4.16}$$

Here, we have considered the three alternatives for the specification of ρ_j described in Section 4.3.1. First, we have specified an autoregressive model of order one, that is, $\rho_j \sim N(\omega\rho_{j-1}, \tau_\rho)$ for $j = 2, \ldots, J$, and $\rho_1 \sim N(0, \tau_\rho)$, with ω being the autoregressive parameter and $\tau_\rho > 0$. Then, a random walk process of order one was considered, so that $\rho_j \sim N(\rho_{j-1}, \tau_\rho)$ for $j = 2, \ldots, J$, and $\rho_1 \sim N(0, \tau_\rho)$, with $\tau_\rho > 0$. Finally, we have also specified an unstructured normal distribution where $\rho_j \sim N(0, \tau_\rho)$, with $\tau_\rho > 0$. Note that ν_i and δ_j are as before.

For each of these specifications, we have fitted two models, one that includes the temporally unstructured random effect δ_j and another one where it was excluded. We have decided to do this in order to be able to compare the models which take into account the temporal correlation only by means of the random coefficient ρ_j for the spatial lag, with models that, besides including this coefficient, also take into account the remaining temporal dependence that could still be present. The results obtained after fitting the six models have been included in Table 4.7.

We should mention here that the prior distribution that will be assigned to the precision parameter $\psi_\rho = 1/\tau_\rho$ is a gamma with a large variance (i.e., $\psi_\rho \sim G(\text{1e-04}, \text{1e-04})$). In addition, in the case of the AR1 process, for the autoregressive parameter ω we will consider a normal prior with a large variance (i.e., $\omega \sim N(0, \text{1e-05})$).

The information criteria values obtained are very similar for all the models, moving from DIC = 10371 to DIC = 10373 and from WAIC = 10341 to WAIC = 10343, values which are also very close to the ones obtained for the spatio-temporal conditional models shown in Table 4.3. Moreover, these values show a very slight improvement when compared to the best fitting model chosen in equation (4.15), which were DIC = 10373 and WAIC = 10344.

With regard to the fixed effect ρ_0, for all the models fitted here, it is statistically significant and its coefficient has a positive estimated mean of approximately $\hat{\rho} = 0.43$, capturing the positive spatial autocorrelation present in the whole time period. The values obtained for ρ_j represent the variation of this spatial correlation on each time unit, which can be better assessed in the plots shown in Figure 4.6. Each of these corresponds to one of the fitted models, where the red line represents the estimated mean obtained for ρ_j, according to the year, and the green bands correspond to its 95% credible interval. Here, we can see how the effect of the spatial lag over the response, which is the number of respiratory hospital admissions, changes with time.

We notice that the credible interval bands are narrower for the models that do not include the temporal random effect δ_j. For these models, we can see that the estimated mean of ρ_j has the largest value for the year 2008 and then, it decreases from that year on. This suggests that in this year is where the strongest spatial autocorrelation is found in the data and, that it becomes weaker for the following years. Only for the year 2009, the effect is nearly zero, meaning that in this year, the spatial dependence is well explained by the fixed parameter ρ_0.

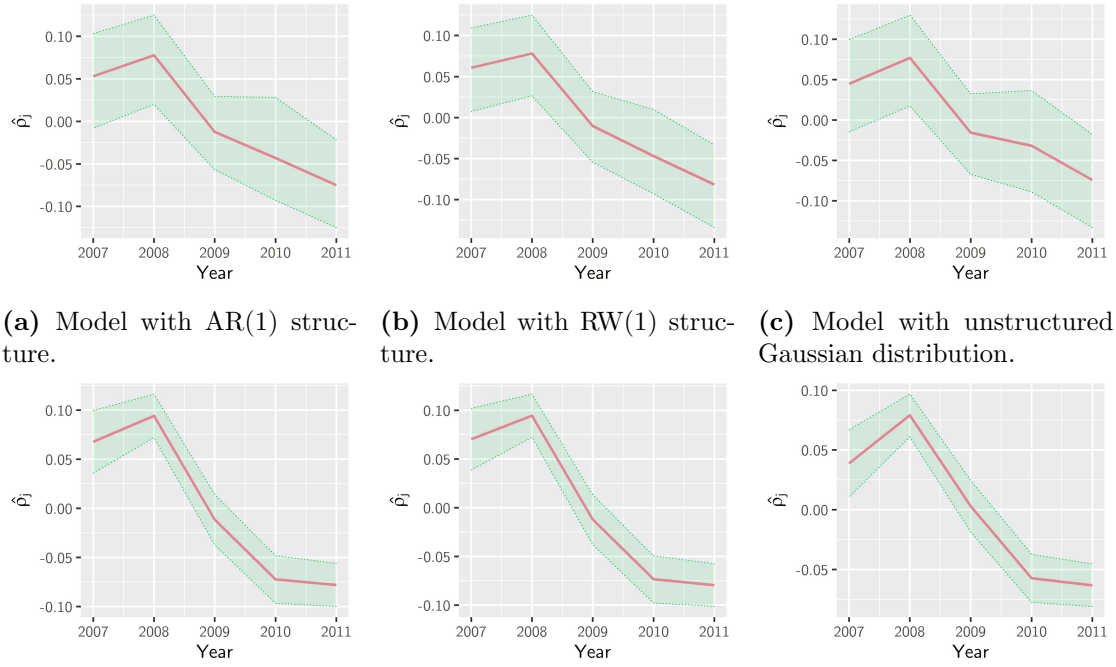
In addition, for the model where ρ_j follows the AR1 process, and the temporal effect is not included, we can see that the autoregressive parameter ω is statistically significant. This indicates a dependence of this term on its past value, which would be properly captured by this effect.

Let us recall that the information criteria values for each of the six fitted models are almost identical, and that the values obtained for the coefficient ρ_j are also very similar. In addition, we have also seen that including the temporal effect does not offer any improvements in model's fitting and that, when including it, the credible intervals for the estimated mean of the temporally varying coefficient for the spatial lag become wider. Therefore, we will consider the model where the simplest structure is specified for ρ_j, that is, where $\rho_j \sim N(0, \tau_\rho)$, $\tau_\rho > 0$, for $j = 1, \ldots, J$, and, where the temporal random effect δ_j is excluded from the linear predictor as the best fitting model for this data set.

In any case, by considering a temporally varying coefficient for the spatial lag term, we have taken a step further in the spatio-temporal conditional model proposals. This has allowed us to have valuable information about the variation of the strength of the spatial autocorrelation according to each time period under study.

Table 4.7: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the temporally varying spatial lag coefficient model, fitted to the respiratory hospital admissions in Glasgow data set.

		AR1 for ρ_j		RW1 for ρ_j		Unstructured Gaussian for ρ_j	
Intercept	Mean	-0.8028	-0.8090	-0.7993	-0.8071	-0.8023	-0.8091
	SD	(0.0835)	(0.0804)	(0.0829)	(0.0802)	(0.0838)	(0.0807)
	95% CI	(-0.9666,-0.6386)	(-0.9670,-0.6512)	(-0.9618,-0.6363)	(-0.9647,-0.6498)	(-0.9667,-0.6376)	(-0.9678,-0.6507)
PM10	Mean	0.0176	0.0183	0.0172	0.0179	0.0176	0.0183
	SD	(0.0058)	(0.0057)	(0.0058)	(0.0056)	(0.0058)	(0.0057)
	95% CI	(0.0062,0.0289)	(0.0071,0.0294)	(0.0059,0.0285)	(0.0068,0.0290)	(0.0061,0.0290)	(0.0071,0.0296)
JSA	Mean	0.0552	0.0554	0.0561	0.0561	0.0554	0.0556
	SD	(0.0054)	(0.0053)	(0.0053)	(0.0053)	(0.0054)	(0.0054)
	95% CI	(0.0445,0.0658)	(0.0450,0.0658)	(0.0455,0.0664)	(0.0456,0.0664)	(0.0447,0.0659)	(0.0450,0.0660)
Price	Mean	-0.1796	-0.1795	-0.1791	-0.1790	-0.1797	-0.1798
	SD	(0.0209)	(0.0208)	(0.0208)	(0.0208)	(0.0208)	(0.0208)
	95% CI	(-0.2204,-0.1386)	(-0.2203,-0.1386)	(-0.2199,-0.1382)	(-0.2198,-0.1381)	(-0.2205,-0.1389)	(-0.2206,-0.1389)
ρ_0	Mean	0.4348	0.4313	0.4319	0.4299	0.4345	0.4303
	SD	(0.0491)	(0.0486)	(0.0488)	(0.0485)	(0.0491)	(0.0486)
	95% CI	(0.3385,0.5314)	(0.3360,0.5268)	(0.3361,0.5278)	(0.3347,0.5252)	(0.3384,0.5310)	(0.3349,0.5258)
ω	Mean	0.3777	0.7008	-	-	-	-
	SD	(0.4072)	(0.2041)	-	-	-	-
	95% CI	(-0.5868,0.9027)	(0.1959,0.9634)	-	-	-	-
τ_ρ	Mean	0.0077	0.0206	0.0086	0.0102	0.0093	0.0110
	SD	(0.0080)	(0.0302)	(0.0119)	(0.0113)	(0.0136)	(0.0102)
	95% CI	(5.705e-04,0.0283)	(0.0023,0.0919)	(5.502e-04,0.0373)	(0.0016,0.0392)	(3.614e-04,0.0422)	(0.0020,0.0384)
τ_ν	Mean	0.0243	0.0242	0.0242	0.0241	0.0242	0.0241
	SD	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0028)
	95% CI	(0.0191,0.0301)	(0.0190,0.0300)	(0.0192,0.0301)	(0.0191,0.0299)	(0.0192,0.0300)	(0.0192,0.0300)
τ_δ	Mean	0.0023	-	0.0015	-	0.0020	-
	SD	(0.0096)	-	(0.0039)	-	(0.0029)	-
	95% CI	(-7.365e-05,0.0159)	-	(4.747e-06,0.0075)	-	(8.294e-05,0.0089)	-
τ_ϵ	Mean	0.0105	0.0105	0.0106	0.0106	0.0105	0.0106
	SD	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
	95% CI	(0.0086,0.0126)	(0.0086,0.0127)	(0.0086,0.0127)	(0.0087,0.0127)	(0.0086,0.0127)	(0.0087,0.0127)
		DIC = 10372 WAIC = 10342	DIC = 10372 WAIC = 10342	DIC = 10372 WAIC = 10341	DIC = 10372 WAIC = 10342	DIC = 10373 WAIC = 10343	DIC = 10371 WAIC = 10342



(a) Model with AR(1) structure.

(b) Model with RW(1) structure.

(c) Model with unstructured Gaussian distribution.

(d) Model with AR(1) structure, excluding δ_j.

(e) Model with RW(1) structure, excluding δ_j.

(f) Model with unstructured Gaussian distribution, excluding δ_j.

Figure 4.6: Estimated mean and its credible interval, obtained for ρ_j after fitting the considered temporally varying spatial lag coefficient models.

4.5 Application to low birth weight in Georgia

The data set we will analyse in this section has been obtained from the Online Analytical Statistical Information System (OASIS) database of the Georgia Department of Public Health (DPH). It consists on the number of live births on each of the $n = 159$ counties in the state of Georgia, USA, for $J = 10$ years, from 2010 up to 2019 (i.e., variable NB) and on the counts of infants with low birth weight (i.e., variable LBW) in these counties, for the same time period. Table 4.8 includes some descriptive statistics for these two variables for the different years under study.

Birth weight is recorded for infants within the first hours after birth and it is considered to be low when it is less than 2500 grams. A low weight at birth is associated with an increased risk of death of these neonates, when compared to those who present a normal birth weight. In addition, it has also been related to the development of a large number of health conditions and chronic diseases (Cutland et al., 2017).

Table 4.8: Descriptive statistics for the variables available in the study of low birth weight in Georgia.

	NB					LBW				
Year	Median	Mean	SD	Min.	Max.	Median	Mean	SD	Min.	Max.
2010	273.0	840.7	1859.9	14.0	12912.0	26.0	82.1	177.5	1.0	1383.0
2011	272.0	831.7	1850.9	20.0	12928.0	26.0	78.1	173.8	3.0	1370.0
2012	272.0	818.3	1816.8	17.0	12622.0	25.0	76.6	169.2	1.0	1263.0
2013	260.0	808.2	1799.1	19.0	12371.0	25.0	76.5	171.5	2.0	1288.0
2014	268.0	822.5	1845.5	11.0	12732.0	27.0	78.0	176.4	1.0	1393.0
2015	269.0	826.0	1856.8	14.0	12593.0	23.0	78.5	176.7	0.0	1298.0
2016	257.0	817.2	1824.9	13.0	12251.0	25.0	80.0	183.5	1.0	1343.0
2017	254.0	812.3	1826.9	13.0	12277.0	26.0	80.5	181.0	1.0	1319.0
2018	259.0	792.8	1752.1	15.0	11668.0	25.0	80.1	181.6	1.0	1339.0
2019	259.0	794.0	1742.4	14.0	11662.0	26.0	79.6	175.6	0.0	1223.0

Similar versions of this data set have been previously studied in the literature. Lawson (2008) carried out a study of very low birth weight (i.e., less than 1500 grams) in the counties of Georgia from 1994 to 2004, fitting a binomial model including temporal and spatial terms with different specifications. Additionally, Blangiardo et al. (2013) fitted a Poisson disease mapping model to low birth weight in these same regions, from 2000 to 2010. In these two applications, the authors found evidence of spatio-temporal correlation in the data sets under study. We believe it is important to mention that here, we analyse a more recent database, as compared to other applications where data from previous years have been studied. For example, this is the case of the data set analysed by Blangiardo et al. (2013).

In Figure 4.7, we can examine the spatial distribution of the number of low birth weight births in the counties of Georgia, from 2010 up to 2019 (i.e., variable LBW). We can see that there is a concentration of large values of this variable in the northern regions of the state, corresponding to Atlanta city and its surroundings. This pattern seems to be constant in time, as very little variation is observed across years.

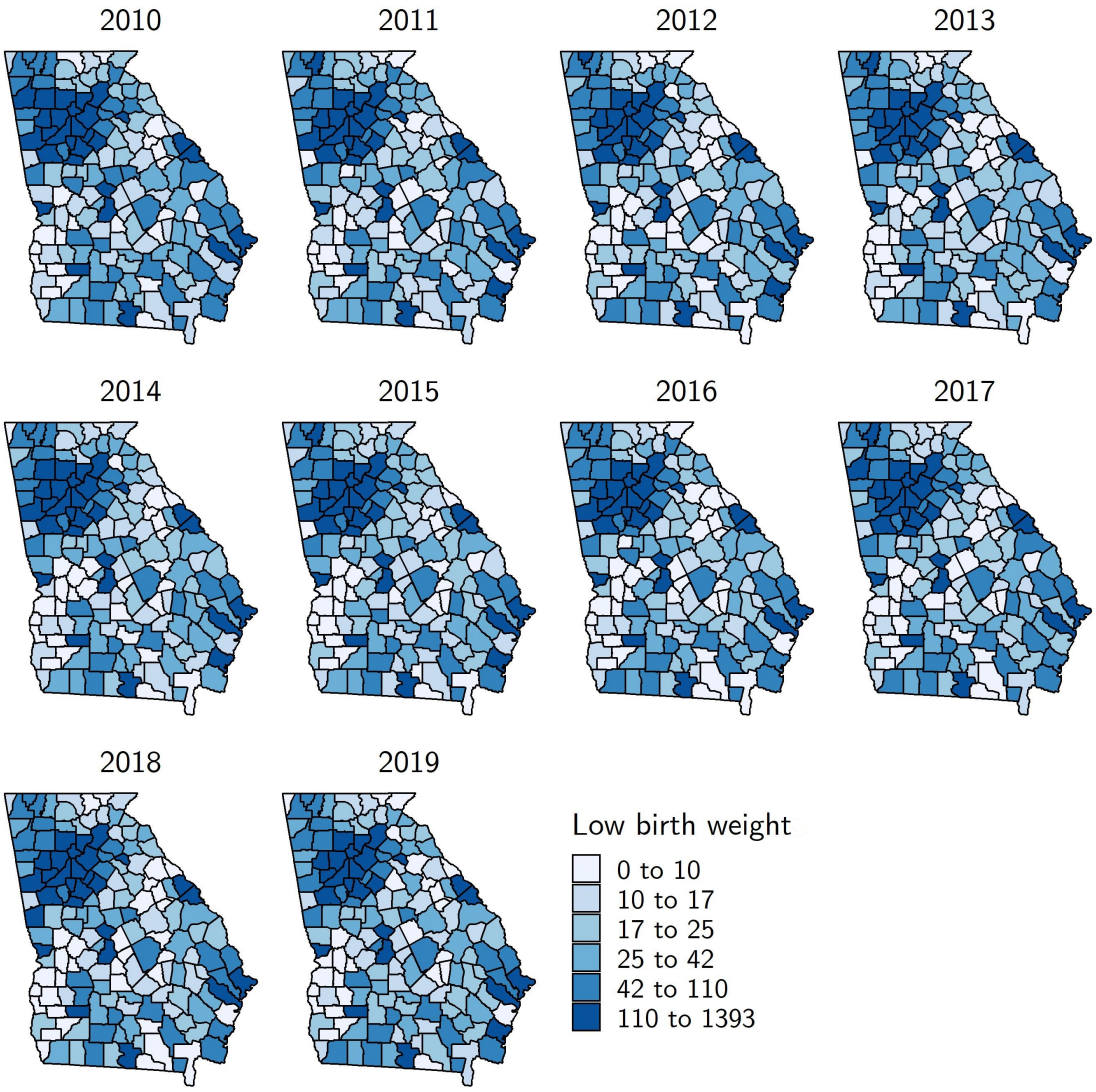


Figure 4.7: Spatial distribution of the low birth weight (LBW) in Georgia by counties and year.

We can also compute the proportion of the number of low weight births over the total of births in each county and year, generating the spatial distribution that can be observed in the maps in Figure 4.8. Here, we can see a different pattern from the one we observed in the total counts, as proportions seem to be more disseminated in space. However, from the year 2016 and on, there seems to be a certain concentration of high values in the southwestern regions of the state.

Therefore, in this section we will study this data set by fitting the spatio-temporal conditional models proposed in Section 4.3. Additionally, we will compare their performance with the Knorr-Held models, discussed in Section 4.2. Note that since there were

no other variables available to be included in the study as fixed effects, we will only take into account spatial terms, and the variability will be explained with structured and unstructured random effects.

All the models will be fitted in R-INLA, where, for the parameters, we will assume the same noninformative prior distributions considered in the previous section for the analysis of the respiratory hospital admissions in Glasgow. That is, normal priors with large variance (i.e., $N(0, 1\text{e-}05)$) for the fixed effects and vague gamma priors (i.e., $G(1\text{e-}04, 1\text{e-}04)$) for the precision parameter of the random effects. The appropriateness of the values chosen for the priors of the precisions will be assessed in a sensitivity analysis, which will be included in Section 4.5.3.

In addition, we compared the results obtained after fitting the models for different spatial weights matrices and found no considerable differences. Therefore, the spatial weights matrix $\mathbf{W}$ we will assume for all models is the one that follows contiguity of order one criterion.

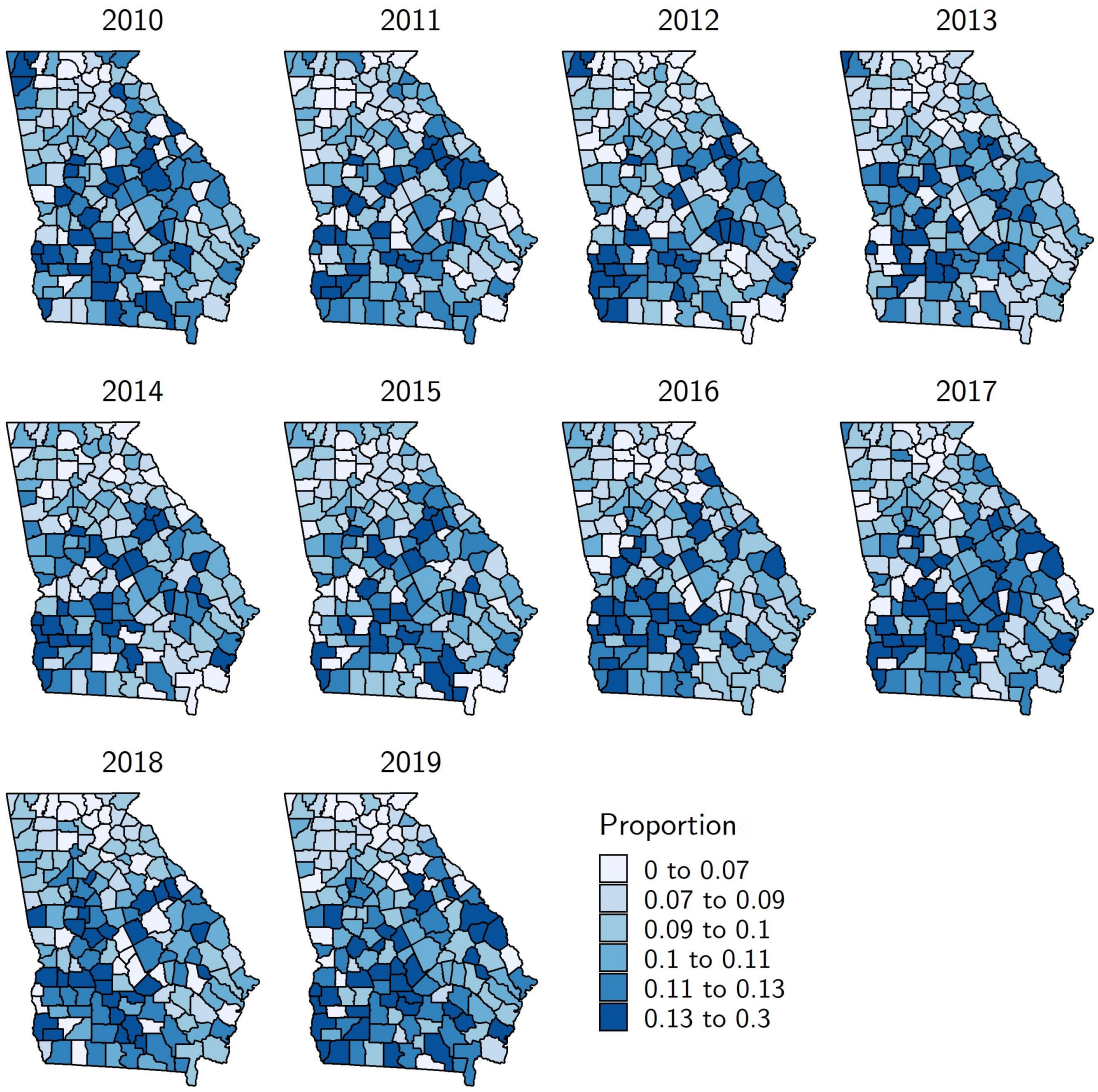


Figure 4.8: Spatial distribution of the proportion of low birth weight (LBW) in Georgia by counties and year.

4.5.1 Fitting of the spatio-temporal models

We assume that the variable representing the counts of low birth weight (i.e., variable LBW) follows a binomial distribution, where the number of births is the number of trials on each region and time period (i.e., variable NB). That is, $\text{LBW}_{ij} \sim \text{Bin}(\text{NB}_{ij}, \pi_{ij})$, for $i = 1, \ldots, n$, and $j = 1, \ldots, J$, with π_{ij} being the probability of success of a trial on each region and time period.

In order to take into account the spatio-temporal autocorrelation, we consider the Knorr-Held (2000) model in equation (4.5). In these models, we assume that the variables

LBW_{ij}, conditioned on the random effects $\nu_i, \eta_i, \delta_j, \phi_j$ and ϵ_{ij}, follow a binomial distribution, that is $(\text{LBW}_{ij}|\nu_i, \eta_i, \delta_j, \phi_j, \epsilon_{ij}) \sim \text{Bin}(n_{ij}, \pi_{ij})$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$, with π_{ij} modelled so that:

$$\text{logit}(\pi_{ij}) = \beta + \nu_i + \eta_i + \delta_j + \phi_j + \epsilon_{ij}, \tag{4.17}$$

where β is an intercept to be estimated, ν_i is an unstructured spatial random effect (i.e., $\nu_i \sim N(0, \tau_\nu)$, $\tau_\nu > 0$) and η_i is an ICAR distributed random effect, (i.e., $\eta_i \sim \text{ICAR}(0, \tau_\eta)$, $\tau_\eta > 0$). In addition, δ_j is an unstructured temporal random effect (i.e., $\delta_j \sim N(0, \tau_\delta)$, $\tau_\delta > 0$) and ϕ_j is a random effect following a random walk process of order one, that is $\phi_j \sim N(\phi_{j-1}, \tau_\phi)$, for $j = 2, \ldots, J$ and $\phi_1 \sim N(0, \tau_\phi)$, with $\tau_\phi > 0$. Finally, ϵ_{ij} is an unstructured interaction effect (i.e., $\epsilon_{ij} \sim N(0, \tau_\epsilon)$, $\tau_\epsilon > 0$).

The results obtained from the fitting of this model and some of its reduced versions have been included in Table 4.9. Here, the lowest information criteria values were obtained for the model in equation (4.17), the one that includes all the random effects (i.e., DIC = 10294 and WAIC = 10325) and for the model excluding the temporally structured effect ϕ_j (i.e., DIC = 10295 and WAIC = 10324). We will consider the simpler model, that does not include ϕ_j, as the best fitting model for further analysis. In particular, in this model, the regression structure specified for π_{ij} is:

$$\text{logit}(\pi_{ij}) = \beta + \nu_i + \eta_i + \delta_j + \epsilon_{ij}, \tag{4.18}$$

where β, ν_i, η_i, δ_j and ϵ_{ij} are as before.

Additionally, in Figure 4.9 we can see the maps of the estimated proportions obtained after fitting the model in equation (4.18). Here we can clearly observe a concentration of high values in the southwestern regions, which is similarly observed in all the years under study. When comparing these estimated proportions with the observed ones, in Figure 4.8, we can see some noticeable differences, mostly given by the spatial concentration of high values found in the predicted proportions, which is not that strongly appreciated in the observed ones.

Table 4.9: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the spatio-temporal binomial model and some of its reduced versions, fitted to the low birth weight in Georgia data set.

		Intercept	τ_ν	τ_η	τ_δ	τ_ϕ	τ_ϵ
DIC = 10500	Mean	-2.2103	0.0413	-	-	-	-
WAIC = 10566	SD	(0.0172)	(0.0053)	-	-	-	-
	95% CI	(-2.2441,-2.1764)	(0.0320,0.0529)	-	-	-	-
DIC = 10491	Mean	-2.2032	0.0115	0.0615	-	-	-
WAIC = 10555	SD	(0.0102)	(0.0041)	(0.0214)	-	-	-
	95% CI	(-2.2237,-2.1832)	(0.0051,0.0210)	(0.0318,0.1146)	-	-	-
DIC = 10394	Mean	-2.2033	0.0123	0.0602	0.0012	-	-
WAIC = 10453	SD	(0.0154)	(0.0049)	(0.0226)	(6.595e-04)	-	-
	95% CI	(-2.2339,-2.1730)	(0.0052,0.0241)	(0.0282,0.1159)	(4.226e-04,0.0029)	-	-
DIC = 10393	Mean	-2.2033	0.0117	0.0614	2.372e-04	5.757e-04	-
WAIC = 10450	SD	(0.0115)	(0.0041)	(0.0209)	(2.417e-04)	(4.762e-04)	-
	95% CI	(-2.2262,-2.1807)	(0.0053,0.0213)	(0.0316,0.1129)	(3.785e-05,8.633e-04)	(1.107e-04,0.0018)	-
DIC = 10294	Mean	-2.2046	0.0116	0.0603	2.658e-04	5.848e-04	0.0033
WAIC = 10325	SD	(0.0117)	(0.0041)	(0.0214)	(2.997e-04)	(5.113e-04)	(7.536e-04)
	95% CI	(-2.2280,-2.1816)	(0.0052,0.0212)	(0.0304,0.1134)	(3.170e-05,0.0010)	(9.468e-05,0.0020)	(0.0020,0.0050)
DIC = 10295	Mean	-2.2047	0.0114	0.0616	9.853e-04	-	0.0034
WAIC = 10324	SD	(0.0146)	(0.0044)	(0.0225)	(5.756e-04)	-	(7.637e-04)
	95% CI	(-2.2336,-2.1759)	(0.0049,0.0218)	(0.0302,0.1172)	(2.894e-04,0.0025)	-	(0.0021,0.0051)
DIC = 10307	Mean	-2.2051	0.0114	0.0605	-	-	0.0048
WAIC = 10331	SD	(0.0106)	(0.0041)	(0.0212)	-	-	(8.368e-04)
	95% CI	(-2.2263,-2.1843)	(0.0051,0.0208)	(0.0310,0.1130)	-	-	(0.0033,0.0066)
DIC = 10303	Mean	-2.2116	0.0412	-	2.721e-04	6.185e-04	0.0033
WAIC = 10335	SD	(0.0180)	(0.0053)	-	(2.992e-04)	(6.084e-04)	(7.548e-04)
	95% CI	(-2.2469,-2.1762)	(0.0318,0.0527)	-	(2.653e-05,0.0010)	(8.588e-05,0.0022)	(0.0020,0.0050)
DIC = 10303	Mean	-2.2116	0.0412	-	0.0010	-	0.0034
WAIC = 10334	SD	(0.0199)	(0.0053)	-	(6.414e-04)	-	(7.648e-04)
	95% CI	(-2.2508,-2.1724)	(0.0318,0.0526)	-	(2.846e-04,0.0027)	-	(0.0021,0.0051)
DIC = 10316	Mean	-2.2120	0.0409	-	-	-	0.0048
WAIC = 10342	SD	(0.0172)	(0.0053)	-	-	-	(8.381e-04)
	95% CI	(-2.2458,-2.1781)	(0.0316,0.0524)	-	-	-	(0.0033,0.0066)
DIC = 10676	Mean	-2.2301	-	-	-	-	0.0476
WAIC = 10606	SD	(0.0074)	-	-	-	-	(0.0029)
	95% CI	(-2.2447,-2.2155)	-	-	-	-	(0.0421,0.0536)

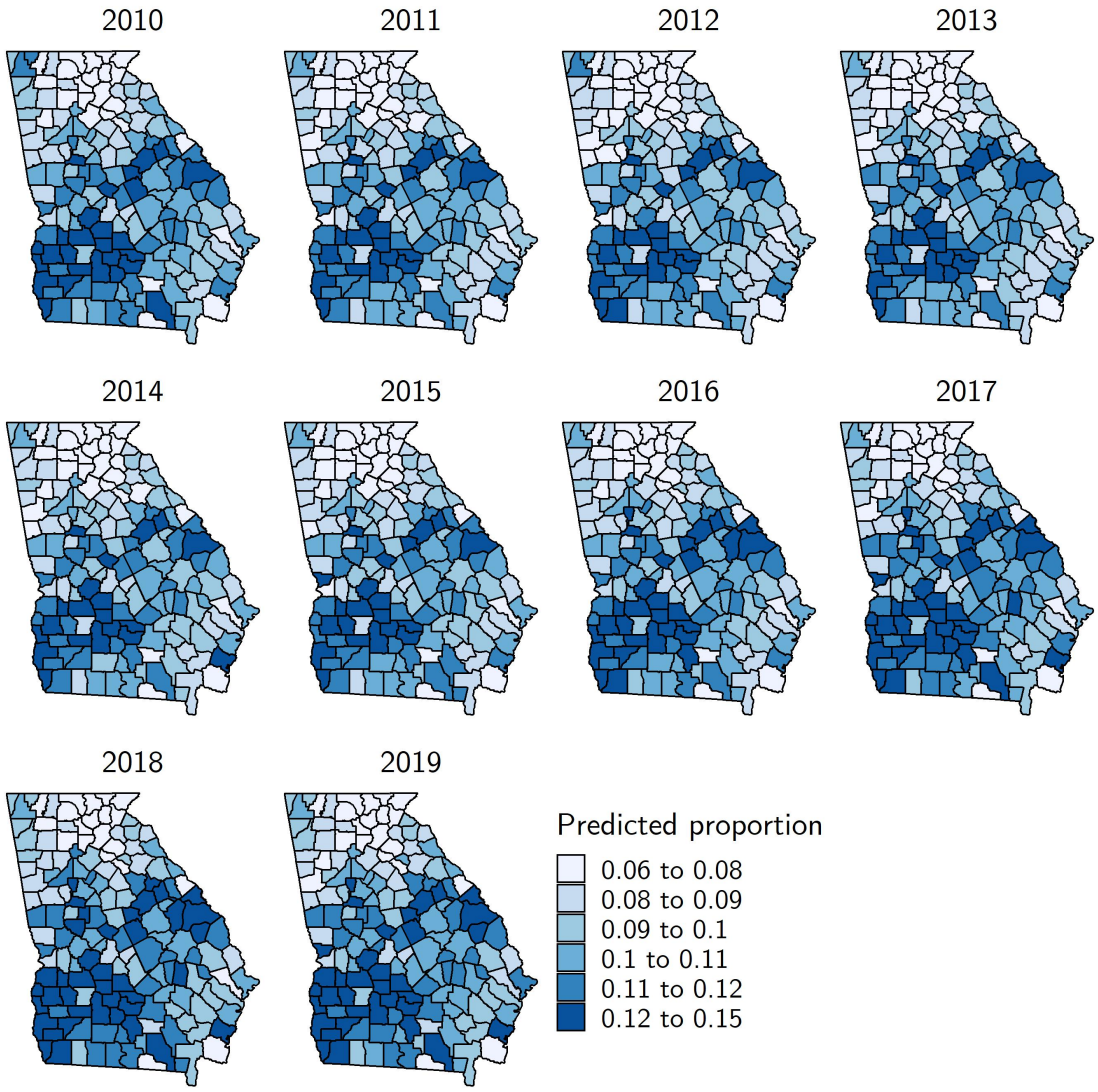


Figure 4.9: Maps for the predicted proportions, for the different years under study, obtained from the spatio-temporal model in equation (4.18), fitted to the low birth weights in Georgia data set.

4.5.2 Fitting of the spatio-temporal conditional models

In this section, we will fit the proposed spatio-temporal conditional models for binomial responses, described in Section 4.3, to the low birth weight data in Georgia. Therefore, we assume that the variables LBW_{ij}, conditioned on the values of all the neighbours of the i-th region, but not including the i-th region itself, and on the random effects ν_i, δ_j, ϕ_j and ϵ_{ij}, follow a binomial distribution, that is $(\text{LBW}_{ij}|\text{LBW}_{\sim ij}, \nu_i, \delta_j, \phi_j, \epsilon_{ij}) \sim \text{Bin}(n_{ij}, \pi_{ij})$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. For the probability of success π_{ij}, we

specify a regression structure in the following way:

$$\text{logit}(\pi_{ij}) = \beta + \rho A_{ij} + \nu_i + \delta_j + \phi_j + \epsilon_{ij}, \tag{4.19}$$

where A_{ij} are the elements of the spatial term $\mathbf{A}$, which is an $n \times J$ matrix with columns $\mathbf{A}_j$, corresponding to the $n \times 1$ vector of spatial terms for the j-th time period. Each of the elements of this vector is defined so that $A_{ij} = \frac{\hat{\pi}_{\sim ij}}{1-\hat{\pi}_{\sim ij}}$, with $\hat{\pi}_{\sim ij} = \frac{\mathbf{W}_i\mathbf{LBW}_j}{\mathbf{W}_i\mathbf{NB}_j}$. Here, $\mathbf{LBW}_j$ and $\mathbf{NB}_j$ are the vectors of observations for the j-th time period for the variables LBW and NB, respectively. Note that ν_i, δ_j, ϕ_j and ϵ_{ij} are as before.

The results obtained after fitting the model in equation (4.19) and some of its reduced versions are shown in Table 4.10. Regarding the spatial parameter ρ, its estimated coefficient is positive and statistically significant for all the fitted models, indicating that there is positive spatial autocorrelation, which is being explained by this term.

If we compare these results with the ones obtained for the spatio-temporal models included in Table 4.9, we can see that they offer a similar fit, in terms of information criteria. For example, for the spatio-temporal model in equation (4.17) these values were DIC = 10294 and WAIC = 10325 and, for the conditional model in equation (4.19), they are DIC = 10296 and WAIC = 10325.

The model with the smallest DIC and WAIC values was the one where all the random effects were included (i.e., DIC = 10296 and WAIC = 10325) and, also, for the one where the structured temporal effect was excluded (i.e., DIC = 10296 and WAIC = 10324). Hence, we will consider the simpler model as the best fitting one, where the probability of success follows the regression structure:

$$\text{logit}(\pi_{ij}) = \beta + \rho A_{ij} + \nu_i + \delta_j + \epsilon_{ij} \tag{4.20}$$

In the maps of the estimated proportions obtained after fitting the model in equation (4.20), included in Figure 4.10, we can see the same concentration of high values in the southwestern regions that was observed in Figure 4.9, for the spatio-temporal model in equation (4.18). Therefore, the predictive accuracy of both models appears to be quite similar.

Table 4.10: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the spatio-temporal conditional binomial model and some of its reduced versions, fitted to the low birth weight in Georgia data set.

		Intercept	ρ	τ_ν	τ_δ	τ_ϕ	τ_ϵ
DIC = 13963	Mean	-2.5010	2.5048	-	-	-	-
WAIC = 13986	SD	(0.0207)	(0.1916)	-	-	-	-
	95% CI	(-2.5417,-2.4603)	(2.1285,2.8806)	-	-	-	-
DIC = 10471	Mean	-2.4985	2.5575	0.0345	-	-	-
WAIC = 10534	SD	(0.0425)	(0.3494)	(0.0045)	-	-	-
	95% CI	(-2.5816,-2.4150)	(1.8703,3.2417)	(0.0266,0.0444)	-	-	-
DIC = 10407	Mean	-2.3785	1.4925	0.0370	9.590e-04	-	-
WAIC = 10469	SD	(0.0475)	(0.3862)	(0.0049)	(5.749e-04)	-	-
	95% CI	(-2.4712,-2.2846)	(0.7325,2.2486)	(0.0284,0.0476)	(2.931e-04,0.0025)	-	-
DIC = 10404	Mean	-2.3733	1.4466	0.0372	1.919e-04	4.196e-04	-
WAIC = 10465	SD	(0.0465)	(0.3841)	(0.0049)	(1.951e-04)	(3.618e-04)	-
	95% CI	(-2.4641,-2.2817)	(0.6902,2.1980)	(0.0284,0.0477)	(3.111e-05,7.123e-04)	(7.500e-05,0.0014)	-
DIC = 10296	Mean	-2.4511	2.1241	0.0351	2.165e-04	3.345e-04	0.0037
WAIC = 10325	SD	(0.0527)	(0.4429)	(0.0047)	(2.425e-04)	(3.589e-04)	(7.841e-04)
	95% CI	(-2.5543,-2.3476)	(1.2544,2.9928)	(0.0269,0.0453)	(3.129e-05,8.460e-04)	(2.801e-05,0.0013)	(0.0023,0.0054)
DIC = 10296	Mean	-2.4638	2.2361	0.0350	5.503e-04	-	0.0039
WAIC = 10324	SD	(0.0536)	(0.4481)	(0.0047)	(4.086e-04)	-	(7.955e-04)
	95% CI	(-2.5685,-2.3583)	(1.3559,3.1140)	(0.0266,0.0449)	(1.118e-04,0.0016)	-	(0.0025,0.0056)
DIC = 10302	Mean	-2.5221	2.7524	0.0338	-	-	0.0046
WAIC = 10327	SD	(0.0499)	(0.4194)	(0.0045)	-	-	(8.248e-04)
	95% CI	(-2.6199,-2.4241)	(1.9283,3.5749)	(0.0258,0.0435)	-	-	(0.0031,0.0064)
DIC = 10605	Mean	-2.7028	4.2769	-	-	-	0.0422
WAIC = 10520	SD	(0.0405)	(0.3596)	-	-	-	(0.0026)
	95% CI	(-2.7824,-2.6235)	(3.5714,4.9828)	-	-	-	(0.0373,0.0475)

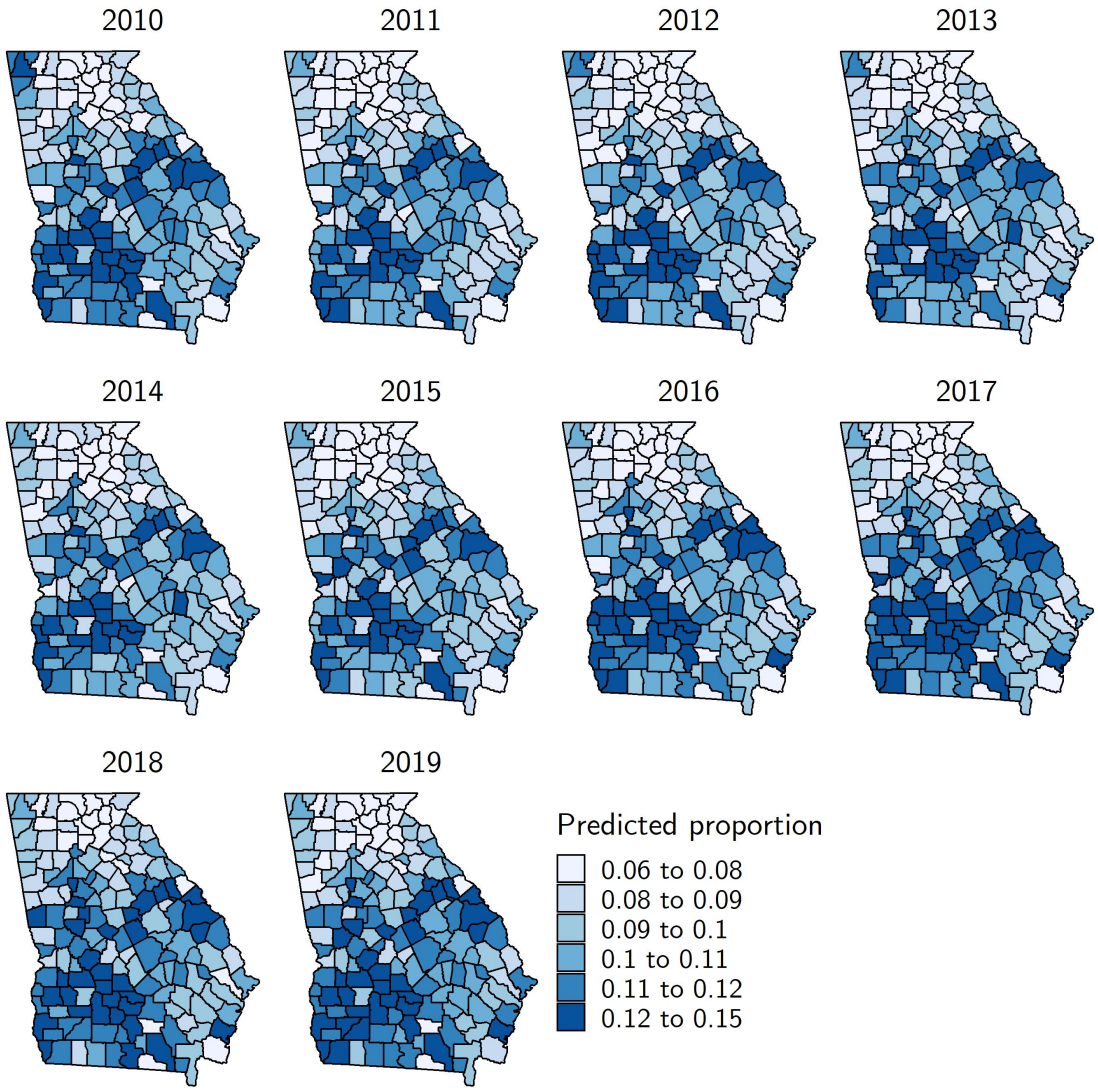


Figure 4.10: Maps for the predicted proportions, for the different years under study, obtained from the spatio-temporal conditional model in equation (4.20), fitted to the low birth weights in Georgia data set.

4.5.3 Sensitivity analysis for the precision of the prior distributions

We have performed a sensitivity analysis in order to properly assess the values specified for the prior distributions of the precision parameters estimated in the models fitted in this study. For brevity of exposition, in this section we only show the results obtained for the model in equation (4.20).

In Tables 4.11, 4.12 and 4.13 we have included the estimates, DIC and WAIC values obtained when fitting the models corresponding to different specifications of α in $\psi_{(.)} \sim$

$G(\alpha, \alpha)$ for the precisions $\psi_\nu = 1/\tau_\nu$, $\psi_\delta = 1/\tau_\delta$, and $\psi_\epsilon = 1/\tau_\epsilon$, respectively. For each of these parameters, the values considered for α were from $\alpha = 0.1$ to $\alpha =$ 1e-08. In each of these three tables we can see some small differences in the estimates obtained. In general, for $\psi_\nu = 1/\tau_\nu$, $\psi_\delta = 1/\tau_\delta$, and $\psi_\epsilon = 1/\tau_\epsilon$, from the value $\alpha =$ 1e-04, differences in the third decimal place are observed for a considerable number of cases, whereas variations in the second decimal place only appear in a few number of cases.

Additionally, in Figure 4.11 we can see the marginal posterior densities of the precision parameters for the different values of α. In particular, in Figure 4.11(b), we can see that for the precision parameter $\psi_\delta = 1/\tau_\delta$ there are some slight differences among the marginal posteriors from the value $\alpha =$ 1e-04 and on. Such differences are not observed for the other parameters analysed and, in any case, they do not seem to represent any major stability issues from the value $\alpha =$ 1e-04 and on. Therefore, we believe that it is a reasonable and well justified choice.

Table 4.11: Posterior means for parameter estimates together with standard deviations, DIC and WAIC values, for the spatio-temporal conditional binomial model in the analysis of low birth weight in Georgia data set with different prior distributions for the precision parameter of the random effect τ_ν.

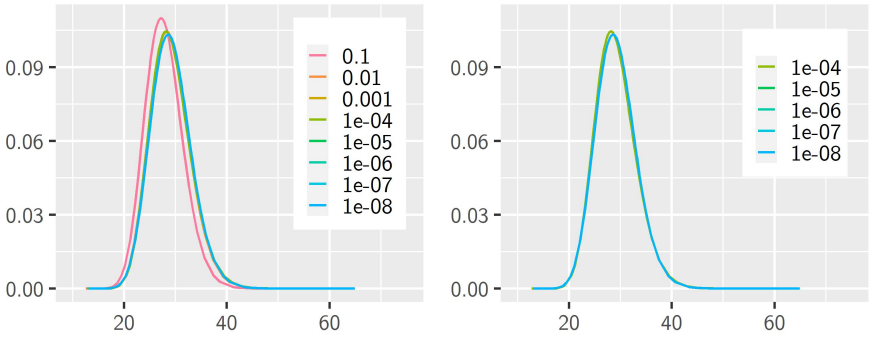
	Intercept	ρ	τ_ν	τ_δ	τ_ϵ	DIC	WAIC
$\alpha = 0.1$	-2.4609	2.2108	0.0366	5.437e-04	0.0038	10296	10324
	(0.0538)	(0.4498)	(0.0048)	(4.239e-04)	(8.022e-04)	10296	10324
$\alpha = 0.01$	-2.4630	2.2295	0.0351	5.592e-04	0.0038	10296	10324
	(0.0537)	(0.4491)	(0.0047)	(4.081e-04)	(8.002e-04)	10296	10324
$\alpha = 0.001$	-2.4637	2.2352	0.0350	5.501e-04	0.0038	10296	10324
	(0.0535)	(0.4479)	(0.0047)	(4.060e-04)	(7.962e-04)	10296	10324
$\alpha =$ 1e-04	-2.4638	2.2361	0.0350	5.503e-04	0.0039	10296	10324
	(0.0536)	(0.4481)	(0.0047)	(4.086e-04)	(7.955e-04)	10296	10324
$\alpha =$ 1e-05	-2.4636	2.2346	0.0349	5.744e-04	0.0038	10296	10324
	(0.0536)	(0.4484)	(0.0047)	(4.265e-04)	(7.959e-04)	10296	10324
$\alpha =$ 1e-06	-2.4636	2.2346	0.0349	5.743e-04	0.0038	10296	10324
	(0.0536)	(0.4484)	(0.0047)	(4.264e-04)	(7.959e-04)	10296	10324
$\alpha =$ 1e-07	-2.4636	2.2346	0.0349	5.744e-04	0.0038	10296	10324
	(0.0536)	(0.4484)	(0.0047)	(4.265e-04)	(7.960e-04)	10296	10324
$\alpha =$ 1e-08	-2.4636	2.2346	0.0349	5.744e-04	0.0038	10296	10324
	(0.0536)	(0.4484)	(0.0047)	(4.265e-04)	(7.959e-04)	10296	10324

Table 4.12: Posterior means for parameter estimates together with standard deviations, DIC and WAIC values, for the spatio-temporal conditional binomial model in the analysis of low birth weight in Georgia data set with different prior distributions for the precision parameter of the random effect τ_δ.

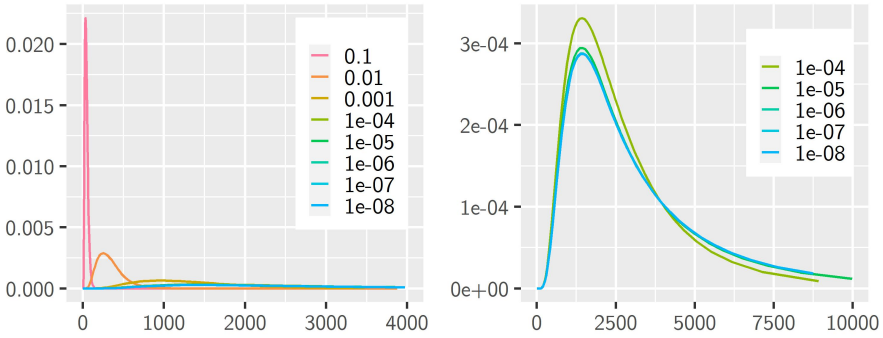
	Intercept	ρ	τ_ν	τ_δ	τ_ϵ	DIC	WAIC
$\alpha = 0.1$	-2.4388	2.0140	0.0354	0.0281	0.0037	10298	10326
	(0.0732)	(0.4429)	(0.0047)	(0.0146)	(7.889e-04)	10297	10326
$\alpha = 0.01$	-2.4432	2.0538	0.0353	0.0037	0.0037	10296	10325
	(0.0555)	(0.4419)	(0.0047)	(0.0019)	(7.895e-04)	10296	10324
$\alpha = 0.001$	-2.4553	2.1614	0.0351	9.367e-04	0.0038	10296	10324
	(0.0533)	(0.4428)	(0.0047)	(5.690e-04)	(7.937e-04)	10296	10324
$\alpha =$ 1e-04	-2.4638	2.2361	0.0350	5.503e-04	0.0039	10296	10324
	(0.0536)	(0.4481)	(0.0047)	(4.086e-04)	(7.955e-04)	10296	10324
$\alpha =$ 1e-05	-2.4658	2.2539	0.0349	5.047e-04	0.0038	10296	10324
	(0.0539)	(0.4517)	(0.0047)	(4.054e-04)	(7.974e-04)	10296	10324
$\alpha =$ 1e-06	-2.4661	2.2564	0.0349	4.964e-04	0.0038	10296	10324
	(0.0540)	(0.4523)	(0.0047)	(4.032e-04)	(7.977e-04)	10296	10324
$\alpha =$ 1e-07	-2.4661	2.2567	0.0349	4.954e-04	0.0038	10296	10324
	(0.0540)	(0.4523)	(0.0047)	(4.028e-04)	(7.976e-04)	10296	10324
$\alpha =$ 1e-08	-2.4661	2.2567	0.0349	4.955e-04	0.0038	10296	10324
	(0.0540)	(0.4523)	(0.0047)	(4.028e-04)	(7.976e-04)	10296	10324

Table 4.13: Posterior means for parameter estimates together with standard deviations, DIC and WAIC values, for the spatio-temporal conditional binomial model in the analysis of low birth weight in Georgia data set with different prior distributions for the precision parameter of the random effect τ_ϵ.

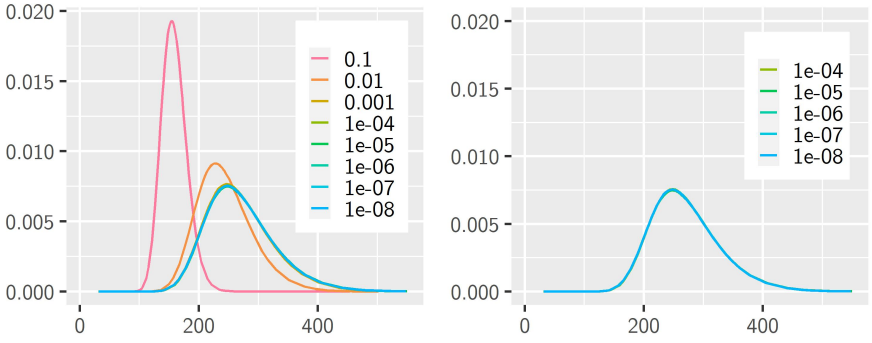
	Intercept	ρ	τ_ν	τ_δ	τ_ϵ	DIC	WAIC
$\alpha = 0.1$	-2.4891	2.4532	0.0341	4.532e-04	0.0064	10292	10289
	(0.0551)	(0.4635)	(0.0046)	(3.828e-04)	(8.512e-04)	10295	10317
$\alpha = 0.01$	-2.4680	2.2725	0.0348	5.384e-04	0.0042	10296	10324
	(0.0538)	(0.4509)	(0.0047)	(3.986e-04)	(7.954e-04)	10296	10324
$\alpha = 0.001$	-2.4642	2.2394	0.0349	5.568e-04	0.0039	10296	10324
	(0.0536)	(0.4488)	(0.0047)	(4.139e-04)	(7.974e-04)	10296	10324
$\alpha =$ 1e-04	-2.4638	2.2361	0.0350	5.503e-04	0.0039	10296	10324
	(0.0536)	(0.4481)	(0.0047)	(4.086e-04)	(7.955e-04)	10296	10324
$\alpha =$ 1e-05	-2.4644	2.2420	0.0349	5.371e-04	0.0039	10296	10324
	(0.0535)	(0.4481)	(0.0047)	(4.120e-04)	(8.068e-04)	10296	10324
$\alpha =$ 1e-06	-2.4647	2.2442	0.0348	5.370e-04	0.0039	10296	10324
	(0.0535)	(0.4481)	(0.0047)	(4.263e-04)	(7.995e-04)	10296	10324
$\alpha =$ 1e-07	-2.4647	2.2444	0.0348	5.350e-04	0.0039	10296	10324
	(0.0536)	(0.4483)	(0.0047)	(4.249e-04)	(8.019e-04)	10296	10324
$\alpha =$ 1e-08	-2.4647	2.2444	0.0348	5.348e-04	0.0039	10296	10324
	(0.0536)	(0.4483)	(0.0047)	(4.248e-04)	(8.024e-04)	10296	10324



(a) Posterior marginal distributions for the precision parameter $\psi = 1/\tau_\nu$, the inverse of the variance parameter τ_ν



(b) Posterior marginal distributions for the precision parameter $\psi = 1/\tau_\delta$, the inverse of the variance parameter τ_δ



(c) Posterior marginal distributions for the precision parameter $\psi = 1/\tau_\epsilon$, the inverse of the variance parameter τ_ϵ

Figure 4.11: Posterior marginal distributions for the precision parameters $\psi_\nu = 1/\tau_\nu$, $\psi_\delta = 1/\tau_\delta$ and $\psi_\epsilon = 1/\tau_\epsilon$, for different values of α, where $\psi_{(.)} \sim G(\alpha, \alpha)$.

4.5.4 Fitting of the temporally varying spatial lag coefficient models

In this section, we will fit the proposed temporally varying spatial lag coefficient model, described in Section 4.3.1, to the low birth weight data in Georgia. In order to do this, we will consider the model in equation (4.20), which is the best fitting spatio-temporal conditional model obtained so far, and specify the following regression structure for π_{ij}:

$$\text{logit}(\pi_{ij}) = \beta_0 + (\rho_0 + \rho_j)A_{ij} + \nu_i + \delta_j + \epsilon_{ij} \tag{4.21}$$

For the random coefficient ρ_j, we will specify different alternatives, which are the same ones considered in Section 4.4.4, for the temporally varying coefficient spatial conditional models fitted to the data corresponding to the respiratory hospital admissions in Glasgow. These are, an autoregressive model of order one (i.e., $\rho_j \sim N(\omega\rho_{j-1}, \tau_\rho)$ for $j = 2, \ldots, J$, and $\rho_1 \sim N(0, \tau_\rho)$, with ω being the autoregressive parameter and $\tau_\rho > 0$), a random walk process of order one (i.e., $\rho_j \sim N(\rho_{j-1}, \tau_\rho)$ for $j = 2, \ldots, J$, and $\rho_1 \sim N(0, \tau_\rho)$, with $\tau_\rho > 0$) and an unstructured Gaussian process (i.e., $\rho_j \sim N(0, \tau_\rho)$, with $\tau_\rho > 0$, for $j = 1, \ldots, J$).

In Table 4.14 we have included the results obtained after fitting the model in equation (4.21) for the three alternatives considered for ρ_j. Here, we have fitted two models for each of the three specifications; one where we include the temporal unstructured random effect δ_j and another one where we exclude it. The DIC and WAIC values obtained are very similar for the six models and also, very close to the values obtained for the spatio-temporal conditional model in equation (4.20).

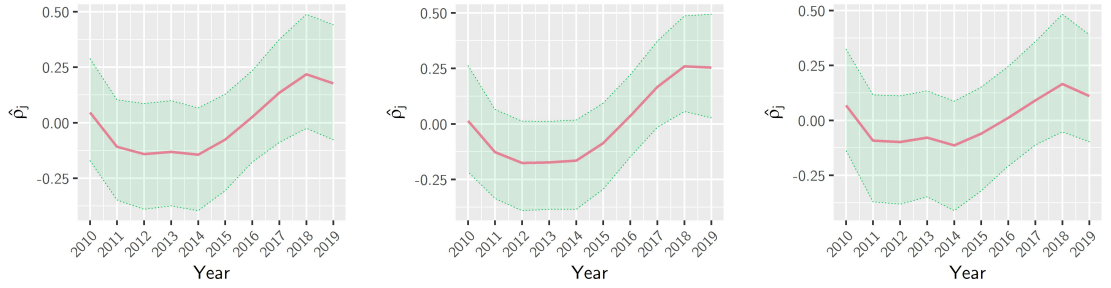
The fixed effect ρ_0 resulted statistically significant and the value of its coefficient was approximately equal to two for all the models. Moreover, the estimates obtained for ρ_j, which represent the temporal variation of the spatial correlation with respect to the one implied by the spatial parameter ρ_0, can be better assessed by examining Figure 4.12. In these plots, corresponding to each one of the six fitted models, the red curve represents the estimated mean value of ρ_j and the green bands, its 95% credible intervals, for the years considered.

Here, we can observe the temporal variation of this coefficient, which behaves in a very similar way in all the six models considered. In general, we see how the value decreases from the year 2011 and it begins to increase from 2014. Then, it slightly decreases again from 2018. This suggests that the strongest spatial dependence in the data is found for the year 2018 and the weakest, approximately, between the years 2011 and 2014. For the years 2010 and 2016 the effect is close to zero, which suggests that the estimated value obtained for the global spatial parameter ρ_0 would properly explain the spatial correlation in these cases.

In addition, note that for the model corresponding to the AR1 process for ρ_j and δ_j is excluded, we have obtained a significant value for the autoregressive parameter ω, indicating that this term might depend on its past value, being properly captured by this effect.

Table 4.14: Parameter estimates, standard deviations and 95% credible intervals in parenthesis for the parameters in the models, and DIC and WAIC values for the temporally varying spatial lag coefficient model, fitted to the low birth weight in Georgia data set.

		AR1 for ρ_j		RW1 for ρ_j		Unstructured Gaussian for ρ_j	
Intercept	Mean	-2.4461	-2.4480	-2.4400	-2.4420	-2.4523	-2.4521
	SD	(0.0535)	(0.0533)	(0.0533)	(0.0531)	(0.0537)	(0.0541)
	95% CI	(-2.5507,-2.3406)	(-2.5524,-2.3431)	(-2.5444,-2.3353)	(-2.5462,-2.3376)	(-2.5575,-2.3466)	(-2.5580,-2.3459)
ρ_0	Mean	2.0763	2.0925	2.0218	2.0403	2.1318	2.1286
	SD	(0.4510)	(0.4512)	(0.4490)	(0.4491)	(0.4519)	(0.4582)
	95% CI	(1.1881,2.9580)	(1.2055,2.9760)	(1.1401,2.9024)	(1.1584,2.9207)	(1.2442,3.0178)	(1.2296,3.0267)
ω	Mean	0.4076	0.6371	-	-	-	-
	SD	(0.3590)	(0.2366)	-	-	-	-
	95% CI	(-0.4595,0.8805)	(0.0617,0.9495)	-	-	-	-
τ_ρ	Mean	0.0444	0.0648	0.0272	0.0276	0.0227	0.0483
	SD	(0.0363)	(0.0641)	(0.0253)	(0.0222)	(0.0299)	(0.0359)
	95% CI	(0.0071,0.1408)	(0.0114,0.2347)	(0.0034,0.0935)	(0.0052,0.0868)	(3.147e-04,0.0997)	(0.0090,0.1423)
τ_ν	Mean	0.0353	0.0352	0.0353	0.0355	0.0350	0.0351
	SD	(0.0047)	(0.0047)	(0.0047)	(0.0048)	(0.0047)	(0.0047)
	95% CI	(0.0270,0.0455)	(0.0269,0.0454)	(0.0270,0.0456)	(0.0268,0.0455)	(0.0268,0.0452)	(0.0268,0.0453)
τ_δ	Mean	3.076e-04	-	2.034e-04	-	5.798e-04	-
	SD	(4.466e-04)	-	(1.986e-04)	-	(8.789e-04)	-
	95% CI	(2.333e-05,0.0014)	-	(3.472e-05,7.320e-04)	-	(4.087e-05,0.0026)	-
τ_ϵ	Mean	0.0036	0.0036	0.0036	0.0036	0.0038	0.0037
	SD	(7.829e-04)	(7.803e-04)	(7.774e-04)	(7.786e-04)	(7.926e-04)	(7.908e-04)
	95% CI	(0.0023,0.0053)	(0.0023,0.0053)	(0.0023,0.0053)	(0.0022,0.0053)	(0.0024,0.0055)	(0.0024,0.0055)
		DIC = 10296 WAIC = 10324	DIC = 10295 WAIC = 10323	DIC = 10294 WAIC = 10324	DIC = 10293 WAIC = 10323	DIC = 10295 WAIC = 10324	DIC = 10295 WAIC = 10323



(a) Model with AR(1) structure.

(b) Model with RW(1) structure.

(c) Model with unstructured Gaussian distribution.

(d) Model with AR(1) structure, excluding δ_j.

(e) Model with RW(1) structure, excluding δ_j.

(f) Model with unstructured Gaussian distribution, excluding δ_j.

Figure 4.12: Estimated mean and its credible interval, obtained for ρ_j after fitting the considered temporally varying spatial lag coefficient models.

In any case, let us recall that the DIC and WAIC values did not favour any of these six models in particular when compared among them, and that they gave very similar estimated values for the coefficients. This leads us to prefer the simpler specification, in this case given by the unstructured Gaussian process. In addition, although these models do not offer great improvement in terms of model fitting when compared to the spatio-temporal conditional binomial models, we believe we can obtain useful interpretation of the spatial parameters given by its variation in time.

4.6 Discussion

In this chapter, we have reviewed some of the most frequently applied regression models in the spatio-temporal count data literature. Moreover, we have proposed some extensions of the spatial conditional models of Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) that allow for the modelling of spatio-temporal data. These proposals have been fitted to real data examples and their performance has been compared with the widely applied spatio-temporal Knorr-Held (2000) model. Model comparison was carried out using the DIC and WAIC information criteria, and by performing posterior predictive

checks.

Our first proposal is a direct extension of the models of Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), also given by the inclusion of a spatial lag term in the regression structure for the conditional mean of the response variable under study. In this case, the spatial lag is computed for each time unit, assuming that the spatial structure, which is represented by a given spatial weights matrix, is the same for all the time units considered. We have denoted these proposals as spatio-temporal conditional models and they have been considered for Poisson and binomially distributed responses, although they could also be specified for any other count data distribution.

In addition, we have also proposed another extension derived from the already proposed spatio-temporal conditional models. Here, we define the spatial lag coefficient as the sum of two terms; one fixed parameter and a temporally structured random coefficient, for which we proposed different specifications. This allows the spatial term parameter to vary in the temporal dimension, leading to the temporally varying spatial lag coefficient model proposals.

The first application presented corresponds to the study of respiratory hospital admissions in the statistical sectors of Glasgow, from the year 2007 to 2011 (Lee, Rushworth and Napier, 2018). Here we assumed that the conditioned response follows a Poisson distribution and we fitted the spatio-temporal conditional models. In this case, lower information criteria values were obtained for the spatio-temporal conditional models, when compared to the spatio-temporal models of the Knorr-Held type, which were also fitted to the same data set. In addition, our results were consistent with those obtained by Lee, Rushworth and Napier (2018).

Furthermore, we have also fitted the proposed temporally varying spatial lag coefficient models to this data set, which also offered a better fit in terms of DIC and WAIC values, when compared to the previously fitted models. This proposal allowed us to identify the year 2008 as the one with the strongest spatial autocorrelation, and that this dependence became weaker from that year on.

The other application included the study of the low birth weight in the counties of Georgia, from 2010 to 2019, which corresponds to a more recent database than the ones analysed in other studies in the literature. The spatio-temporal and the conditional models were fitted to these data, assuming a binomial distribution for the response variable. These models offered similar results in terms of information criteria, as well as in terms of predictive accuracy. In this case, the temporally varying spatial lag coefficient models suggested that the spatial autocorrelation began to increase around the year 2012, and that it was the strongest for the year 2018.

In general terms, we have found that the proposed models were able to provide a similar, and in some specific cases, a better fit in terms of information criteria than some of the already existing models in the literature, such as the ones proposed by Knorr-Held (2000). In the two applications presented, both the spatio-temporal and the conditional models offered similar predictive accuracy. Moreover, the temporally varying spatial lag coefficient models allowed us to have a better idea of the behaviour and of the variation of the spatial autocorrelation across the time units under study.

Chapter 5

Spatial autoregressive modelling of COVID data: Assessment of weights matrix neighbourhood alternatives

5.1 Introduction

The analysis of spatial data has become widely spread in epidemiology, specially because location can be an important surrogate for lifestyle, environment, as well as genetic and other factors and, therefore, it can provide important insights for public health data analysis. In order to be able to take the possible existing spatial correlation into account, very common spatial regression models for count data make use of a spatially structured random effect, which is structured according to a given spatial weights matrix.

In this context, two of the most popular models in spatial disease mapping are the Besag-York-Mollié (BYM) model (Besag, York and Mollié, 1991) and the BYM2 model (Riebler et al., 2016). These models were described in Chapter 2, where they were applied to the analysis of infant mortality rates and to mother's postnatal period screening test in Colombia. Let us recall that the BYM model incorporates spatial dependence by means of two unobserved latent effects, namely a spatially unstructured random effect and a spatially structured random effect following an Intrinsic Conditional Autoregressive (ICAR) prior (Besag, 1974). In the BYM2 model the latent effect is a weighted average of these two random effects. Another random effects model frequently found in the literature is the Leroux model (Leroux, Lei and Breslow, 2000). These models are generally estimated using Bayesian inferential methods.

An alternative to these models is given by the spatial conditional overdispersion models, also denoted as spatial auto-regressive models (see Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018 and Morales-Otero and Núñez-Antón, 2021). These models include a spatial lag of the response variable in the regression model specification, which

allows to capture the spatial dependence on the spatial neighbouring regions. These were also described in Chapter 2 and applied to the same data sets mentioned before.

The conditionally autoregressive model (CAR) or auto-Poisson scheme was previously proposed by Besag (1974). However, in the way it was originally defined, there was the need to impose the restriction that the spatial autoregressive parameter is non positive, so that the conditional distribution exists. This is a consequence of the fact that, if the range of the response variable considered in the study is infinite, the spatial terms could cause the model to be explosive. Zeger and Qaqish (1988) and Held, Höhle and Hofmann (2005) proposed several models for analysing time series data. In the case of counts, they considered a Poisson model which included the logarithm of the past counts in the mean regression specification. These models are able to account for both positive and negative temporal autocorrelation. In this paper, we will elaborate on the latter models in the context of spatial epidemiology.

In the aforementioned models, the relationship between two regions is described by a spatial weights matrix, for which several different specifications have been developed (see Anselin, 2002). In most cases, this matrix is fixed and previously specified, a choice that may have an impact on the results of the analysis. Therefore, it is very important for researchers to be able to study how to best describe the spatial structure of the data. Traditionally, spatial weights matrices are based on the adjacency of regions or on the distance among regions. However, there may be situations where the association is not given by the geographical proximity but, instead, it depends on some other connectivity structure or even on the specific characteristics of the regions under study.

In this sense, Earnest et al. (2007) studied the influence of different specifications of spatial weights matrices on the smoothing properties of the CAR model. The authors fitted the models to a data set corresponding to birth defects in Australia, obtaining considerable differences in the results, which provided clear evidence about the importance of the proper choice of the spatial structure. Case, Hines and Rosen (1993) explored spatial autocorrelation in a data set corresponding to government expenditure in the USA from 1970 to 1985. The authors implemented different spatial weights matrices based on contiguity and distance and, in addition, they proposed the use of a similarity matrix based on the inverse of the difference of the values that a given covariate takes in each state. They concluded that the similarity matrix based on the variable that represents the percentage of population which is black considerably improved the performance of the model. Ejigu and Wencheko (2020) proposed a weights matrix that took into account geographical proximity and covariate information simultaneously. The authors fitted an autoregressive spatial model to the Meuse river heavy metals data set, and compared the performance of the proposed weights with other alternatives frequently used in the literature. They concluded that their proposed weights led to a better justification and motivation of the spatial structure present in the data under study.

After the beginning of the pandemic, a considerable number of research manuscripts about the spatial modelling of COVID-19 have been published. For example, D'Angelo, Abbruzzo and Adelfio (2021) studied its spread in the Northern Italian provinces from February to April 2020, corroborating the effectiveness of the implementation of the

lockdown in this country by applying a BYM model to the two periods from before and after the decree of the measure. In addition, the authors fitted spatio-temporal models for the number of cases in all the Italian provinces from February to October 2020, finding that the temporal evolution was independent of the spatial correlation. In addition, they were able to identify areas with the most elevated risk of infection. Note that no covariates were included in the study, since the authors did not intend to determine risk factors in the population.

Johnson, Ravi and Braneon (2021) studied the monthly number of cases and deaths related to COVID-19 between March and December 2020 in USA counties and their relation with social vulnerability of the population in each of the regions. They fitted spatio temporal Poisson models for each of these two outcomes, including socio-economic and environmental variables, and assumed a BYM2 model for the spatial effects. The authors characterized the spatio-temporal pattern of the spread of the disease and the fatalities across the country and identified risk factors in the population such as the proportion of non-white people and the proportion of people without a higher education level, among others.

Konstantinoudis et al. (2022) analysed the weekly number of deaths for several regions in Europe during the period going from 2015 to 2019. They fitted a hierarchical Poisson model to this data, where the spatial autocorrelation was captured by means of a BYM2 model. In this way, they were able to predict the deaths in 2020 and evaluate the excess of mortality in each of the regions for this year.

Natalia et al. (2022) studied the evolution of COVID-19 cases per two weeks in the municipalities of Belgium from June to December 2020. The authors used a spatio-temporal Poisson model with a BYM model specification to take into account the spatial dependence. As proxies for the socio-economic status, they used the mean income and the number of students having a higher education level. Their conclusions suggested that there was an increased incidence after the reopening of higher education facilities.

In this chapter, we propose an alternative to the auto-Poisson models that is able to account for positive spatial autocorrelation. It is a modification of the spatial conditional models in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) to account for the spatial autocorrelation that might be present in the data. The spatial conditional auto-Poisson model is described in Section 5.2.1, and the extension is motivated and introduced in Section 5.2.2. Additionally, we also investigate the use of several spatial weights matrices in the computation of the spatial lag and propose some new possible structures to be implemented, which are discussed in Section 5.2.3. An illustration of the methodology and comparison of the different models is provided in Section 5.3. In addition, a comparison with the BYM2 model is included in Section 5.4 and a simulation study is included in Section 5.5. We end with a discussion in Section 5.6.

We believe it is important to mention that the work presented in this chapter is a collaboration with Christel Faes and Vicente Núñez-Antón. It has been included in a paper which was submitted for publication and it is currently under review.

5.2 Methodology

This section reviews the spatial conditional overdispersion models proposed by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) (see Chapter 2 for a more detailed description of these models) in the context of epidemiological applications. Thereafter, we propose an extension of the autoregressive model and discuss possible weights matrices that could describe the underlying spatial dependency structure.

5.2.1 Review of the spatial conditional autoregressive model

The spatial conditional overdispersion models were developed to fit spatial count data, allowing to capture overdispersion and to explain the spatial dependence that may exist in the data, as suggested by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018). These authors assume that the dependent variable Y_i, for regions $i = 1, \ldots, n$, follows a conditional distribution $f(y_i \mid y_{\sim i})$, where y_i represents the observed count in region i and, $y_{\sim i}$, the values in all of the neighbouring regions of the i-th region (without including the i-th region itself). A spatial autoregressive term, more specifically, the lag of the response variable, is incorporated in the regression model specification for the conditional mean $\mathrm{E}(Y_i \mid Y_{\sim i})$. The inclusion of such spatial dependence in the model can explain part of the overdispersion.

In an epidemiological context, interest often goes towards the modelling of the rates of a disease. In this case, Morales-Otero and Núñez-Antón (2021) assumed that the conditioned response variable $(Y_i \mid Y_{\sim i}, \nu_i)$, the total number of cases for $i = 1, \ldots, n$, follows a Poisson distribution, with conditional mean μ_i so that $\mathrm{E}(Y_i|Y_{\sim i}, \nu_i) = \mu_i = P_i r_i$, with P_i being the population size and r_i representing the disease rate in the i-th region, for $i = 1, \ldots, n$. They proposed the following regression structure for the conditioned means:

$$\log(\mu_i) = \log(P_i) + \mathbf{x}_i^\top \boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{r} + \nu_i, \tag{5.1}$$

where an autoregressive component is included for the rates, i.e., $\mathbf{W}_i\mathbf{r} = \sum_{j=1}^{n} w_{ij} r_j$, which is a weighted average of the observed rates $r_i = y_i/P_i$, with weights specified by the spatial weights matrix $\mathbf{W}$. Here, $\mathbf{x}_i$ is a vector of explanatory variables for the i-th observation, $\boldsymbol{\beta}$ a vector of unknown regression parameters that need to be estimated and ρ the unknown spatial autoregressive parameter. In addition, a normally distributed random effect $\nu_i \sim N(0, \tau)$, with $\tau > 0$, is included to allow for additional unstructured overdispersion in the counts. Note that the assumed spatial structure is given by the matrix $\mathbf{W}$, where its elements, w_{ij}, are weights that represent the strength of the relationship between regions i and j. Section 5.2.3 includes a detailed description about the different ways these weights can be defined.

There is a certain similarity between the spatial conditional model considered in this work and the auto-Poisson model (Besag, 1974), with the disadvantage that the latter only allows for negative autocorrelation. We address this issue in the next section.

5.2.2 Geometric mean spatial conditional model

Zeger and Qaqish (1988) proposed several models to account for temporal autocorrelation in time series data, including one for count data, where they suggested the use of a Poisson model that incorporates the logarithm of the past counts in the regression model for the logarithm of the mean instead of the past counts. In particular, they assumed a Poisson distribution for the variables Y_j, representing counts for $j = 1, \ldots, J$ time periods, so that $Y_j \sim \text{Poi}(\mu_j)$, with means μ_j following the regression model:

$$\log(\mu_j) = \mathbf{x}_j^\top \boldsymbol{\beta} + \sum_{l=1}^{q} \theta_l [\log(y^*_{j-l}) - \mathbf{x}_{j-l}^\top \boldsymbol{\beta}], \tag{5.2}$$

where q represents all the previous counts and θ_l, for $l = 1, \ldots, q$, are unknown parameters that need to be estimated. To ensure the existence of the logarithm when there are zero counts, the authors propose the use of the term $y^*_{j-l} = \max(y_{j-l}, c)$, with $0 < c < 1$ a parameter that is assumed to be known. In addition, $\mathbf{x}_j$ and $\boldsymbol{\beta}$ are as before.

The authors also propose an alternative to the linear predictor in equation (5.2), where the model for the means would be:

$$\log(\mu_j) = \mathbf{x}_j^\top \boldsymbol{\beta} + \sum_{l=1}^{q} \theta_l \{\log(y_{j-l} + c) - \log[\exp(\mathbf{x}_{j-l}^\top \boldsymbol{\beta})]\}, \tag{5.3}$$

Knorr-Held and Richardson (2003) proposed the use the term $\log(y_{j-1} + 1)$ in order to overcome the issue of the nonexistence of the logarithm, so that it is equal to zero when there are no cases. Held, Höhle and Hofmann (2005) proposed to regress the mean directly on the past counts instead, but assuming an identity link.

Following the ideas in Zeger and Qaqish (1988) and Knorr-Held and Richardson (2003), we propose the following geometric mean spatial conditional model for count data. As before, we assume a Poisson model for the conditioned response outcomes, that is $(Y_i | Y_{\sim i}, \nu_i) \sim \text{Poi}(\mu_i)$, with conditional mean $\text{E}(Y_i | Y_{\sim i}, \nu_i) = \mu_i = P_i r_i$, following the regression model:

$$\log(\mu_i) = \log(P_i) + \mathbf{x}_i^\top \boldsymbol{\beta} + \rho \mathbf{W}_i \log(\mathbf{r}) + \nu_i \tag{5.4}$$

This model closely resembles the model in equation (5.1), but here the autoregressive component is a weighted average of the logarithms of the rates, instead of the rates. Zeger and Qaqish (1988) indicated that a model such as this does allow for positive autocorrelation, which is indeed common in spatial epidemiology. It can be easily seen that the smoothed estimates of the rates are estimated as:

$$\begin{aligned} \widehat{r_i} &= \exp(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}) \exp\left(\frac{1}{n_i} \sum_{j=1}^{n} w^*_{ij} \log(r_j) \right)^{\hat{\rho}} \exp(\nu_i) \\ &= \exp(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}) \overline{\mathbf{r}}_i^{\hat{\rho}} \exp(\nu_i), \end{aligned} \tag{5.5}$$

with w_{ij}^* representing the non-standardized spatial weights, n_i being the number of neighbours of region i, and $\bar{\bar{\mathbf{r}}}_i$ being the geometric mean of the rates included in the vector of rates $\mathbf{r}$. Note that the geometric mean of a sample $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$ is defined as $(\prod_{i=1}^{n} x_i)^{\frac{1}{n}}$, which can also be expressed as $\exp\left[\frac{1}{n}\sum_{i=1}^{n}\log(x_i)\right]$, when $x_i > 0$, for $i = 1, \ldots, n$.

Here, the estimated value obtained for the spatial parameter ρ would represent how the incidence rate in one region resembles the geometric mean of the rates in its neighbours. Therefore, the use of the logarithm of the rates in the autoregressive component has an important epidemiological interpretation.

5.2.3 Spatial weights matrices

This section discusses different possible choices for specifying the weights w_{ij} used in the proposed model in equation (5.4).

Spatial weights matrices based on contiguity

The spatial structure based on contiguity or adjacency is defined by the spatial weights matrix $\mathbf{W}$, where $w_{ij} = 1$, if region i is adjacent or a neighbour to region j, and $w_{ij} = 0$, otherwise. Different criteria can be assumed to specify whether two regions are adjacent, for example, the Queen contiguity criterion assumes that regions i and j are neighbours if they share at least one point in their boundaries. Most commonly the spatial weights matrix is standardized by rows, so that if region i is adjacent to region j, then $w_{ij} = 1/n_i$, where n_i is the number of neighbours region i has. In this way, the spatial lag $\mathbf{W}_i\mathbf{y}$ can be viewed as a spatial average of the values that the variable takes in all of its neighbouring locations.

First order contiguity is specified when we consider that regions i and j are neighbours if they share at least one point in their boundaries. Extending this criteria by considering that i and j are neighbours if they share a common neighbour, we can define second order contiguity. Third order contiguity can be specified the same way, when it is assumed that regions i and j are adjacent if they share a common neighbour of order two. Contiguity of higher order is also possible to specify by following these ideas.

Spatial weights matrices based on distance

An alternative way to define a spatial structure is to consider a spatial weights matrix where its elements are defined as a function of the distance among the central points of the polygons representing the regions, called the centroids, s_i $(i = 1, \ldots, n)$. Inverse distance weights are specified as $w_{ij} = 1/\|s_i - s_j\|$, with $\|s_i - s_j\|$ being the Euclidean distance between regions i and j. In addition, in the negative exponential criteria the weights are defined so that $w_{ij} = \exp(-\|s_i - s_j\|)$.

Finally, we can also define the distance band weights, with band width given by a critical threshold h. In particular, it is considered that regions i and j are neighbours if their centroid lies within the chosen band. Let s_i be the centroids of the regions under

study, for a given threshold h, then $w_{ij} = 1$ if the Euclidean distance between s_i and s_j is smaller than h, that is $\|s_i - s_j\| < h$, and $w_{ij} = 0$ otherwise.

Covariate-based similarity (or difference) matrices

Ejigu and Wencheko (2020) proposed a weights matrix $\mathbf{W}$, which not only takes into account geographical proximity, but also a specific covariate's information. Given an environmental variable e_i, for $i = 1, \ldots, n$, regions with centroids s_i, they define the following structure for the weights:

$$w_{ij} = \exp\{-[\alpha|e_i - e_j| + (1-\alpha)\|s_i - s_j\|]\}, \tag{5.6}$$

where α is a previously fixed chosen value between zero and one, $|e_i - e_j|$ is the absolute difference in the value of the environmental covariate between regions i and j and $\|s_i - s_j\|$ is the Euclidean distance between the centroids of regions i and j. The elements in the diagonal of this matrix are zero and it is row standardized. As α approaches zero, the weights give more relevance to the geographical distance, and, when it approaches one, the covariate differences receive more importance.

Following this idea, we also propose an alternative covariate-based similarity matrix, where we will consider both environmental and socio-economic variables to impact the weight among regions. Let $\mathbf{W}$ be a traditional weights matrix based on contiguity, distance, or any other criteria, with elements w_{ij}, and $\mathbf{D}$ an $n \times n$ matrix with elements $d_{ij} = 0$ if $i = j$ and:

$$d_{ij} = \exp(-|e_i - e_j|), \text{ for } i \neq j, \tag{5.7}$$

We then propose the use of the matrix $\mathbf{W} \circ \mathbf{D}$, which is the Hadamard (or element-wise) product of matrices $\mathbf{W}$ and $\mathbf{D}$. In this way, small weights are given to neighbouring regions with large differences in the values of the covariate and to distant regions, while large weights are given to neighbouring regions with similar covariate information and that are geographically close to each other.

Mobility matrix

The previous proposals presented here for the weights matrices are a representation of how close (in space) and/or how similar (in terms of covariate information) regions are. Another characteristic to define the weights matrix is to assess how much contact there was among individuals in the different areas. This is of special interest when considering, for example, an outcome that depends on the contact behaviour, such as is the case in infectious disease incidence. As a proxy for the contact behaviour, and based on mobile phone data (Ensoy-Musoro et al., 2022), the mobility among regions can be used. That is, each element m_{ij} in the mobility matrix $\mathbf{M}$ is defined as the mean proportion of time that people from region i have spent in region j in a given time period. This matrix would then clearly represent a different type of connectivity structure among regions.

5.2.4 Review of the BYM2 model

In order to offer some comparison of the proposed methods with other models employed in the disease mapping literature, we will also consider the fitting of the BYM2 model (Riebler et al., 2016). In this model, it is assumed that $(Y_i \mid \nu_i, \eta_i) \sim \text{Poi}(\mu_i)$, with conditional mean $\text{E}(Y_i \mid \nu_i, \eta_i) = \mu_i = P_i r_i$ following the regression structure:

$$\log(\mu_i) = \log(P_i) + \mathbf{x}_i^\top \boldsymbol{\beta} + \frac{1}{\sqrt{\tau_s}} \left(\sqrt{1-\phi_s}\nu_i + \sqrt{\phi_s}\eta_i \right), \tag{5.8}$$

where ν_i and η_i are unstructured normally and intrinsic conditionally autoregressive (ICAR) distributed random effects, respectively, but with variance scaled to approximately one. In addition, τ_s is a precision parameter that controls for the variance contribution from the sum of the two random effects and ϕ_s is a mixing parameter that captures the proportion of the variance explained by the spatially structured random effect. Note that $1 - \phi_s$ represents the proportion of the variance explained by the unstructured random effect.

In this model, the spatial neighbourhood structure that is usually assumed is the one based on contiguity of first order. Although some other structure might be specified, this model requires the spatial matrix to be symmetrical (Wall, 2004). In addition, for the parameters τ_s and ϕ_s, penalized complexity priors are generally assumed (Simpson et al., 2017).

Comparison of the spatial conditional model with the BYM and the BYM2 models has been previously performed by Morales-Otero and Núñez-Antón (2021), research already included in Chapter 2. Their results showed that, when compared to the spatial conditional model, they offered a similar fit in terms of information criteria. However, the BYM and BYM2 models did not provide additional information about the type and strength of spatial autocorrelation that was present in the data.

5.2.5 Model estimation and selection

All models considered here are fitted using the integrated nested Laplace approximation (INLA) approach, in the R-INLA package. It should be noted, however, that, in general, any software methodology that allows for estimation of a generalized linear mixed model can be used to implement this model. This is a great advantage of the proposed method, as one is not restricted to complex estimation tools for fitting spatial models. Model comparison is carried out by using the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) and the Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010), where their smallest values indicate the best fitting model.

5.3 Illustration of methodology

5.3.1 Data Exploration

We investigate the spatial distribution of COVID-19 from September 2020 until January 2021 among the Flemish municipalities. Figure 5.1 shows the observed incidence of

COVID-19 per 100,000 inhabitants in Flanders' municipalities in the time period considered, which was the time of the second wave in Belgium. It can be observed that not all areas presented the same impact in the second COVID-19 wave.

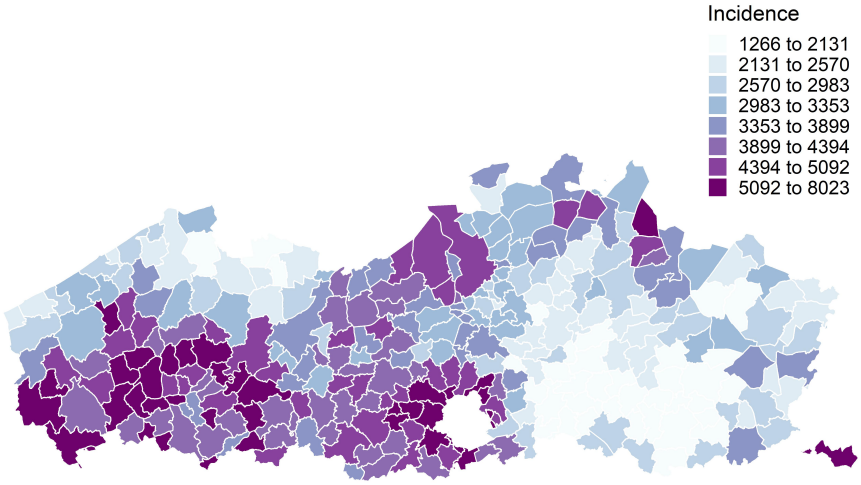


Figure 5.1: Spatial distribution of the incidence of COVID-19 per 100,000 inhabitants in Flanders' municipalities from 2020-09-01 to 2021-01-31.

The data under analysis includes information on the 300 municipalities of the Flanders area in Belgium, which is available at the website of the Belgian Institute for Public Health (Sciensano) (https://epistat.wiv-isp.be/covid/). Table 5.1 includes some descriptive statistics for the variables available across municipalities. The outcome of interest is the number of confirmed COVID-cases from 2020-09-01 to 2021-01-11, summarized by the variable N.cases. The population size in the area is denoted as P, and incidence is the number of COVID-19 cases in this time period per 100,000 inhabitants. There are also two additional variables available which can be considered as proxies for the socio-economic status and demography of the municipality. These are the percentage of households with a discount on the electricity meter (budgetmeters) and the percentage of single-parent households (single_house).

Table 5.1: Descriptive statistics for the variables available across municipalities.

	Median	Mean	SD	Min.	Max.
N.cases	544.00	801.24	1570.14	1.00	24387.00
P	15036.50	22097.14	36156.40	79.00	529247.00
incidence	3358.87	3564.18	1235.74	1265.82	8023.07
single_house	7.88	8.02	1.25	5.30	15.30
budgetmeters	1.12	1.23	0.68	0.00	6.86

5.3.2 Model Estimates

The COVID-19 incidence map in Figure 5.1 suggests the presence of spatial autocorrelation in the data, as municipalities with similar values of COVID-19 incidence are grouped together in space. Therefore, we will further analyse this data by implementing spatial models that account for the spatial dependence. Additionally, we will explore different choices for the spatial weights matrix to be included in the fitted models.

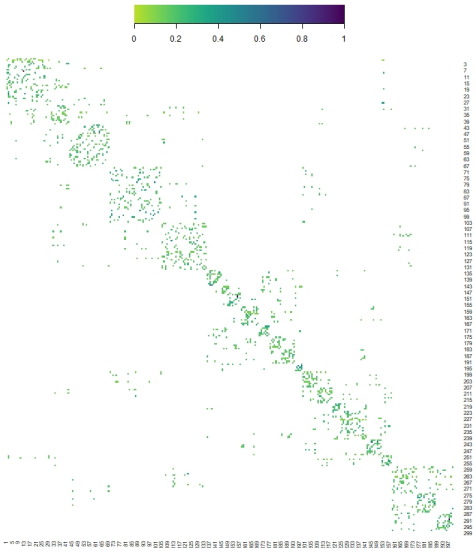
We fit both the spatial conditional normal Poisson model in equation (5.1) (Section 5.2.1) and the proposed geometric mean spatial conditional normal Poisson model in equation (5.4) (Section 5.2.2). As we do not wish to pre-specify the weights in the weights matrix $\mathbf{W}$, we use the different weights matrices described in Section 5.2.3, and compare the fitting of the different models by using their DIC and WAIC values. Note that, in this specific application, we do not include any covariates in the linear predictor, as we focus on the spatial modelling by means of the autoregressive terms and on the comparison of the performance of such models.

We believe it is important to mention the fact that, at the beginning of this research, the variables available were included in the model as covariates. However, the results obtained suggested that they did not offer any improvements in models' fitting in terms of information criteria. Therefore, in this specific study, we decided to only employ them when computing the proposed weights matrices based on similarities. It should also be noted that, in this study, we do not aim to identify any risk factor in the spreading of the infection, but to investigate the spatial correlation that may exist in the data and find the structures that best accommodate it.

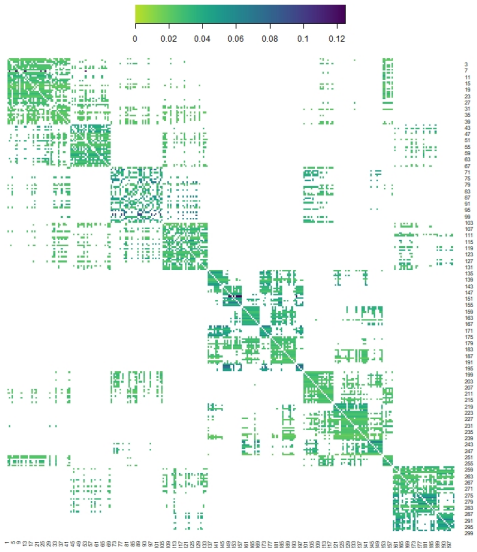
The results obtained for the fitting of these models are included in Tables 5.2 and 5.3, which were fitted by considering ten different options for the weights matrix. First, we have used the spatial weights matrices based on the adjacency among regions (contiguity of first and third order). Second, weights were based on the distance among the centroids of the regions (inverse distance, negative exponential distance and distance band method). Third, weights matrix were based on the product between covariate differences and traditional spatial weights, as proposed in Section 5.2.3. For these similarity matrices, the spatial weights matrices considered are the ones based on contiguity of first order, and that based on the distance band. The variables used to measure whether regions have a similar socio-economic status are single_house and budgetmeters. Finally, the mobility matrix was also considered.

Figure 5.2 shows the heatmaps of the weights matrices considered here. Heatmaps use colours to represent the values of the weights for each matrix. Thus, white would indicate that the weights are zero for those municipalities. That is, heatmaps are graphical representations for the individual weights in each of the different weights matrix structures. More specifically, matrices following the inverse and the negative exponential distance only have zeros in their diagonal, for the weights w_{ii}, $i = 1, \ldots, n$, whereas the rest of the matrices have a larger percentage of weights that are zero. Moreover, matrices presenting the largest number of connected areas are the ones following the contiguity of order three and the mobility matrix. Finally, matrices where the contiguity of order one criterion is considered are the ones that present the smallest number of

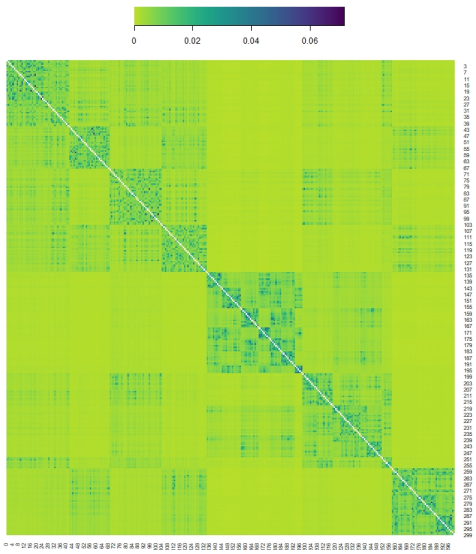
connected regions and, hence, the ones having the largest percentage of weights that are zero.



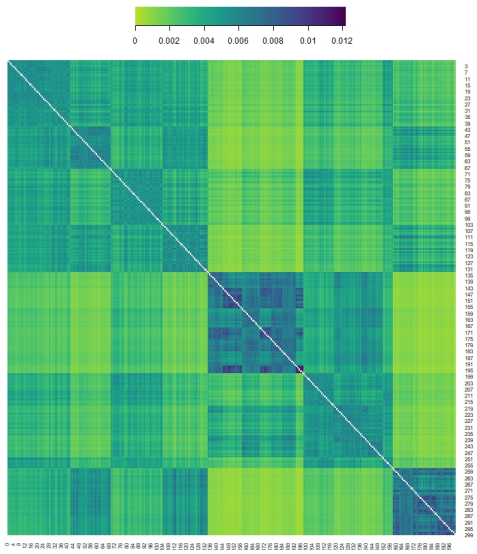
(a) Heatmap of the spatial weights matrix following contiguity of order 1.



(b) Heatmap of the spatial weights matrix following contiguity of order 3.



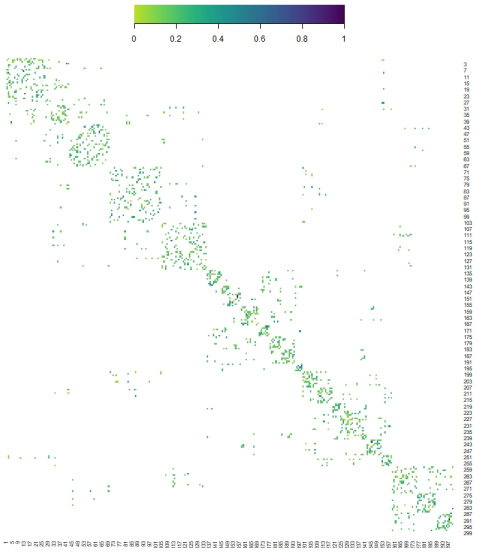
(c) Heatmap of the spatial weights matrix following inverse distance.



(d) Heatmap of the spatial weights matrix following negative exponential.

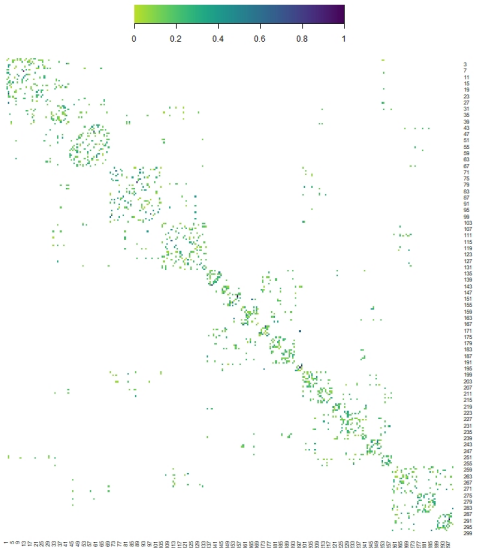
Figure 5.2: Heatmaps of the spatial weights matrices considered.

(e) Heatmap of the spatial weights matrix following distance band.



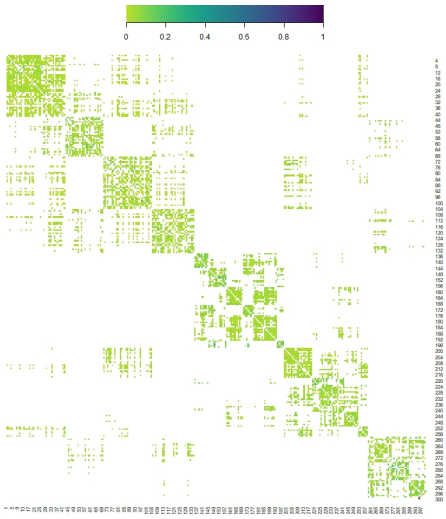
(f) Heatmap of the similarity spatial weights matrix combining contiguity order 1 and the variable **budgetmeters**.

(g) Heatmap of the similarity spatial weights matrix combining distance band and the variable **budgetmeters**.

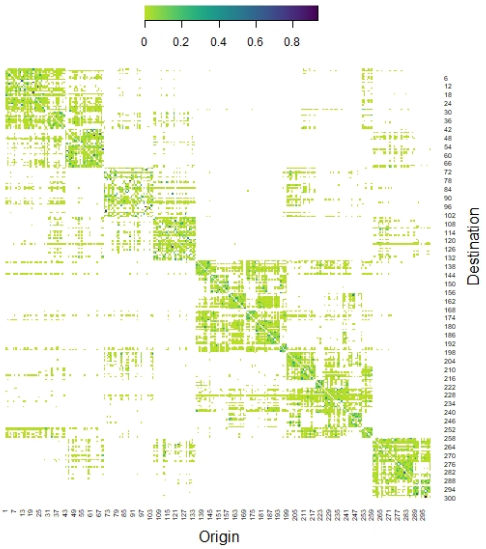


(h) Heatmap of the similarity spatial weights matrix combining contiguity of order 1 and the variable **single_house**.

Figure 5.2: Heatmaps of the spatial weights matrices considered (Continued).



(i) Heatmap of the similarity spatial weights matrix combining distance band and the variable **single_house**.



(j) Heatmap of the mobility matrix.

Figure 5.2: Heatmaps of the spatial weights matrices considered (Continued).

Table 5.2: Results obtained after fitting the spatial conditional normal Poisson models to the COVID-19 incidence data in Flanders, for the different weights matrices considered.

Weights matrix			$\hat{\beta}$	$\hat{\rho}$	$\hat{\tau}$
Contiguity of order 1	DIC = 3015.6	Mean	-4.3779	27.9278	27.8299
	WAIC = 2947.7	SD	(0.0414)	(1.1205)	(2.4400)
		95% CI	(-4.4593,-4.2967)	(25.7277,30.1287)	(23.2868,32.8680)
Contiguity of order 3	DIC = 3019.1	Mean	-4.4245	29.2200	18.0993
	WAIC = 2942.7	SD	(0.0597)	(1.6368)	(1.5422)
		95% CI	(-4.5419,-4.3074)	(26.0055,32.4342)	(15.2166,21.2709)
Inverse distance	DIC = 3018	Mean	-5.6489	63.0826	19.3191
	WAIC = 2941.2	SD	(0.1202)	(3.3331)	(1.6462)
		95% CI	(-5.8853,-5.4131)	(56.5380,69.6293)	(16.2424,22.7046)
Negative exponential	DIC = 3022.1	Mean	-6.0063	73.1322	12.6599
	WAIC = 2941.4	SD	(0.2207)	(6.1524)	(1.0624)
		95% CI	(-6.4402,-5.5731)	(61.0492,85.2137)	(10.6724,14.8430)
Distance band	DIC = 3013.5	Mean	-4.5140	31.4797	24.6629
	WAIC = 2938.4	SD	(0.0505)	(1.3703)	(2.1204)
		95% CI	(-4.6134,-4.4149)	(28.7899,34.1716)	(20.7059,29.0307)
$W \circ D$ single_house and Contiguity of order 1	DIC = 3015.9	Mean	-4.3490	27.1208	26.9563
	WAIC = 2946.9	SD	(0.0411)	(1.1121)	(2.3552)
		95% CI	(-4.4300,-4.2684)	(24.9376,29.3054)	(22.5668,31.8139)
$W \circ D$ single_house and Distance band	DIC = 3012.6	Mean	-4.4899	30.9698	27.6089
	WAIC = 2940.3	SD	(0.0459)	(1.2463)	(2.3962)
		95% CI	(-4.5802,-4.4000)	(28.5236,33.4184)	(23.1419,32.5503)
$W \circ D$ budgetmeters and Contiguity of order 1	DIC = 3015.7	Mean	-4.3598	27.4999	29.0054
	WAIC = 2949.4	SD	(0.0396)	(1.0736)	(2.5564)
		95% CI	(-4.4378,-4.2821)	(25.3922,29.6088)	(24.2465,34.2842)
$W \circ D$ budgetmeters and Distance band	DIC = 3012.2	Mean	-4.5130	31.6566	27.0085
	WAIC = 2938.7	SD	(0.0474)	(1.2925)	(2.3358)
		95% CI	(-4.6063,-4.4200)	(29.1197,34.1958)	(22.6523,31.8233)
Mobility	DIC = 3029.9	Mean	-4.2698	25.1133	22.0095
	WAIC = 2972.6	SD	(0.0445)	(1.2129)	(1.9735)
		95% CI	(-4.3571,-4.1824)	(22.7263,27.4908)	(18.3415,26.0931)

Table 5.3: Results obtained after fitting the geometric mean spatial conditional normal Poisson models to the COVID-19 incidence data in Flanders, for the different weights matrices considered.

Weights matrix			$\hat{\beta}$	$\hat{\rho}$	$\hat{\tau}$
Contiguity of order 1	DIC = 3018	Mean	-0.7863	0.7695	22.4881
	WAIC = 2945.1	SD	(0.1223)	(0.0360)	(1.9380)
		95% CI	(-1.0265,-0.5462)	(0.6988,0.8402)	(18.8708,26.4785)
Contiguity of order 3	DIC = 3020.5	Mean	-0.8845	0.7402	15.7244
	WAIC = 2942.1	SD	(0.1616)	(0.0476)	(1.3309)
		95% CI	(-1.2020,-0.5672)	(0.6468,0.8336)	(13.2348,18.4591)
Inverse distance	DIC = 3017.9	Mean	3.7527	2.1075	19.1440
	WAIC = 2941	SD	(0.3804)	(0.1122)	(1.6297)
		95% CI	(3.0056,4.4999)	(1.8872,2.3279)	(16.0966,22.4940)
Negative exponential	DIC = 3022.1	Mean	4.9186	2.4510	12.7087
	WAIC = 2941.4	SD	(0.6950)	(0.2050)	(1.0668)
		95% CI	(3.5537,6.2834)	(2.0485,2.8535)	(10.7116,14.9004)
Distance band	DIC = 3013.3	Mean	0.2414	1.0705	25.0003
	WAIC = 2938.6	SD	(0.1569)	(0.0461)	(2.1514)
		95% CI	(-0.0666,0.5495)	(0.9799,1.1612)	(20.9854,29.4322)
$W \circ D$ single_house and Contiguity of order 1	DIC = 3018	Mean	-0.8108	0.7621	22.5347
	WAIC = 2945.3	SD	(0.1209)	(0.0356)	(1.9427)
		95% CI	(-1.0484,-0.5733)	(0.6922,0.8320)	(18.9093,26.5355)
$W \circ D$ single_house and Distance band	DIC = 3012.3	Mean	0.2155	1.0617	28.1287
	WAIC = 2940.4	SD	(0.1436)	(0.0422)	(2.4449)
		95% CI	(-0.0663,0.4975)	(0.9789,1.1446)	(23.5714,33.1695)
$W \circ D$ budgetmeters and Contiguity of order 1	DIC = 3017.8	Mean	-0.7804	0.7703	23.6219
	WAIC = 2946.3	SD	(0.1181)	(0.0347)	(2.0452)
		95% CI	(-1.0124,-0.5485)	(0.7021,0.8385)	(19.8088,27.8381)
$W \circ D$ budgetmeters and Distance band	DIC = 3012	Mean	0.2676	1.0764	27.5007
	WAIC = 2938.9	SD	(0.1478)	(0.0434)	(2.3800)
		95% CI	(-0.0226,0.5579)	(0.9912,1.1616)	(23.0584,32.4024)
Mobility	DIC = 3033.8	Mean	-1.7658	0.4823	14.8418
	WAIC = 2960.6	SD	(0.1148)	(0.0338)	(1.2750)
		95% CI	(-1.9925,-1.5414)	(0.4155,0.5484)	(12.4600,17.4657)

When comparing the models' fit related to the different weights matrices included in Table 5.2, it can be seen that parameter estimates can differ considerably. The estimated value for the autoregressive parameter ρ is large and statistically significant, according to its 95% credible interval, in all models, an indication that there is a clear sign for the existence of spatial autocorrelation. Interpretation of the value of the estimated parameter is difficult, however.

The information criteria values obtained for the fitting of these models indicate that the best fit among the models accounting only for contiguity or distance among regions is for the distance band spatial weights (DIC = 3013.5 and WAIC = 2938.4). The next best fitting model is the one using the contiguity of order one criterion (DIC = 3015.6 and WAIC = 2947.7). As for the models taking into account the similarity in socio-economic status, the combination of single_house or budgetmeters and distance bands are the best fitting models (DIC = 3012.6 and WAIC = 2940.3, and DIC = 3012.2 and WAIC = 2938.7, respectively). For the model considering the mobility matrix, we can conclude that, according to the information criteria, this model did not provide a good fit.

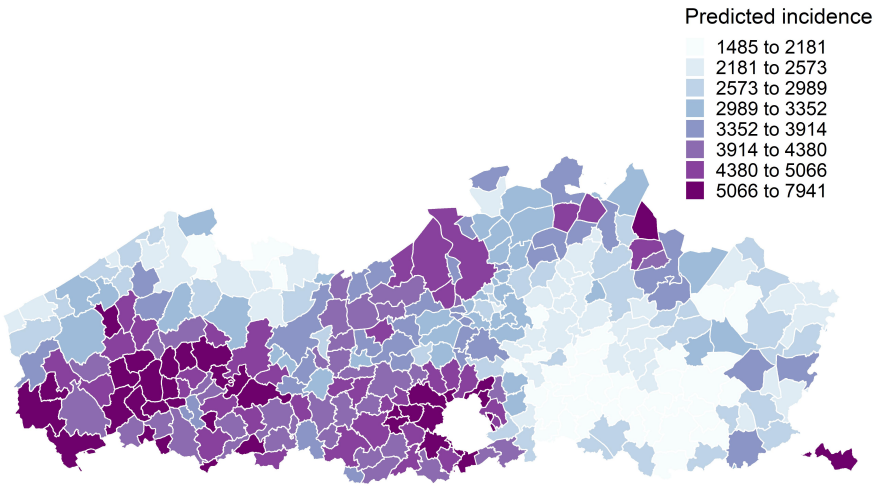
Similar results are observed in Table 5.3, where the fitting of these models appears to be very similar to the ones in Table 5.2, according to the information criteria values. Here, the models with the smallest DIC and WAIC values were the ones using the similarity matrix of the distance band and single_house or budgetmeters (DIC = 3012.3 and WAIC = 2940.4, and DIC = 3012 and WAIC = 2938.9, respectively). In these weights matrices, larger weights are specified for regions that lie within the distance band and have similar values of these variables. Therefore, the fitting of these models suggests that this structure could be properly explaining the underlying spatial dependence, assuming that the variables considered represent the socio-economic or demographic characteristics of the population in these areas.

Regarding the spatial autoregressive parameter ρ, here the spatial lag is also significant for all the fitted models, indicating that the spatial autocorrelation is being properly captured. Moreover, the interpretation of this parameter can be useful in order to quantify how much the spatial structure considered can influence the resemblance of the incidence rate in a region to the geometric mean of the incidence rates of its neighbours. In the models where the distance band matrix was used, the parameter ρ has posterior mean approximately equal to 1, and, thus, in this setting, we find that the rate in a municipality is close to the geometric mean of the rates in the municipalities within the distance band. For the models where the specified weights matrix was either the exponential or the inverse distance, the estimated values of ρ was approximately equal to two, suggesting that the rate in a municipality is the square of the geometric mean of the rates of its neighbours. For the remaining models, this parameter's estimated value was smaller than one. For example, in the model with the mobility matrix, it was $\hat{\rho} = 0.4823$, suggesting that, for this connectivity structure, the rate in a municipality is approximately the squared root of the geometric mean of the rates of its neighbours.

In Figure 5.3, we include the maps of the predicted incidence obtained after fitting some of the geometric mean spatial conditional normal Poisson models considered. If

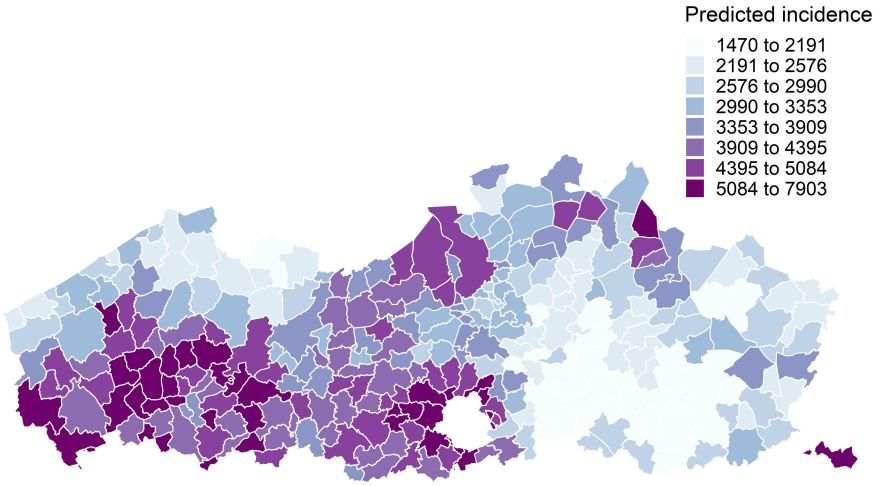
we compare these maps with the observed incidence map shown in Figure 5.1, we can see that, in general, the predictions are quite accurate, as they are very similar to the observed incidence. In addition, when compared to each other, we note that the predictions obtained differ only for a small number of municipalities.

Furthermore, scatterplots of the observed versus the predicted rates, obtained from the fitting of these models are included in Figure 5.4, where it can be seen that the fitted models show high accuracy in the prediction of the incidence rates. In the plots included in Figure 5.4, we can see a point at the beginning of the line which always seems to fall away from it. This point corresponds to the municipality of Herstappe, which has only one positive COVID-19 case for the whole period under analysis and a population of only 79 individuals. Note that, after Herstappe, the next municipality with fewer cases is Zuienkerke, with a total number of positive cases of 60 and a population of 2709 individuals. This may be the reason that could explain the fact that the predicted rate of Herstappe is mostly dominated by the information provided by its neighbours.

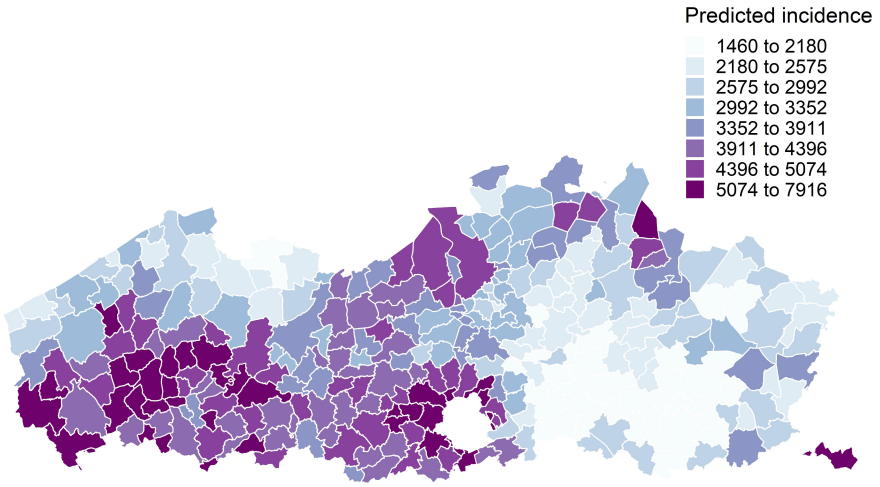


(a) Predicted incidence obtained from the model using the spatial matrix following the contiguity of order one criterion, fitted to the COVID-19 data in Flanders.

Figure 5.3: Predicted incidence obtained from some of the geometric mean spatial conditional normal Poisson models considered, fitted to the COVID-19 data in Flanders.

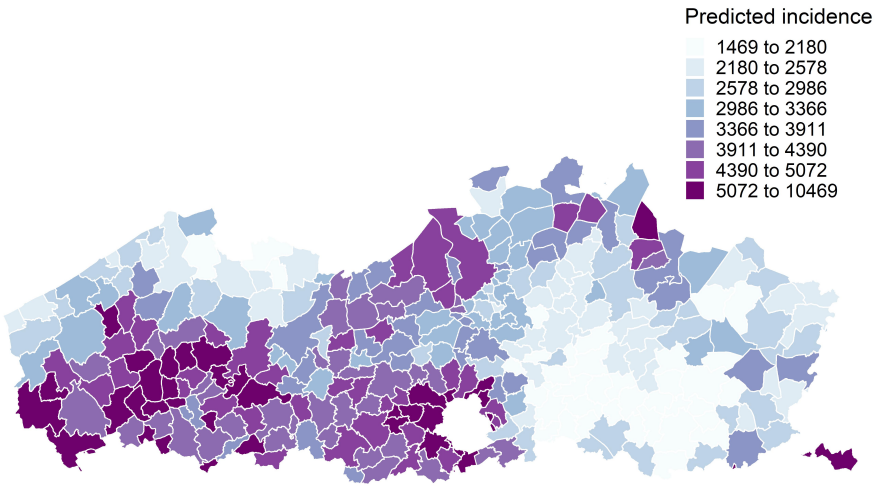


(b) Predicted incidence obtained from the model using the spatial matrix following the distance band criterion, fitted to the COVID-19 data in Flanders.



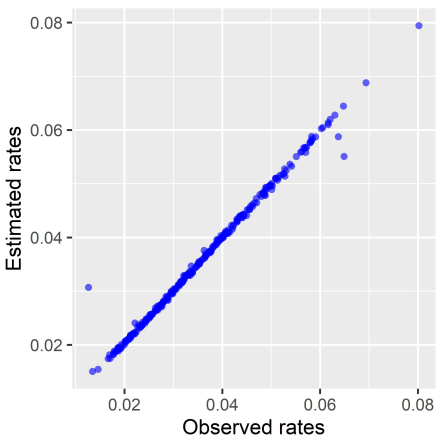
(c) Predicted incidence obtained from the model using the similarity spatial matrix combining the differences in the variable budgetmeters and the distance bands criterion, fitted to the COVID-19 data in Flanders.

Figure 5.3: Predicted incidence obtained from some of the geometric mean spatial conditional normal Poisson models considered, fitted to the COVID-19 data in Flanders (Continued).

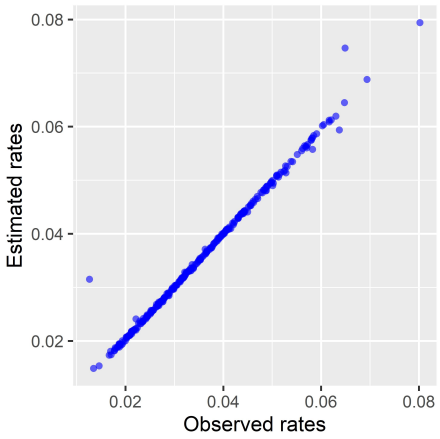


(d) Predicted incidence obtained from the model using the mobility matrix, fitted to the COVID-19 data in Flanders.

Figure 5.3: Predicted incidence obtained from some of the geometric mean spatial conditional normal Poisson models considered, fitted to the COVID-19 data in Flanders (Continued).

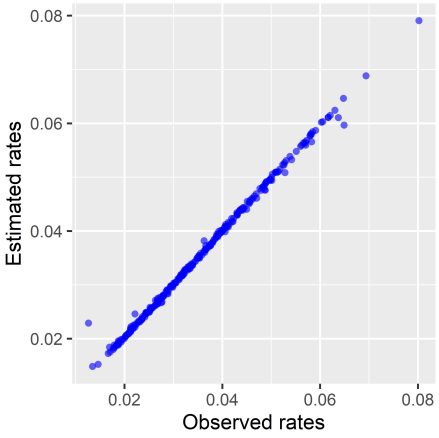


(a) Model where the spatial matrix follows the contiguity of order one criterion.

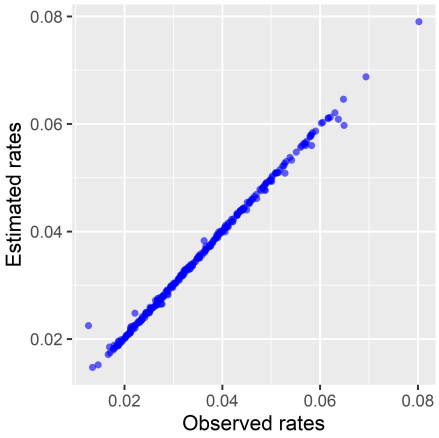


(b) Geometric mean model where the spatial matrix follows the contiguity of order three criterion.

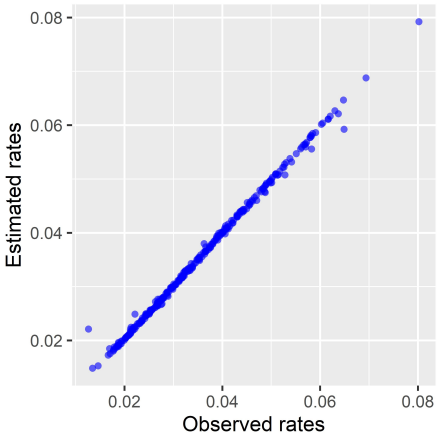
Figure 5.4: Scatterplots of the observed versus the predicted rates obtained for some of the fitted models.



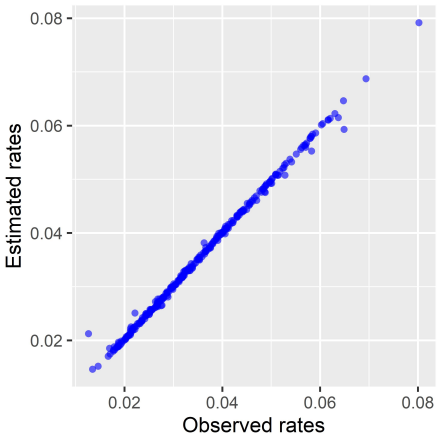
(c) Model where the spatial matrix follows the negative exponential distance criterion.



(d) Geometric mean model where the spatial matrix follows the distance bands criterion.

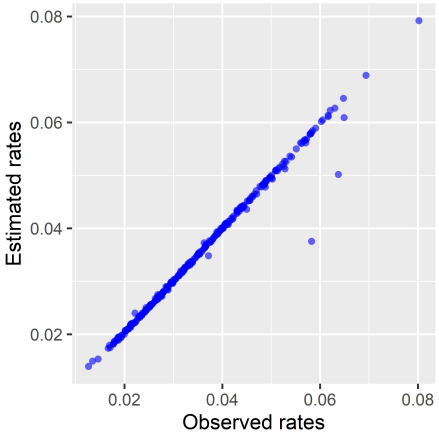


(e) Model for the similarity spatial weights matrix combining distance band and the variable **budgetmeters**.

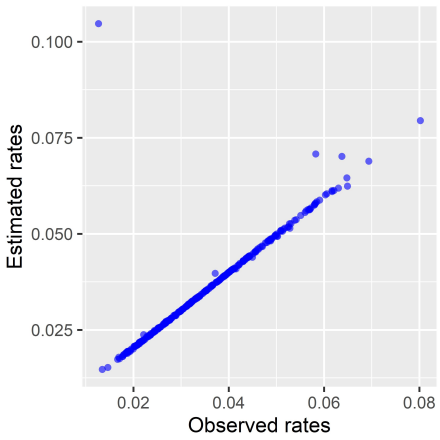


(f) Geometric mean model for the similarity spatial weights matrix combining distance band and the variable **budgetmeters**.

Figure 5.4: Scatterplots of the observed versus the predicted rates obtained for some of the fitted models (Continued).



(g) Model for the mobility matrix.



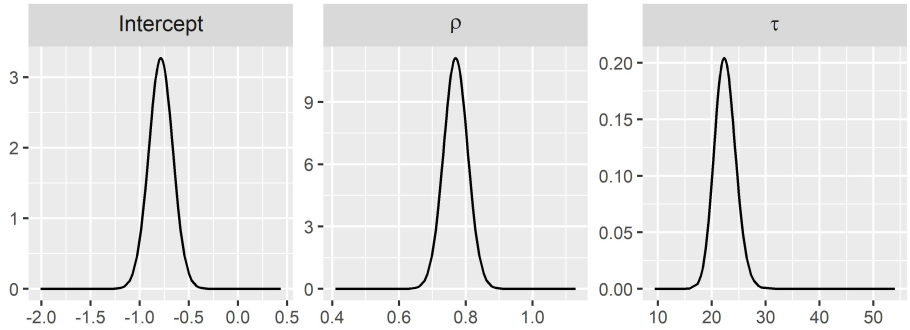
(h) Geometric mean model for the mobility matrix.

Figure 5.4: Scatterplots of the observed versus the predicted rates obtained for some of the fitted models (Continued).

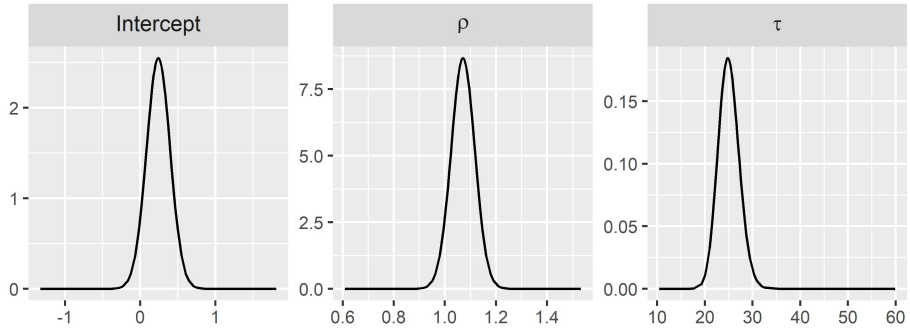
In addition, Figure 5.5 shows the marginal posterior distribution of the parameters estimated from some of the fitted models, where it can be verified that the normality assumption holds.

We can also check the distributional assumptions in the fitted models, which is a Poisson distribution, where the overdispersion is accommodated by means of the inclusion of a random effect in the regression for the mean. This can be achieved by using the distribution_check function from the R package inlatools (Onkelinx, 2019). Here, simulations are drawn from the model and the empirical cumulative distribution function (eCDF) is computed for the observed response and for the simulated data, so that they can be compared.

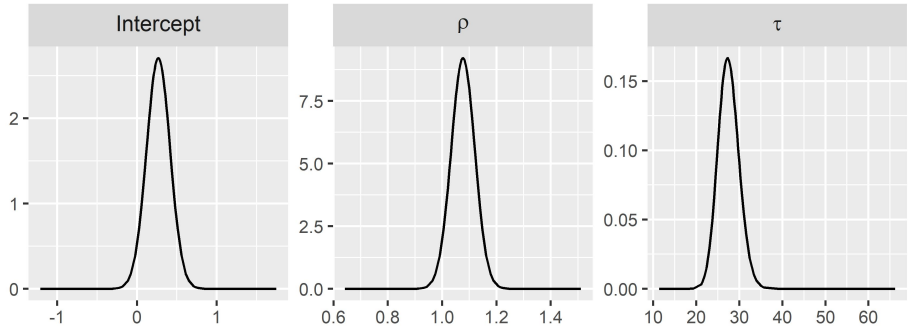
Figure 5.6 includes the plots which illustrate these comparison results. In each figure, the black line is the result of dividing the eCDF of the observed data by the median of the eCDF's of the simulated data sets, and the grey bands represent the 95% credible intervals of the simulated data. In addition, the dotted horizontal line placed at 100% indicates where the ratio of the eCDF's is equal to one. If the eCDF is inside the credible intervals, which is the case for all of the models fitted here, the assumed distribution in the model seems to be a plausible one. Moreover, given that the eCDF is quite close to the reference line, these results suggest that the data is well modelled with this distribution.



(a) Geometric mean model where the spatial matrix follows the contiguity of order one criterion.

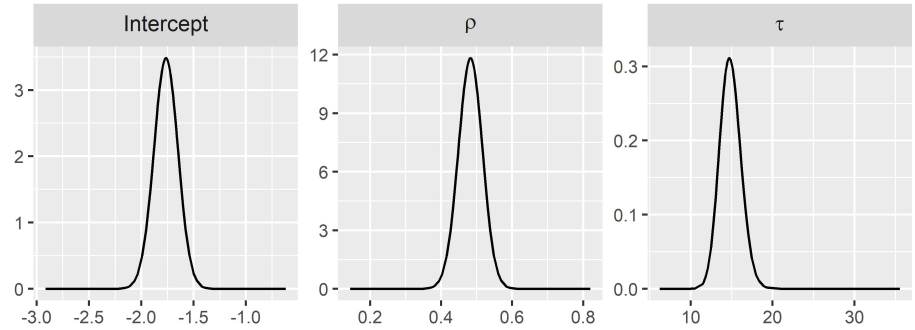


(b) Geometric mean model where the spatial matrix follows the distance band criterion.



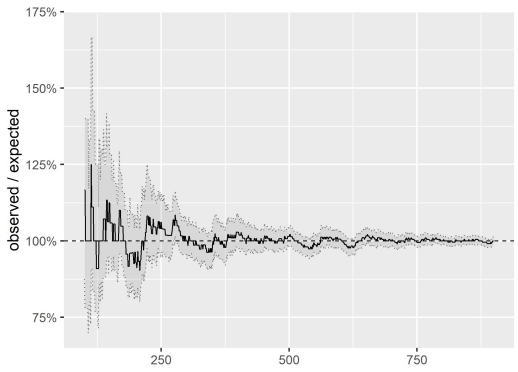
(c) Geometric mean model for the similarity spatial weights matrix combining distance band and the variable **budgetmeters**.

Figure 5.5: Posterior densities from the parameters estimated for some of the fitted models.

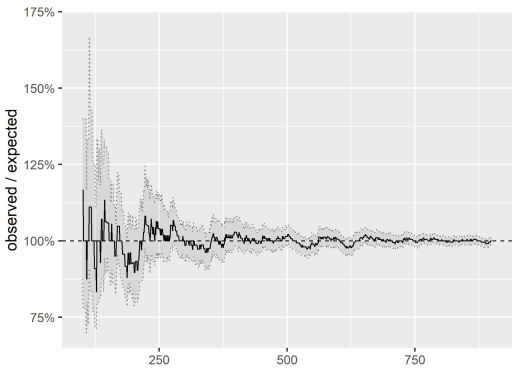


(d) Geometric mean model for the mobility matrix.

Figure 5.5: Posterior densities from the parameters estimated for some of the fitted models (Continued).

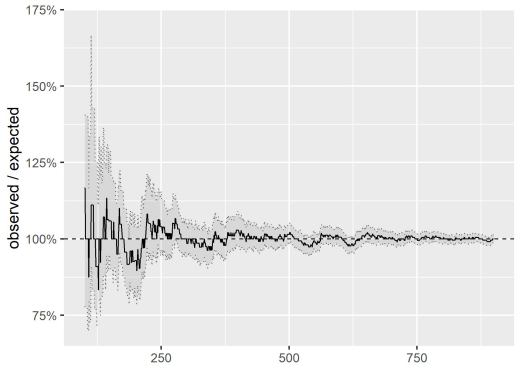


(a) Geometric mean model where the spatial matrix follows the contiguity of order one criterion.

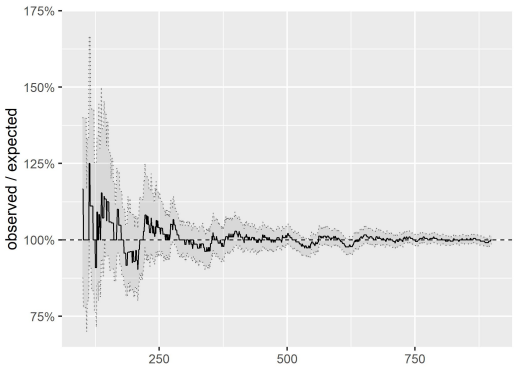


(b) Geometric mean model where the spatial matrix follows the distance band criterion.

Figure 5.6: Distribution check for some of the fitted models.



(c) Geometric mean model for the similarity spatial weights matrix combining distance band and the variable **budgetmeters**.



(d) Geometric mean model for the mobility matrix.

Figure 5.6: Distribution check for some of the fitted models (Continued).

After examining the results obtained in this section, we could conclude that, on the one hand, with the proposed model we present an appealing interpretation of the spatial parameter, given by the geometric mean of the incidence rates. We have shown how this interpretation can change for the different fitted models, indicating how much the spatial structure considered explains the spatial autocorrelation by means of the geometric mean of the rates in the neighbouring regions. On the other hand, by examining different weights matrices, we can have a better idea of the underlying spatial dependence structure of the data. When the similarity matrices based on the distance band were used, the information criteria values were similar to the model considering the traditional distance band matrix. Therefore, taking into account that they provide similar predictions and similar fit, we believe that, for the specific data set considered, this weights matrix could represent a proper choice for modelling the spatial underlying structure of the data.

5.4 Comparison to the BYM2 model

In this section, we will fit the BYM2 models to the COVID-19 data in Flanders. However, we should stress here that one of our main goals in this thesis is to present the geometric mean proposal as a new extension of the spatial conditional Poisson model in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018). In these models, the interpretation of the spatial parameters is different from that of the BYM2 model. Furthermore, the spatial conditional and the geometric mean models offer the possibility of specifying any weights matrix in a straightforward way, as it is used for computing a spatial lag. In our view, this feature makes these models more appealing for investigating different spatial structures, which is another one of our goals here. In the case of the BYM2 model, this is not straightforward due to its limitations, where the assumed spatial structure needs to be

symmetric, which is not the case, for example, for the mobility matrix we have employed before.

Nevertheless, we believe it can be useful to compare the performance of the proposed methods with that of the BYM2 model, often employed in disease mapping applications. Therefore, we consider the model in equation (5.8), where, in order to specify the penalized complexity priors and following Simpson et al. (2017), for the precision parameter τ_s we assume that $\text{Prob}(1/\sqrt{\tau_s} > 0.2/31) = 0.01$ and, for the mixing parameter ϕ_s, $\text{Prob}(\phi_s < 0.5) = 2/3$. The results obtained after fitting this model to the COVID-19 data in Flanders are included in Table 5.4.

Table 5.4: Results obtained after fitting the BYM2 model to the COVID-19 incidence data in Flanders.

	Mean	SD	95% CI
$\hat{\beta}$	-3.3924	0.0039	(-3.4002,-3.3848)
$\hat{\tau_s}$	12.0567	1.1313	(9.8847,14.3180)
$\hat{\phi_s}$	0.9757	0.0202	(0.9231,0.9979)
DIC = 3006.9 WAIC = 2932.9			

Let us recall that, in the previous section, the smallest DIC was obtained for the geometric mean model using the similarity matrix of the distance band and budgetmeters (DIC = 3012) and the smallest WAIC was given for the model using the distance band criterion (WAIC = 2938.6). For the BYM2 model, we can see that the DIC and WAIC values obtained are slightly lower than those obtained for the previous models. In addition, the value obtained for the mixing parameter, $\hat{\phi_s} = 0.9757$, suggests that more than 97% of the variability in the data is being explained by the spatially structured effect.

Regarding the predictive accuracy of this model, Figure 5.7 includes the map of the predicted incidence obtained from its fitting, where we can see that the predictions are very accurate when compared to the map of the observed incidence in Figure 5.1, and also very similar to the ones obtained in the previous section for our proposed methods (see Figure 5.3). The scatterplot of the observed versus the predicted incidence rates is included in Figure 5.8, showing some issues in some of the municipalities.

Despite the fact that the information criteria values favoured the BYM2 model and that its predictive accuracy is similar to the one from the geometric mean model, we restate our goal here of presenting the geometric mean proposal, which can be viewed as an alternative to the BYM and BYM2 models, and to investigate the weights matrices which best reflect the spatial underlying process.

There are situations where the spatial conditional models might offer a better fit than the BYM and BYM2 models, or viceversa. We believe that the choice of the model to fit should depend on the specific objective of the study. For example, Morales-Otero and Núñez-Antón (2021) reported that, given by the information criteria values obtained, the spatial conditional and the BYM and BYM2 models offered a very similar fitting to the infant mortality data they studied. In addition, in Morales-Otero, Gómez-Rubio and

Núñez-Antón (2022), the spatial conditional models were employed in order to illustrate a new fitting approach in INLA.

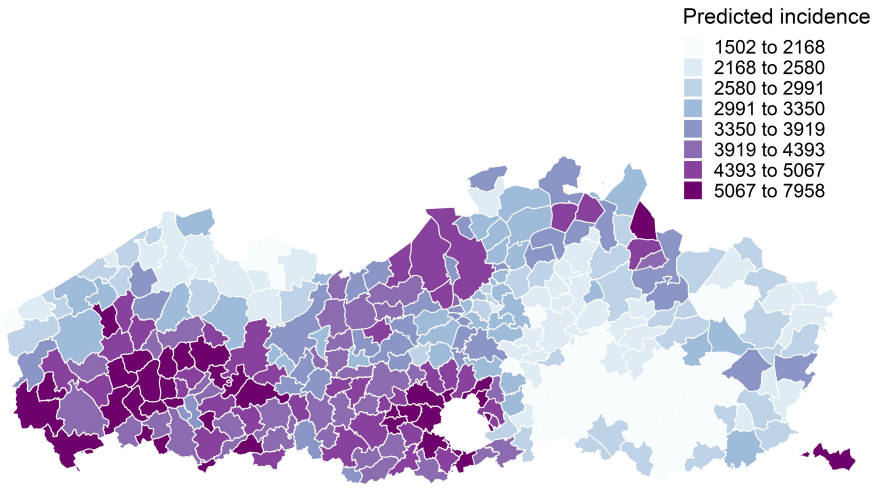


Figure 5.7: Predicted incidence obtained from the fitting of the BYM2 to the COVID-19 data in Flanders.

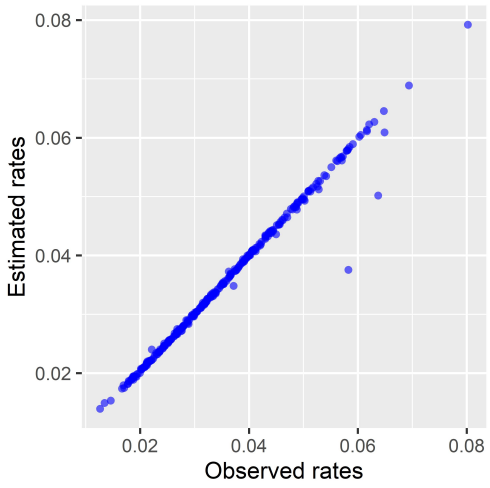


Figure 5.8: Scatterplot of the observed versus the predicted rates obtained from the fitting of the BYM2 to the COVID-19 data in Flanders.

5.5 Simulation study

Based on the previous results, the fitted models suggest a strong spatial correlation in the data, which is clearly better explained by the distance band spatial weights matrix.

This implies that, for the data under study, the underlying spatial process is closely related to this spatial structure. We wish to investigate the performance of the selection of the weights matrix, and study the sensitivity of the parameters to a misspecified neighbourhood matrix.

Therefore, we have carried out a simulation study, where we induce correlation in the response variable following the mobility matrix structure. For this purpose, we have implemented a Gibbs sampling algorithm, which allowed us to generate spatially autocorrelated Poisson data by repeatedly sampling from conditional distributions (see Jackson and Sellers, 2008). In our specific case, we define a set of initial values for the parameters β, ρ and τ and, on each iteration, we draw Poisson samples, where the mean is conditioned on the values of the previous iteration. For $k = 1$, we draw Poisson samples from an uncorrelated mean:

$$\begin{array}{rcll} \mu_i^{(1)} & = & \exp(\beta + \log(P_i) + \nu_i^{(1)}) & \\ y_i^{(1)} & \sim & \text{Poi}(\mu_i^{(1)}) & \\ r_i^{(1)} & = & \frac{(y_i^{(1)}+1)}{P_i} & \text{for } i = 1, \ldots, n. \end{array} \tag{5.9}$$

Then, if we perform a total number of S^* iterations, for $k = 2, \ldots, S^*$, the algorithm follows, so that:

$$\begin{array}{rcl} \mu_1^{(k)} & = & \exp\{\beta + \log(P_1) + \rho \sum_{j=1}^{n} w_{1j} \log(r_j^{(k-1)}) + \nu_1^{(k)}\} \\ y_1^{(k)} & \sim & \text{Poi}(\mu_1^{(k)}) \\ r_1^{(k)} & = & \frac{(y_1^{(k)}+1)}{P_1} \\ & & \\ \mu_2^{(k)} & = & \exp\{\beta + \log(P_2) + \rho(w_{21} \log(r_1^{(k)}) + \sum_{j=2}^{n} w_{2j} \log(r_j^{(k-1)})) + \nu_2^{(k)}\} \\ y_2^{(k)} & \sim & \text{Poi}(\mu_2^{(k)}) \\ r_2^{(k)} & = & \frac{(y_2^{(k)}+1)}{P_2} \\ \vdots & & \\ \mu_h^{(k)} & = & \exp\{\beta + \log(P_h) + \rho(\sum_{j=1}^{h-1} w_{hj} \log(r_j^{(k)}) + \sum_{j=h}^{n} w_{hj} \log(r_j^{(k-1)})) + \nu_h^{(k)}\} \\ y_h^{(k)} & \sim & \text{Poi}(\mu_h^{(k)}) \\ r_h^{(k)} & = & \frac{(y_h^{(k)}+1)}{P_h} \\ \vdots & & \\ \mu_n^{(k)} & = & \exp\{\beta + \log(P_n) + \rho \sum_{j=1}^{n-1} w_{nj} \log(r_j^{(k)}) + \nu_n^{(k)}\} \\ y_n^{(k)} & \sim & \text{Poi}(\mu_n^{(k)}) \\ r_n^{(k)} & = & \frac{(y_n^{(k)}+1)}{P_n} \end{array} \tag{5.10}$$

We have defined twelve different scenarios, given the true values for the parameters, which can be consulted in the first column to the left in Table 5.5. For each case, we have simulated $S^* = 500$ data sets (with the number of areas $n = 300$), and discarded half of them, so that $S = 250$ simulations for each scenario remained. We have fitted

three models to each simulated data set, one using the mobility matrix to compute the spatial lag, another one using the contiguity of order one spatial weights matrix, and a third one using the inverse distance spatial weights matrix.

In order to assess the performance of the models fitted to the simulated data, and compare the results, we have computed the bias $\left[(\sum_{s=1}^{S}(\hat{\theta}_s - \theta)/S\right]$, the variance $\left[\sum_{s=1}^{S}(\hat{\theta}_s - \overline{\hat{\theta}})^2/(S-1)\right]$, and the mean squared error (MSE) $\left[\sum_{s=1}^{S}(\hat{\theta}_s - \theta)^2/S\right]$ of the estimations, where θ represents the true value of a parameter, $\hat{\theta}_s$ its resulting estimate for data set s and $\overline{\hat{\theta}}$ is the arithmetic mean of the resulting estimates. The values obtained are included in Table 5.5. Here, we can see how, in general, the bias is smaller for the models where the mobility matrix was used.

For the scenarios where the parameters' true values were $\beta = -2$ and $\rho = 0.5$ (i.e., first two scenarios), the smallest bias was obtained for the estimations for the model using the mobility matrix, indicating that this is the model where the resulting estimates are closer to the true values of the parameters. However, when the true value for β changed to -0.5 (i.e., third and fourth scenarios), the smallest bias was obtained for the model using the contiguity criterion for the weights matrix, which seems to suggests that the value given to the intercept β is having a significant impact on the results.

In the scenarios where the true value for ρ is set to 0.2, the bias of the estimates considerably increases when using the mobility matrix. In fact, the estimations with smallest bias are obtained for the model using the inverse distance criterion for the spatial matrix. This can be due to the fact that here we are setting a small value for the spatial parameter and, thus, forcing the mobility connectivity structure to have a smaller relevance in the simulated data.

In addition, for the parameters' true values $\beta = -2$ and $\rho = 0.9$, the smallest bias of the estimates was also obtained for the mobility matrix. Given that, in this case, we are setting a large value for the spatial parameter, more relevance is given to this structure. However, when $\beta = -0.5$ and $\rho = 0.9$, the models using the contiguity and the mobility matrix produce similar values for the bias of the estimations, meaning that, for this specific setting, the spatial structure is not so clearly defined.

Finally, from the results included in Table 5.5, for the precision parameter τ, no significant changes were observed in the bias of the estimations when changing this parameter's value from 5 to 15.

Table 5.5: Results obtained from the models using different weight matrices, fitted to the simulated data sets.

Fitted model:		Mobility			Contiguity			Inverse distance		
True value		β	ρ	τ	β	ρ	τ	β	ρ	τ
$\beta = -2$,	Bias	0.172	0.044	0.806	-0.718	-0.187	-6.863	12.064	3.049	-6.362
$\rho = 0.5$,	Variance	0.001	9.554e-05	0.198	0.009	5.979e-04	0.060	0.220	0.014	0.077
$\tau = 15$	MSE	0.031	0.002	0.846	0.524	0.036	47.160	145.754	9.312	40.547
$\beta = -2$,	Bias	0.347	0.088	0.319	-0.487	-0.130	-1.303	10.892	2.751	-1.217
$\rho = 0.5$,	Variance	0.002	1.412e-04	0.011	0.008	5.422e-04	0.007	0.155	0.010	0.008
$\tau = 5$	MSE	0.123	0.008	0.113	0.245	0.017	1.705	118.791	7.577	1.489
$\beta = -0.5$,	Bias	0.232	0.233	0.406	0.037	0.027	-0.219	2.623	2.639	-0.296
$\rho = 0.5$,	Variance	5.188e-05	4.861e-05	4.743e-04	4.612e-05	4.658e-05	6.047e-04	8.466e-04	7.994e-04	6.370e-04
$\tau = 5$	MSE	0.054	0.054	0.166	0.001	7.769e-04	0.049	6.883	6.966	0.088
$\beta = -0.5$,	Bias	0.187	0.189	1.181	-0.010	-0.018	-1.285	2.612	2.632	-1.367
$\rho = 0.5$,	Variance	1.230e-04	1.219e-04	0.014	1.306e-04	1.354e-04	0.017	0.003	0.003	0.017
$\tau = 15$	MSE	0.035	0.036	1.409	2.215e-04	4.489e-04	1.669	6.824	6.932	1.886
$\beta = -2$,	Bias	0.138	0.055	0.855	-0.539	-0.219	-0.318	0.224	0.088	-0.317
$\rho = 0.2$,	Variance	0.002	2.455e-04	0.052	0.002	3.419e-04	0.053	0.233	0.038	0.053
$\tau = 15$	MSE	0.020	0.003	0.783	0.293	0.048	0.153	0.282	0.045	0.153
$\beta = -2$,	Bias	0.229	0.092	0.307	-0.501	-0.203	0.045	-0.066	-0.029	0.045
$\rho = 0.2$,	Variance	0.002	2.518e-04	0.003	0.001	2.337e-04	0.003	0.126	0.020	0.003
$\tau = 5$	MSE	0.054	0.009	0.097	0.252	0.042	0.005	0.130	0.021	0.005
$\beta = -0.5$,	Bias	0.083	0.131	0.308	-0.095	-0.156	0.171	-0.071	-0.118	0.169
$\rho = 0.2$,	Variance	2.894e-05	6.806e-05	3.599e-04	2.482e-05	5.930e-05	4.238e-04	0.001	0.003	4.254e-04
$\tau = 5$	MSE	0.007	0.017	0.095	0.009	0.024	0.030	0.006	0.017	0.029

Table 5.5: Results obtained from the models using different weights matrices, fitted to the simulated data sets (Continued).

Fitted model:		Mobility			Contiguity			Inverse distance		
True value		β	ρ	τ	β	ρ	τ	β	ρ	τ
$\beta = -0.5$,	Bias	0.082	0.130	0.928	-0.104	-0.171	0.440	-0.058	-0.095	0.438
$\rho = 0.2$,	Variance	8.460e-05	2.060e-04	0.007	6.376e-05	1.633e-04	0.010	0.003	0.008	0.010
$\tau = 15$	MSE	0.007	0.017	0.869	0.011	0.029	0.203	0.007	0.017	0.201
$\beta = -2$,	Bias	0.106	0.012	110.710	0.584	0.117	-14.665	63.681	7.041	-14.651
$\rho = 0.9$,	Variance	0.002	6.859e-05	26846.151	0.301	0.004	7.995e-04	30.127	0.357	8.334e-04
$\tau = 15$	MSE	0.013	2.227e-04	38995.408	0.640	0.018	215.050	4085.262	49.935	214.643
$\beta = -2$,	Bias	0.191	0.020	13.815	0.410	0.093	-4.615	61.910	6.859	-4.664
$\rho = 0.9$,	Variance	0.002	7.673e-05	84.823	0.339	0.004	0.403	28.125	0.336	7.499e-04
$\tau = 5$	MSE	0.039	4.853e-04	275.350	0.506	0.013	21.701	3860.889	47.386	21.756
$\beta = -0.5$,	Bias	0.353	0.075	0.327	0.295	0.054	-2.981	15.885	3.387	-3.219
$\rho = 0.9$,	Variance	4.726e-04	2.239e-05	0.026	0.002	1.038e-04	0.002	0.106	0.005	0.002
$\tau = 5$	MSE	0.125	0.006	0.132	0.089	0.003	8.891	252.426	11.474	10.362
$\beta = -0.5$,	Bias	0.174	0.037	0.647	0.194	0.033	-11.888	15.331	3.287	-12.256
$\rho = 0.9$,	Variance	3.587e-04	1.640e-05	0.401	0.002	1.107e-04	0.007	0.379	0.018	0.006
$\tau = 15$	MSE	0.031	0.001	0.819	0.040	0.001	141.331	235.405	10.822	150.208

Regarding the predictive accuracy of the models, we can evaluate it by computing the mean squared predictive error (MSPE) of the simulated rates for each simulated data set $\left[\text{MSPE}_s = \sum_{i=1}^{n}(r_i^{(s)} - \hat{r}_i^{(s)})^2/n\right]$ (Carroll et al., 2015). In this way, we can obtain an average for the model fitted for each of the 250 data sets generated for each scenario, so that $\text{MSPE} = \sum_{\text{s}=1}^{\text{S}}(\text{MSPE}_\text{s})/\text{S}$. Note that the models with the lowest values of the MSPE would be considered as the best fitting ones. The results obtained are included in Table 5.6, where we can see that, in general, the MSPE is small in every scenario, but the smallest values are mostly obtained for the models in which the mobility matrix was used to compute the spatial lag of the log-rates.

Table 5.6: Average of the MSPE values obtained from the models using different weights matrices, fitted to the simulated data sets.

True values	Mobility	Contiguity	Inverse distance
$\beta = -2, \rho = 0.5, \tau = 15$	2.323e-06	5.173e-05	5.601e-05
$\beta = -2, \rho = 0.5, \tau = 5$	9.659e-07	2.994e-05	3.190e-05
$\beta = -0.5, \rho = 0.5, \tau = 5$	1.765e-06	1.223e-05	1.212e-05
$\beta = -0.5, \rho = 0.5, \tau = 15$	2.292e-06	4.135e-05	4.204e-05
$\beta = -2, \rho = 0.2, \tau = 15$	2.764e-06	1.793e-05	1.796e-05
$\beta = -2, \rho = 0.2, \tau = 5$	1.683e-06	1.004e-05	1.005e-05
$\beta = -0.5, \rho = 0.2, \tau = 5$	3.518e-06	6.521e-06	6.540e-06
$\beta = -0.5, \rho = 0.2, \tau = 15$	5.497e-06	1.594e-05	1.598e-05
$\beta = -2, \rho = 0.9, \tau = 15$	6.441e-06	9.049e-06	9.917e-06
$\beta = -2, \rho = 0.9, \tau = 5$	4.254e-06	9.247e-06	9.594e-06
$\beta = -0.5, \rho = 0.9, \tau = 5$	8.752e-07	5.207e-05	4.737e-05
$\beta = -0.5, \rho = 0.9, \tau = 15$	2.551e-06	1.071e-04	9.862e-05

Moreover, we have counted the number of times that the information criteria values were smaller in each case so that we can check how many times the "correct" model was selected as the best fitting one. These results are included in Table 5.7. Most of the times, with a very few exceptions, the model where the mobility matrix was used, was selected with the smallest DIC and WAIC values. This indicates that we can indeed, based on the model selection criteria, select the underlying true neighbourhood matrix.

Table 5.7: Number of times that the information criteria values selected each of the models fitted to the simulated data sets as the best fitting ones.

True values		Mobility	Contiguity	Inverse distance
$\beta = -2, \rho = 0.5, \tau = 15$	DIC	250	0	0
	WAIC	250	0	0
$\beta = -2, \rho = 0.5, \tau = 5$	DIC	250	0	0
	WAIC	250	0	0
$\beta = -0.5, \rho = 0.5, \tau = 5$	DIC	250	0	0
	WAIC	250	0	0
$\beta = -0.5, \rho = 0.5, \tau = 15$	DIC	250	0	0
	WAIC	250	0	0
$\beta = -2, \rho = 0.2, \tau = 15$	DIC	243	0	7
	WAIC	247	0	3
$\beta = -2, \rho = 0.2, \tau = 5$	DIC	247	0	3
	WAIC	247	0	3
$\beta = -0.5, \rho = 0.2, \tau = 55$	DIC	250	0	0
	WAIC	250	0	0
$\beta = -0.5, \rho = 0.2, \tau = 15$	DIC	244	3	3
	WAIC	244	4	2
$\beta = -2, \rho = 0.9, \tau = 15$	DIC	250	0	0
	WAIC	248	0	2
$\beta = -2, \rho = 0.9, \tau = 5$	DIC	250	0	0
	WAIC	230	1	19
$\beta = -0.5, \rho = 0.9, \tau = 5$	DIC	250	0	0
	WAIC	210	9	21
$\beta = -0.5, \rho = 0.9, \tau = 15$	DIC	250	0	0
	WAIC	249	0	1

From the results obtained in the simulation study, we can conclude that it is essential to evaluate whether the spatial structure used in a study is the most adequate one. For most of the spatial modelling applications, the spatial weights matrix employed to describe the spatial structure of the region under study is the one following the contiguity of order one criterion. However, we believe it has been clearly shown that this is not always necessarily the best choice.

In this specific study, it has been shown that when the mobility matrix is the underlying structure, and the model is misspecified, in general, the bias of the estimations is larger than the bias obtained for the model using the mobility matrix. Moreover, information criteria values such as the DIC and WAIC and, also predictive accuracy measures such as the MSPE, have favoured the correctly specified model, selecting it as the best fitting one in almost all cases.

5.6 Discussion

In this chapter, we have studied the geographical spread of COVID-19 cases in the municipalities of the Flanders region in Belgium during the period going from September 2020 to January 2021. In order to be able to fit these data, we have considered the Bayesian spatial conditional model proposals (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018; Morales-Otero and Núñez-Antón, 2021), which assume the incidence of cases is conditional on the incidence of cases in the other regions. These models offer a great flexibility and also the possibility that considering different weights matrices can be done in a direct and very simple way.

We have proposed a geometric mean spatial conditional model, where the logarithm of the rates is employed for computing the spatial lag component. This model offers an interpretation of the spatial parameter ρ based on the geometric mean, representing how the incidence rate in one region resembles the geometric mean of the rates in its neighbours. We have compared these proposed models with the ones in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), finding that our proposal provides a similar fit, but offers a particular and straightforward interpretation within the context of the specific data set under analysis.

We have fitted these models by using different definitions for the weights matrices employed to compute the spatial lag, such as the classical ways of accounting for spatial autocorrelation based on contiguity and distance. Moreover, we have proposed alternative weights matrices based on a combination of the similarity of a certain variable in the different locations and the distance between these regions. Additionally, we have also studied the use of the mobility matrix in modelling the COVID-19 incidence data in Flanders, which is given by the proportion of time individuals from one municipality spent in a different one.

In order to offer a comparison of the proposed methods with other commonly used models employed in disease mapping applications, we have fitted the BYM2 model to the data set under analysis. Here, we should highlight the flexibility of the proposed geometric mean over the BYM2 model, given by the straightforward way in which the weights matrix can be included in the model, allowing different structures to be specified in a very simple manner.

Results suggest a strong spatial correlation in the data, which is best explained by the distance band spatial weights matrix. This implies that, for the data under study, the underlying spatial process is well explained and modelled by this spatial structure. Nevertheless, we wished to further investigate the performance of the proposed methods when the correlation among the different municipalities is given by another connectivity pattern, such as, for example, the mobility matrix. To accomplish this additional objective, a simulation study was carried out, where we induce correlation in the response variable based on the mobility matrix, and we have been able to appropriately verify that the models are able to identify the correct spatial structure for most of the cases under study.

Chapter 6

Fitting double hierarchical generalized linear models with the integrated nested Laplace approximation

6.1 Introduction

When fitting the generalized overdispersion models in Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón (2012) and Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) to the infant mortality data and to the mother's postnatal period screening test data, in Chapter 2, we found that they could not be directly implemented in INLA. Moreover, as these models can be included in the class of double hierarchical generalized linear models (DHGLM) (Lee and Nelder, 2006), we have developed a method that allows us to fit DHGLM in this R package.

DHGLM provide a unique approach to modelling highly structured data sets allowing for additional flexibility, particularly when modelling the dispersion parameters. The class of DHGLM encompasses a large variety of models, such as standard generalized linear models (GLM), mixed effects models, random coefficient models, semiparametric models and many others. A typical DHGLM includes a linear mixed-effects term to model the mean, as well as several terms to model the scale parameters of the likelihood and/or random effects present in the model.

Estimation of DHGLM can be approached in different ways. Lee and Nelder (2006) proposed the use of the H-likelihood for model fitting and Rönnegård et al. (2010) used penalized quasi-likelihood. Bayesian inference on DHGLM allows us to estimate the different effects and parameters in the model and estimate their uncertainty by means of the joint posterior distribution.

Because of the different structured terms and effects in DHGLM, model fitting can become a daunting task. Popular methods such as Markov chain Monte Carlo (MCMC)

could take a long time and many simulations to converge. The integrated nested Laplace approximation (INLA) (Rue, Martino and Chopin, 2009) approach is an appealing option because of its short computation time. However, DHGLM are a class of models that are not currently implemented in the R-INLA package for the R programming language.

In order to fit models that cannot be currently implemented in INLA, given their specific structure, there have been some developments to combine INLA with other methods, such as Markov chain Monte Carlo (MCMC) methods (Gómez-Rubio and Rue, 2018; Gómez-Rubio, 2020), and importance sampling (IS) and adaptive multiple IS (AMIS) algorithms (Berild et al., 2022).

In this chapter, we propose fitting DHGLM with the use of the AMIS-INLA approach (Berild et al., 2022). We show how this class of models can be fitted in this way, providing specific details for the implementation of the algorithms in the cases where variables following Gaussian, Poisson and negative binomial distributions are modelled. Lee and Noh (2012) described ways of modelling the variance of the random effects for DHGLM. We have focused on modelling the precision instead, but the approach presented here can also be used to model the variance or standard deviation, if required.

In Section 6.2, we describe the class of models that DHGLM encompass. In Section 6.3, we offer some details about the INLA estimation approach. In Section 6.4, we present the proposed method that uses the IS/AMIS with INLA to fit DHGLM. Section 6.5 includes simulation studies for three different models considered to illustrate our proposals. Section 6.6 includes some applications to real data examples. In these two sections, we also include some diagnostics and, for each case, specific details on the implementation of the proposed method. In Section 6.7, we provide comparison on the computation times for the two approaches obtained for each of the fitted models. We end this chapter with a discussion in Section 6.8.

Finally, we believe it is important to mention that the work presented in this chapter is a collaboration with Virgilio Gómez-Rubio and Vicente Núñez-Antón and that it has been recently published (Morales-Otero, Gómez-Rubio and Núñez-Antón, 2022).

6.2 Double Hierarchical Generalized Linear Models

Suppose Y_i, for $i = 1, \ldots, n$, are random variables following a distribution from the exponential family (McCullagh and Nelder, 1989). That is, their probability distribution function can be written as:

$$f(y_i; \theta_i, \phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right\}, \tag{6.1}$$

where y_i is the observation corresponding to the variable Y_i, θ_i is a parameter, ϕ is a known positive constant value labelled as the scale or dispersion parameter, and $b(\cdot)$ and $c(\cdot)$ are given known functions.

It is known that $\mathrm{E}(Y_i) = \mu_i = b'(\theta_i)$ and that $\mathrm{Var}(Y_i) = \phi V(\mu_i)$, with $V(\mu_i) = b''(\theta_i)$ being a variance function. Different forms for ϕ and $V(\mu_i)$ for some known distributions are included in Table 6.1, where σ^2 is the variance parameter for the normal distribution,

n_i is the number of observations on each trial for the binomial distribution and k is the dispersion parameter or size of the negative binomial distribution.

Table 6.1: Different form of ϕ and $V(\mu_i)$ for some known distributions.

Distribution	ϕ	$V(\mu_i)$
Normal	$\phi = \sigma^2$	$V(\mu_i) = 1$
Poisson	$\phi = 1$	$V(\mu_i) = \mu_i$
Negative binomial	$\phi = 1$	$V(\mu_i) = \mu_i + k^{-1}\mu_i^2$
Binomial	$\phi = 1$	$V(\mu_i) = \mu_i \left(\frac{n_i - \mu_i}{n_i}\right)$

For example, for a variable Y_i having a negative binomial distribution, its probability mass function can be specified as:

$$f(y_i; p_i, k) = P(Y_i = y_i) = \binom{y_i + k - 1}{y_i} p_i^k (1 - p_i)^{y_i}, \tag{6.2}$$

where p_i is the probability of success on a Bernoulli trial, with $0 < p_i < 1$, for $i = 0, \ldots, n$, and y_i would represent the number of failures before the k-th success occurs. If the parameter k is considered fixed, this distribution belongs to the exponential family (Agresti, 2002).

In order to illustrate how to obtain ϕ and $V(\mu_i)$ for this distribution, let us take the exponential of the logarithm of $f(y_i; p_i, k)$, so that:

$$\exp\{\log(f(y_i; p_i, k))\} = \exp\left\{k \log(p_i) + y_i \log(1 - p_i) + \log\binom{y_i + k - 1}{y_i}\right\} \tag{6.3}$$

Here, if we consider that $\theta_i = \log(1 - p_i)$, then we would have that $\log(p_i) = \log(1 - e^{\theta_i})$ and the equation above becomes:

$$\exp\{\log(f(y_i; p_i, k))\} = \exp\left\{y_i\theta_i + k \log(1 - e^{\theta_i}) + \log\binom{y_i + k - 1}{y_i}\right\} \tag{6.4}$$

From equation (6.4), we can obtain $b(\theta_i)$ and ϕ, so that $b(\theta_i) = -k \log(1 - e^{\theta_i})$ and $\phi = 1$. Taking into account that $\theta_i = \log(1 - p_i)$, the first derivative of $b(\theta_i)$ is:

$$b'(\theta_i) = \frac{ke^{\theta_i}}{1 - e^{\theta_i}} = \frac{k(1 - p_i)}{p_i} \tag{6.5}$$

In addition, the second derivative is:

$$b''(\theta_i) = \frac{ke^{\theta_i}}{(1 - e^{\theta_i})^2} = \frac{k(1 - p_i)}{p_i^2} \tag{6.6}$$

Let us recall that, from equation (6.1), for distributions belonging to the exponential family, the mean is given by $\mathrm{E}(Y_i) = \mu_i = b'(\theta_i)$, hence, in this case we have that $\mathrm{E}(Y_i) =$

$\mu_i = \frac{k(1-p_i)}{p_i}$. Additionally, the variance was $\text{Var}(Y_i) = \phi V(\mu_i)$, with $V(\mu_i) = b''(\theta_i)$, so that, for the negative binomial we have $V(\mu_i) = \frac{k(1-p_i)}{p_i^2}$. If we take into account that the mean is $\mu_i = \frac{k(1-p_i)}{p_i}$, then we can write p_i in terms of μ_i, so that $p_i = \frac{k}{\mu_i+k}$. Therefore, $V(\mu_i)$ can be rewritten in terms of μ_i as:

$$V(\mu_i) = \frac{k(1-p_i)}{p_i^2} = \frac{\mu_i}{p_i} = \frac{\mu_i}{k/(\mu_i+k)} = \frac{\mu_i(\mu_i+k)}{k} = \mu_i + k^{-1}\mu_i^2 \tag{6.7}$$

Note that, if we consider $p_i = \frac{k}{\mu_i+k}$ in the distribution function of the negative binomial in equation (6.2), we obtain this distribution in the same form as was shown in Section 2.2, in Chapter 2. Also note that, in Chapter 2, we labelled the dispersion parameter for the negative binomial as τ, whereas here we have labelled it in a different manner (i.e., k), so that it can be clearly distinguished from the precision parameters that we will employ in Sections 6.5 and 6.6.

Let us also illustrate this process for the binomial distribution, with probability mass function given by:

$$f(y_i; p_i, n_i) = P(Y_i = y_i) = \binom{n_i}{p_i} p_i^{y_i}(1-p_i)^{n_i - y_i}, \tag{6.8}$$

where $0 < p_i < 1$, for $i = 0, \ldots, n$. Following the same process as before, that is, taking the exponential of the logarithm of $f(y_i; p_i, n_i)$, we obtain:

$$\begin{aligned} \exp\{\log(f(y_i; p_i, n_i))\} &= \exp\left\{ y_i \log(p_i) + (n_i - y_i)\log(1-p_i) + \log\binom{n_i}{p_i} \right\} \\ &= \exp\left\{ y_i \log\left(\frac{p_i}{1-p_i}\right) + n_i \log(1-p_i) + \log\binom{n_i}{p_i} \right\} \end{aligned} \tag{6.9}$$

If we consider here that $\theta_i = \log\left(\frac{p_i}{1-p_i}\right)$, then $p_i = \frac{e^{\theta_i}}{e^{\theta_i}+1}$ and $\log(1-p_i) = \log\left(\frac{1}{e^{\theta_i}+1}\right) = -\log(1+e^{\theta_i})$. Therefore, equation (6.9) becomes:

$$\exp\{\log(f(y_i; p_i, n_i))\} = \exp\left\{ y_i\theta_i - n_i \log(1+e^{\theta_i}) + \log\binom{n_i}{p_i} \right\}$$

From the equation above, we can obtain $b(\theta_i)$ and ϕ, so that $b(\theta_i) = n_i \log(1+e^{\theta_i})$ and $\phi = 1$. The first derivative of $b(\theta_i)$ is:

$$b'(\theta_i) = \frac{n_i e^{\theta_i}}{1+e^{\theta_i}} = n_i p_i \tag{6.10}$$

In addition, the second derivative of $b(\theta_i)$ is:

$$b''(\theta_i) = \frac{n_i e^{\theta_i}(1+e^{\theta_i}) - e^{\theta_i} n_i e^{\theta_i}}{(1+e^{\theta_i})^2} = \frac{n_i e^{\theta_i}}{(1+e^{\theta_i})^2} = n_i p_i (1-p_i) \tag{6.11}$$

Since, for the distributions belonging to the exponential family, the mean is $\mathrm{E}(Y_i) = \mu_i = b'(\theta_i)$, then, for the binomial distribution we have that $\mathrm{E}(Y_i) = \mu_i = n_i p_i$. Additionally, the variance is $\mathrm{Var}(Y_i) = \phi V(\mu_i)$, with $V(\mu_i) = b''(\theta_i)$, hence, here we have $V(\mu_i) = n_i p_i(1-p_i) = \mu_i(1-p_i)$. Taking into account that the mean is $\mu_i = n_i p_i$, we can write $p_i = \frac{\mu_i}{n_i}$, so that we obtain:

$$V(\mu_i) = \mu_i(1-p_i) = \mu_i\left(1 - \frac{\mu_i}{n_i}\right) = \mu_i\left(\frac{n_i - \mu_i}{n_i}\right) \tag{6.12}$$

In order to obtain the form of ϕ and $V(\mu_i)$ for the normal and the Poisson distributions, shown in Table 6.1, similar processes to the ones we have illustrated for the case of the negative binomial and the binomial distributions can be carried out.

A generalized linear model (GLM) (McCullagh and Nelder, 1989) is defined when a regression model is specified for the mean via a link function $g(\cdot)$, obtaining a linear predictor for the i-th observation, so that:

$$g(\mu_i) = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \tag{6.13}$$

where $\boldsymbol{x}_i$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of unknown regression parameters to be estimated.

GLMs were further extended by Lee and Nelder (2006) by proposing the DHGLM, which are specified given two sets of random effects $u_i^{(\mu)}$ and $u_i^{(\phi)}$, for $i = 1, \ldots, n$, so that the conditional mean and variance of the response variables Y_i are $\mathrm{E}(Y_i \mid u_i^{(\mu)}, u_i^{(\phi)}) = \mu_i$ and $\mathrm{Var}(Y_i \mid u_i^{(\mu)}, u_i^{(\phi)}) = \phi_i V(\mu_i)$, respectively, for $i = 1, \ldots, n$. The random effects depend on the variance (or precision) parameters $\lambda_i > 0$ and $\alpha_i > 0$, for $i = 1, \ldots, n$, i.e., $(u_i^{(\mu)}(\lambda_i), u_i^{(\phi)}(\alpha_i))$. Here, regression models for the mean, for the dispersion parameters and for the parameters of the random effects are specified, so that:

$$\begin{array}{rcl} g^{(\mu)}(\mu_i) & = & \boldsymbol{x}_i^{(\mu)\top}\boldsymbol{\beta}^{(\mu)} + \boldsymbol{z}_i^{(\mu)\top} u_i^{(\mu)} \\ g^{(\lambda)}(\lambda_i) & = & \boldsymbol{x}_i^{(\lambda)\top}\boldsymbol{\beta}^{(\lambda)} \\ g^{(\phi)}(\phi_i) & = & \boldsymbol{x}_i^{(\phi)\top}\boldsymbol{\beta}^{(\phi)} + \boldsymbol{z}_i^{(\phi)\top} u_i^{(\phi)} \\ g^{(\alpha)}(\alpha_i) & = & \boldsymbol{x}_i^{(\alpha)\top}\boldsymbol{\beta}^{(\alpha)}, \end{array} \tag{6.14}$$

where $\boldsymbol{x}_i^{(\cdot)}$ are vectors of explanatory variables included in the regression structures, with corresponding vectors of unknown coefficients $\boldsymbol{\beta}^{(\cdot)}$, for μ, λ, ϕ and α. In addition, $\boldsymbol{z}_i^{(\cdot)}$ are vectors of explanatory variables included in the regression structures for μ and ϕ.

As we have previously mentioned, estimation of this model can be done by using the H-likelihood proposed by Lee and Nelder (2006), and also penalized quasi-likelihood proposed by Rönnegård et al. (2010). Bayesian methods have been widely employed to fit highly parametrized hierarchical models in the context of DHGLM (see, for example, Bonner et al., 2021, and references therein). In Cepeda-Cuervo, Córdoba and Núñez-Antón (2018) and Morales-Otero and Núñez-Antón (2021), the authors used MCMC methods to fit generalized overdispersion models, where regression structures depending on some covariates were specified both for the mean and for the dispersion parameters.

6.3 Integrated Nested Laplace Approximation

The integrated nested Laplace approximation (INLA) was first proposed by Rue, Martino and Chopin (2009) to provide fast approximate Bayesian inference for latent Gaussian Markov random field (GMRF) models. Given a set of n observed variables Y_i, for $i = 1, \ldots, n$, usually with a distribution from the exponential family, the density of Y_i, $i = 1, \ldots, n$, may depend on some vector of hyperparameters $\boldsymbol{\theta_1}$. In addition, the mean of Y_i, $\mathrm{E}(Y_i)$, will be linked to a linear predictor η_i on the covariates using a convenient link function $g(\cdot)$ so that $g[\mathrm{E}(Y_i)] = \eta_i$.

The linear predictor may include different terms, as fixed and/or random effects, so that the distribution of all these terms is a GMRF with zero mean and precision matrix $Q(\boldsymbol{\theta_2})$, that may depend on some other vector of hyperparameters $\boldsymbol{\theta_2}$. To simplify notation, we will often use $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ to refer to the vector of all the hyperparameters. In addition, the vector of latent effects will be denoted by $\boldsymbol{\zeta}$.

In a Bayesian framework, the aim is to compute the posterior distribution of the latent effects and hyperparameters, $p(\boldsymbol{\zeta}, \boldsymbol{\theta} \mid \mathcal{D})$, and to be able to make inference about them. Here, $\mathcal{D}$ represents the available data, which will include the vector of observations for the response, $\mathbf{y} = (y_1, \ldots, y_n)$, and, possibly, other covariates required to define the fixed and random effects in the latent GMRF. Using Bayes' rule, this joint posterior distribution can be written as:

$$p(\boldsymbol{\zeta}, \boldsymbol{\theta} \mid \mathcal{D}) \propto L(\mathcal{D} \mid \boldsymbol{\zeta}, \boldsymbol{\theta}) p(\boldsymbol{\zeta}, \boldsymbol{\theta})$$

Here, $L(\mathcal{D} \mid \boldsymbol{\zeta}, \boldsymbol{\theta})$ represents the likelihood of the data, while $p(\boldsymbol{\zeta}, \boldsymbol{\theta})$ is the joint prior distribution of the latent effects and hyperparameters. This is often expressed as $p(\boldsymbol{\zeta}, \boldsymbol{\theta}) = p(\boldsymbol{\zeta} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$. Note that $p(\boldsymbol{\zeta} \mid \boldsymbol{\theta})$ is a GMRF and $p(\boldsymbol{\theta})$ is often defined as the product of univariate distributions as hyperparameters are considered to be independent a priori.

The joint posterior distribution $p(\boldsymbol{\zeta}, \boldsymbol{\theta} \mid \mathcal{D})$ is often highly multivariate and difficult to estimate. For this reason, Rue, Martino and Chopin (2009) focused on estimating the marginal posterior distributions of each of the latent effects ζ_j in $\boldsymbol{\zeta}$ and hyperparameters θ_l in $\boldsymbol{\theta}$. In this way, approximations $\tilde{p}(\zeta_j \mid \mathcal{D})$ and $\tilde{p}(\theta_l \mid \mathcal{D})$ for $p(\zeta_j \mid \mathcal{D})$ and $p(\theta_l \mid \mathcal{D})$, respectively, are obtained.

In addition, INLA can be used to obtain an approximation to the marginal likelihood of the model, $p(\mathcal{D})$, which is often difficult to compute. Other important quantities for model selection and model choice are available in the R-INLA package that implements the INLA method (Gómez-Rubio, 2020).

As discussed in the next section, INLA cannot fit DHGLM directly, but INLA can be embedded into the model fitting process to be able to easily fit these models.

6.4 Model Fitting

DHGLM do not fall into the class of models that INLA can fit due to their particular structure, which includes different hierarchies on the mean and scale parameters. How-

ever, as explained below, DHGLM can be expressed as conditional latent GMRF models after conditioning on some model parameters. This idea of fitting conditional models with INLA has been exploited by several authors (see, for example, Gómez-Rubio and Rue, 2018) to increase the number of models that can be fitted with INLA.

In particular, the vector of hyperparameters $\boldsymbol{\theta}$ can be decomposed into two sets of parameters $\boldsymbol{\theta}_c$ and $\boldsymbol{\theta}_{-c}$, so that the model, conditional on $\boldsymbol{\theta}_c$, can be fitted with INLA. The posterior distribution of $\boldsymbol{\theta}_c$ can then be expressed as

$$p(\boldsymbol{\theta}_c \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{\theta}_c) p(\boldsymbol{\theta}_c)$$

Here, $p(\boldsymbol{\theta}_c)$ is the prior on $\boldsymbol{\theta}_c$, which is known, and $p(\mathcal{D} \mid \boldsymbol{\theta}_c)$ is the *conditional* (on $\boldsymbol{\theta}_c$) marginal likelihood, as this is obtained after integrating out all the other hyperparameters and latent effects. This quantity can be easily obtained with INLA, so that the posterior distribution of $p(\boldsymbol{\theta}_c \mid \mathcal{D})$ can be then estimated.

Regarding the other hyperparameters $\boldsymbol{\theta}_{-c}$ and the latent effects, their marginal posterior distributions can be obtained by noting that:

$$p(\cdot \mid \mathcal{D}) = \int p(\cdot, \boldsymbol{\theta}_c \mid \mathcal{D}) d\boldsymbol{\theta}_c = \int p(\cdot \mid \boldsymbol{\theta}_c, \mathcal{D}) p(\boldsymbol{\theta}_c \mid \mathcal{D}) d\boldsymbol{\theta}_c$$

The conditional posterior marginal $p(\cdot \mid \boldsymbol{\theta}_c, \mathcal{D})$ is provided by INLA when fitting the model (after conditioning on $\boldsymbol{\theta}_c$).

In practice, an approximation for $p(\cdot \mid \mathcal{D})$ is obtained by weighing the posterior conditional marginals, that is:

$$\tilde{p}(\cdot \mid \mathcal{D}) = \sum_{m=1}^{M} \tilde{p}(\cdot \mid \boldsymbol{\theta}_c^{(m)}, \mathcal{D}) w_m$$

Here, M represents a number of ensembles of values of $\boldsymbol{\theta}_c^{(m)}$, $\{\boldsymbol{\theta}_c^{(m)}\}_{m=1}^{M}$, which is used for numerical integration. In addition, w_m are weights that can be computed in different ways, depending on how the values of $\boldsymbol{\theta}_c$ have been obtained.

For this specific purpose, Gómez-Rubio and Rue (2018) used the Metropolis-Hastings algorithm to estimate the distribution of $\boldsymbol{\theta}_c$, and also used the resulting values to estimate the remainder of the latent effects and hyperparameters. This algorithm requires fitting a model with INLA at each iteration of the Metropolis-Hastings algorithm, which makes it less appealing in practice. Moreover, the average of the posterior conditional marginal distributions is computed by using the corresponding weights $w_m = 1/M,\ m = 1, \ldots, M$.

Similarly, Berild et al. (2022) used the importance sampling (IS) algorithm instead, which can be run in parallel and provides reduced computing times. In this particular case, samples of $\boldsymbol{\theta}_c$ are obtained by using an importance distribution $s(\cdot)$ to obtain $\{\boldsymbol{\theta}_c^{(m)}\}_{m=1}^{M}$. For each value $\boldsymbol{\theta}_c^{(m)}$, a conditional model is fitted with INLA, so that integration weights w_m are obtained as follows:

$$w_m \propto \frac{p(\mathcal{D} \mid \boldsymbol{\theta}_c) p(\boldsymbol{\theta}_c^{(m)})}{s(\boldsymbol{\theta}_c^{(m)})}$$

Weights are re-scaled so that they sum up to one. Furthermore, Berild et al. (2022) described the use of the adaptive multiple importance sampling (AMIS) (Corneut et al., 2012) algorithm that provides a more robust sampling method that updates the importance distribution $s(\cdot)$.

Regarding model fitting of DHGLM with INLA, we can use IS and AMIS with INLA by conditioning on some of the model hyperparameters or latent effects. These will depend on the way in which the DHGLM is defined. Both Gómez-Rubio and Rue (2018) and Berild et al. (2022) discussed different approaches on how to best select the parameters in $\boldsymbol{\theta}_c$. In the simplest cases, the choice of $\boldsymbol{\theta}_c$ will be clear as just a few parameters will need to be fixed to obtain a conditional latent GMRF model. For highly structured models, it may happen that, after conditioning on some hyperparameters or latent effects, two or more conditionally independent submodels appear (see, for example, Lázaro, Armero and Gómez-Rubio, 2020). These submodels can be fitted independently with INLA. All the different cases are illustrated in Section 6.5, where different simulations studies are developed in detail on different types of models.

However, in order to provide a more general approach to the choice of $\boldsymbol{\theta}_c$, we propose the use of a graphical representation of the model. This graphical model encodes conditional independence relationships among the model parameters, so that its structure can be exploited to be able to select the best possible choice of the parameters to be included in $\boldsymbol{\theta}_c$ (see, for example, Cowell et al., 1999). See Section 6.5.4 for more details and a thorough discussion about this graphical representation using the examples in the simulation study conducted in Section 6.5.

Regarding the sampling distribution for $\boldsymbol{\theta}_c$, Berild et al. (2022) suggested choosing a multivariate Gaussian distribution or a multivariate t distribution with a small number of degrees of freedom for continuous variables. Note that some of the variables in $\boldsymbol{\theta}_c$ may need to be re-scaled (e.g., a precision will be sampled in the log scale). Hence, the mean and precision of these distributions are updated at each adaptive step. For discrete variables, the choice is not so clear. When the variables are dichotomous, Berild et al. (2022) suggested using a binomial distribution for each of them, so that their probabilities depend on some fixed effects (which are the parameters updated after each adaptive step).

The choice of the parameters of the sampling distribution is crucial to obtain a good performance of the proposed methodology. The initial parameters of the distribution could be based on summary statistics of the observed data as rough estimates. For example, for continuous data, when the sampling distribution is a multivariate normal, the mean can be set to the sample mean of the observed data and the precision can be diagonal with *large* values in the diagonal. Here *large* must be put into context according to the scale of the parameters. Too large values of the precision will imply that the parameter space is not conveniently explored, while too small values will imply that samples with a very small posterior density will be sampled too often. In both cases, bad estimates will be obtained at the adaptive steps that can result in the algorithm requiring more steps to produce reliable estimates. This issue is thoroughly discussed in the simulation studies in Section 6.5, and the examples in Section 6.6.

In addition, as a general guidance, the conditional model can be fitted with INLA given the set of possible values for the mean of the sampling distribution before running AMIS with INLA. Different sets of values can be tested and the marginal likelihoods compared. The one with the highest value of the marginal likelihood may be a better candidate as it improves model fitting. This will help to be able to choose an initial sampling distribution whose mode is close to the posterior mode of $\boldsymbol{\theta}_c$, so that less adaptive steps (and, hence, simulations) are required to obtain good estimates.

Another way of assessing the performance of IS with INLA is to compute the effective sample size and conduct graphical diagnostics, as discussed in Berild et al. (2022). The effective sample size can be estimated as:

$$n_e = \frac{(\sum_{m=1}^{M} w_m)^2}{\sum_{m=1}^{M} w_m^2}$$

Note that this effective sample size will be the same for all the components of $\boldsymbol{\theta}_c$ as it is only based on the weights and not on the sampled values.

Graphical diagnostics can be produced for each variable in $\boldsymbol{\theta}_c$ by re-ordering the sampled values in ascending order and comparing the estimated cumulative probability (i.e., the cumulative sum of the re-ordered weights) with the empirical cumulative probabilities $1/M, \ldots, M/M$, respectively. A straight line means that the estimated posterior marginal of that specific parameter is reliable.

Monitoring the convergence of the algorithm could be conducted in a number of ways. First of all, the effective sample size could be computed and the algorithm can be stopped once the desired sample size has been attained. The conditional marginal likelihood fitted at the mean of the sampling distribution after each adaptive step could also be monitored to assess whether it keeps increasing or approaches a certain value (at this point the algorithm can be stopped). It is worth noting that more samples could be obtained when needed by simply resuming the simulations using updated estimates of the parameters of the sampling distribution.

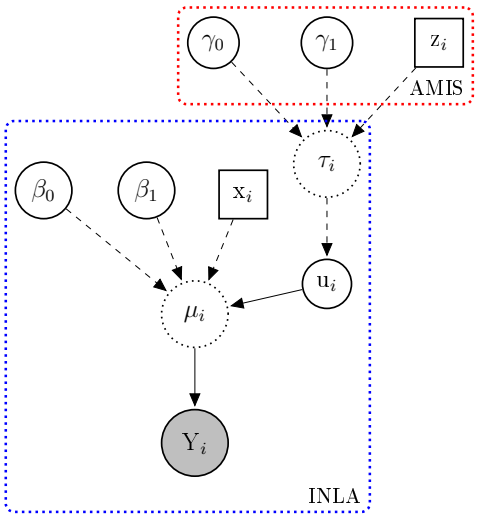
6.5 Simulation study

In this section, we carry out three different simulation studies to illustrate model fitting of hierarchical models with different structures. In Section 6.5.1, we fit a Poisson log-linear model with random effects, in which the log-precision of the random effects is modelled using a linear term; in Section 6.5.2, we fit a negative binomial model in which the log-size parameter is modelled using a linear term; and in Section 6.5.3, we fit a Gaussian model to grouped data in which the log-precision of each group is modelled using a linear mixed-effects model. In all cases, models are fit using MCMC and AMIS with INLA. IS with INLA has not been considered because Berild et al. (2022) showed that, in general, AMIS-INLA has a better performance than IS with INLA.

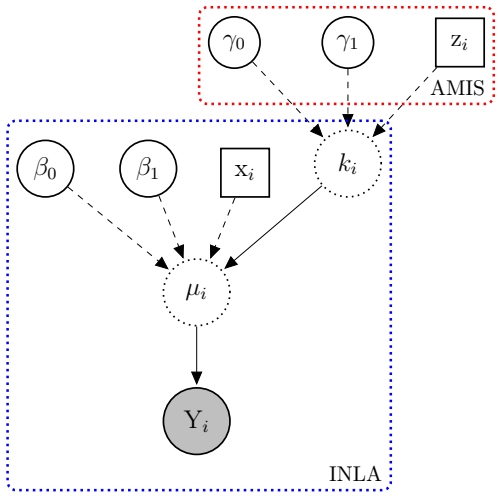
The aim of these simulation studies is twofold. On the one hand, we would like to illustrate the way in which IS and AMIS with INLA can be implemented and how the conditioning effects $\boldsymbol{\theta}_c$ can be chosen. On the other hand, it is important to compare the

results obtained with these methods to a gold standard. In our case, we have fitted the models using Markov chain Monte Carlo (Brooks et al., 2011) using the JAGS software via the R-package rjags (Plummer, 2021).

Figure 6.1 shows the representation of these models as graphical models. In addition to the different elements of the model, the conditioning parameters have been highlighted (using a red dotted box) to illustrate which parameters are estimated using AMIS. The marginals of all the other parameters are obtained by averaging the conditional marginals resulting after fitting the conditional models with INLA. Nodes in a shaded solid circle represent the observed response, nodes in a solid square represent the observed independent variables, nodes in a white solid circle represent model effects and parameters and nodes in a dotted white circle represent deterministic nodes (i.e., their values are fully determined by the values at their parent nodes). Parameters fitted with AMIS with INLA are in a red dotted box, whereas the conditional model fitted with INLA is in a blue dotted box.

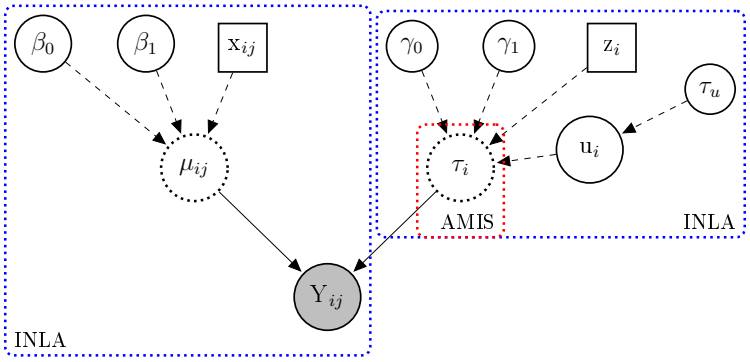


(a) Poisson model with random effects with different precisions.



(b) Negative binomial model with regression on the log-sizes.

Figure 6.1: Graphical representation of the models fitted in the simulation study in Section 6.5.



(c) Gaussian model with regression model on the likelihood log-precisions.

Figure 6.1: Graphical representation of the models fitted in the simulation study in Section 6.5 (Continued).

When implementing AMIS, the importance distribution $s(\cdot)$ is assumed as a multivariate Gaussian in all examples. Note that this means that some parameters may be transformed so that simulations are feasible. For example, precisions will be sampled in the log-scale, so that samples from the log-precision are obtained. In all cases we have computed results using a very vague distribution (with zero mean and large precision) and another distribution based on a rough estimate of the parameters of the sampling distribution from the observed data. This should reduce the number of iterations required to obtain a reliable model fitting. In all cases, the estimates from AMIS with INLA are compared with MCMC estimates.

In all the examples presented below the same number of simulations has been used. When fitting the model using AMIS with INLA, 5000 iterations have been used in the initial step, followed by 10 new adaptive steps with 1000 simulations each. For MCMC, a burn-in of 10000 simulations is used, plus 100000 simulations of which only one in 100 is retained, leading to a final number of 1000 samples. In addition, in the Gaussian example in Section 6.5.3, different scenarios have also been tested (see below for details). Finally, simulations have been carried out on a Linux Ubuntu 18.2 cluster using 60 cores Intel(R) Xeon(R) CPU E5-2683 v4 @ 2.10GHz.

6.5.1 Poisson model with random effects with different precisions

The first simulation study is based on a Poisson log-linear model with fixed and random effects, so that the precision of the random effects is modelled using a linear term with covariates. In particular, the model is:

$$\begin{aligned} Y_i \mid u_i &\sim \text{Poi}(\mu_i),\ i = 1, \ldots, n \\ \log(\mu_i) &= \beta_0 + \beta_1 x_i + u_i \\ u_i &\sim N(0, \tau_i),\ \tau_i > 0 \\ \log(\tau_i) &= \gamma_0 + \gamma_1 z_i \\ \beta_0, \beta_1 &\sim N(0, 0.001) \\ \gamma_0, \gamma_1 &\sim N(0, 0.001) \end{aligned}$$

Note that the Gaussian distribution $N(\cdot, \cdot)$ is defined in terms of the mean and precision, so that τ_i represents the precision of the Gaussian distribution of the random effects. A Poisson distribution with random effects is often used to model overdispersed data (Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón, 2012).

This model can be expressed as a latent GMRF by conditioning on $\boldsymbol{\theta}_c = \boldsymbol{\gamma} = (\gamma_0, \gamma_1)$, resulting in a Poisson model with random effects with different precisions. This is illustrated in the graphical representation of the model in Figure 6.1(a). Hence, this model will be fitted using AMIS with INLA and values of $\boldsymbol{\gamma}$ will be obtained by simulation. Estimates of the posterior distribution can be obtained by using importance weights and the posterior marginals of β_0 and β_1 will be obtained by weighting their conditional marginals.

For the simulated data, we have used $n = 1000$, $\beta_0 = 1$, $\beta_1 = 0.25$, $\gamma_0 = 0$ and $\gamma_1 = 0.5$. Covariate x_i has been simulated using a uniform distribution between 0 and 1, and covariate z_i has been simulated using a standard Gaussian distribution. Once these values have been set, the observed value y_i has been obtained by sampling from a Poisson distribution with the resulting mean μ_i.

The sampling distribution for $\boldsymbol{\gamma}$ is a bivariate Gaussian distribution. The initial value of the mean is vector $(0, 0)$ and the initial value of the variance matrix is a diagonal matrix with entries equal to 5 in the diagonal. This choice provides ample initial variability to explore the parametric space of $\boldsymbol{\gamma}$ conveniently, so that accurate estimates are obtained at the adaptive and final steps of AMIS with INLA.

Table 6.2 summarizes the estimates using the different methods and Figure 6.2 shows the posterior marginal estimates obtained with both methods. Here, the dashed vertical lines represent the true values of the parameters specified for the simulated data. As can be seen, the estimates obtained with AMIS with INLA and MCMC are very similar. The effective sample size n_e obtained with AMIS with INLA in this case is 9900.914.

Table 6.2: Summary of the estimates of the Poisson model with random effects with different precisions used in the simulation study.

		AMIS		MCMC	
Parameter	True value	Mean	St. dev.	Mean	St. dev.
β_0	1	1.0531	0.0736	1.049	0.0729
β_1	0.25	0.2302	0.1254	0.2347	0.1253
γ_0	0	-0.0210	0.0655	-0.0484	0.0654
γ_1	0.5	0.4830	0.0622	0.4787	0.0636

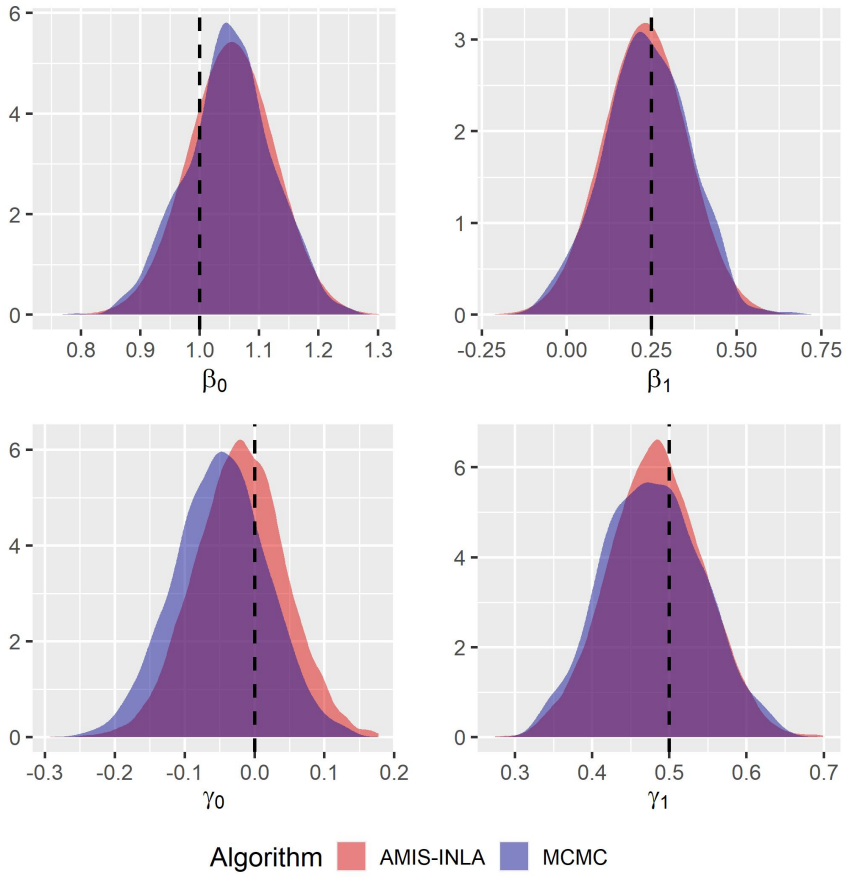


Figure 6.2: Posterior marginals of the estimated parameters obtained by fitting the Poisson model with random effects, using both the MCMC and AMIS-INLA methods. Vertical lines represent the actual values of the parameters used when simulating the data.

Finally, in Figure 6.3 we have included the graphical diagnostics for the parameters in the model fitted here. As we stated in Section 6.4, a straight line in this graphical diagnostic means that the estimates obtained for the parameters are reliable. Therefore, in this figure, it can be seen that model fitting seems to be reasonably good.

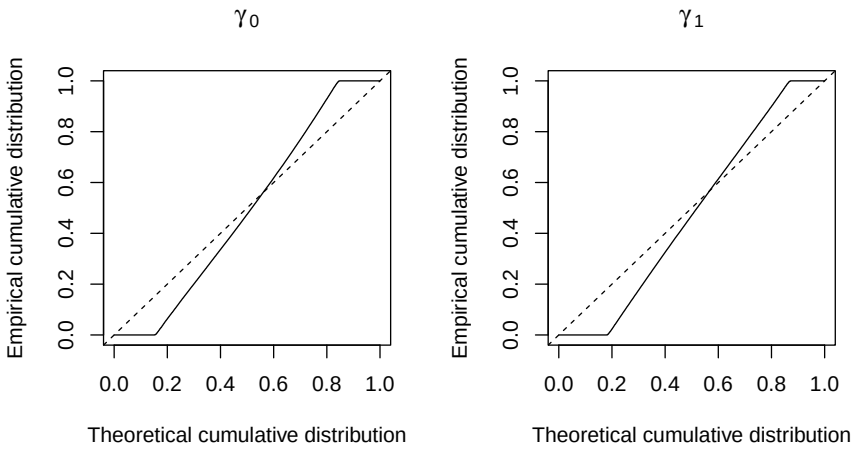


Figure 6.3: Graphical diagnostics for the Poisson model used in the simulation study.

6.5.2 Negative binomial with different sizes

The negative binomial distribution is also used to model overdispersed count data. In this simulation study the logarithm of the size parameter k_i of the negative binomial distribution depends on a linear term with covariates, which in turns makes the probability to be different across observations. In particular, the model is as follows:

$$\begin{aligned} Y_i &\sim \text{NB}(p_i, k_i) \\ p_i &= \tfrac{k_i}{k_i+\mu_i} \\ \log(\mu_i) &= \beta_0 + \beta_1 x_i \\ \log(k_i) &= \gamma_0 + \gamma_1 z_i \\ \beta_0, \beta_1 &\sim N(0, 0.001) \\ \gamma_0, \gamma_1 &\sim N(0, 0.001) \end{aligned}$$

Similarly, as in the previous example, this model can be expressed as a latent GMRF by conditioning on $\boldsymbol{\theta}_c = \boldsymbol{\gamma} = (\gamma_0, \gamma_1)$, resulting in a negative binomial model with different sizes. This is illustrated in the graphical representation of the model in Figure 6.1(b). When fitting the model with AMIS with INLA, values of $\boldsymbol{\gamma}$ will be obtained by simulation and their estimates will be computed using the importance weights. The posterior marginals of β_0 and β_1 will be obtained by weighting their conditional marginals.

For our study, $n = 500$ observations have been simulated. Covariate x_i is simulated from a uniform between 10 and 20, and covariate z_i has been simulated from a uniform between 0 and 20. Values of z_i have then been standardized before simulating the data. Regarding the model parameters, we have used $\beta_0 = 1$, $\beta_1 = 0.25$, $\gamma_0 = 0$ and $\gamma_1 = 5$. Once the mean and size of the negative binomial have been computed, the values of the response variable have been sampled using a negative binomial distribution.

Likewise, as in the previous simulation study, the sampling distribution for $\boldsymbol{\gamma}$ is a bivariate Gaussian distribution. The initial value of the mean is vector $(0, 0)$ and the

initial value of the variance matrix is a diagonal matrix with entries equal to 5 in the diagonal. This a convenient choice for this example as well and it provides good estimates of the model parameters (see below).

Table 6.3 summarizes the estimates using the different methods and Figure 6.4 shows the posterior marginal estimates obtained with both methods. As can be seen, the estimates obtained with AMIS with INLA and MCMC are very similar. The effective sample size n_e obtained with AMIS with INLA in this case is 9737.075.

Table 6.3: Summary of the estimates of the negative binomial model with different sizes used in the simulation study.

		AMIS		MCMC	
Parameter	True value	Mean	St. dev.	Mean	St. dev.
β_0	1	0.9875	0.0541	0.9893	0.0545
β_1	0.25	0.2506	0.0033	0.2505	0.0033
γ_0	0	-0.0879	0.0926	-0.0862	0.0931
γ_1	5	4.8594	0.1861	4.8568	0.1836

Figure 6.5 shows the graphical diagnostics for the parameters estimated in the model, where we can see that model fitting seems to be reasonably good for γ_0, whereas it is perhaps not as good for γ_1, since it slightly deviates from the dotted line. This may be due to the fact that, in this case, the initial parameters assumed for the sampling distribution (i.e., a multivariate Gaussian with mean vector $(0, 0)$) might be relatively far from the true value of γ_1 (i.e., $\gamma_1 = 5$). Hence, the first samples obtained for this parameter, corresponding to the first iterations of the algorithm, would also be far from its true value. As a consequence, the empirical cumulative distribution would differ from the theoretical one for these initial samples. In any case, as can be seen in Figure 6.5, the empirical cumulative distribution for γ_1, at some point after the value 0.4 for the theoretical distribution, gets considerably close to the dotted line. This would indicate that the algorithm has been able to obtain reasonable approximations of this parameter, a fact that can be corroborated by the results reported in Table 6.3, where these initial values provided estimates close to the true values of each of the two parameters.

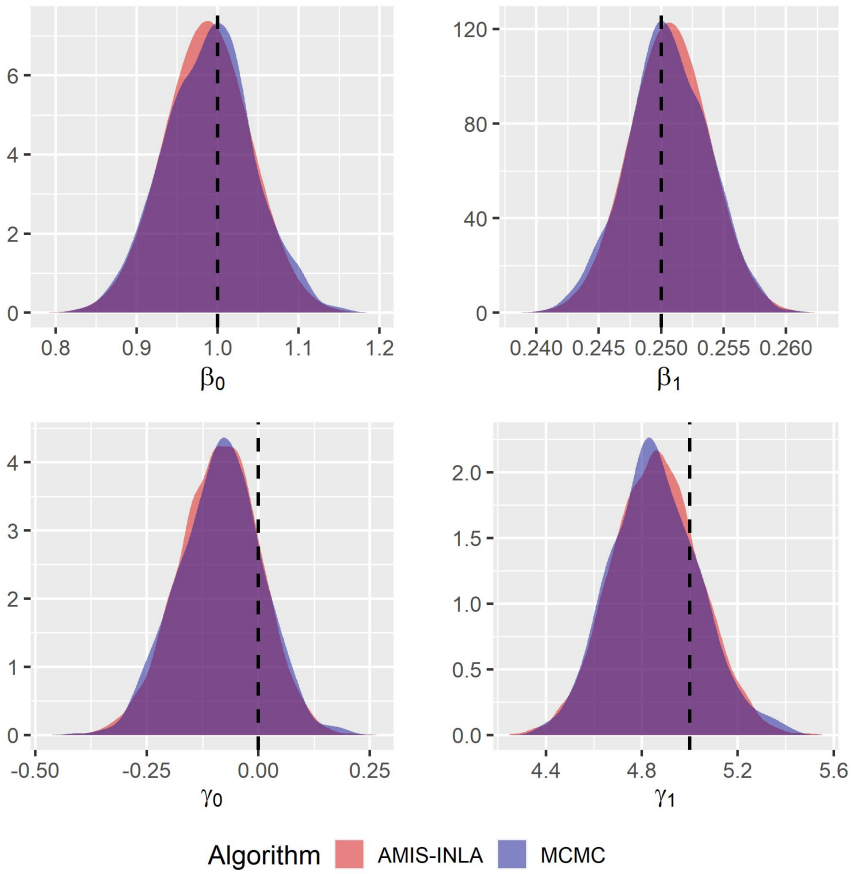


Figure 6.4: Posterior marginals of the estimated parameters obtained by fitting the negative binomial model with different sizes, using both the MCMC and AMIS-INLA methods. Vertical lines represent the actual values of the parameters used when simulating the data.

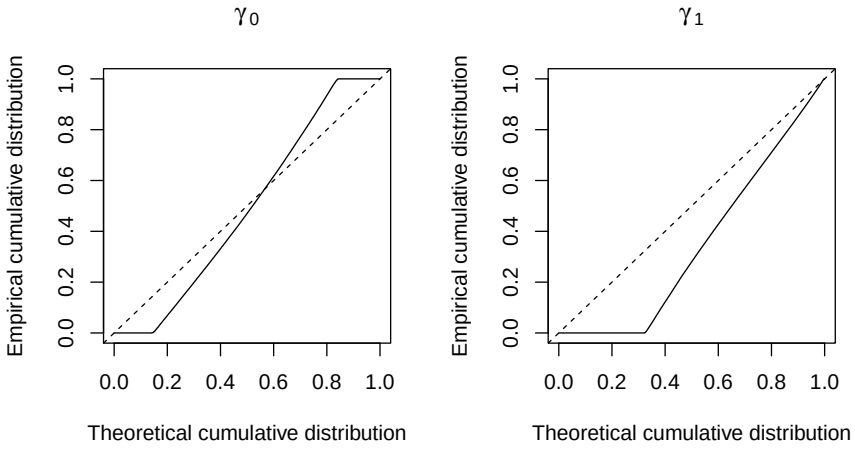


Figure 6.5: Graphical diagnostics for the negative binomial model used in the simulation study.

6.5.3 Gaussian model with different scale parameters

In the last simulation study we have considered the case of grouped Gaussian data so that each group has a different precision and the log-precision is modelled on a mixed-effects model. In particular, we consider the model:

$$\begin{aligned}
Y_{ij} \mid u_i &\sim N(\mu_{ij}, \tau_i),\ \tau_i > 0\ ;\ i = 1, \ldots, p;\ j = 1, \ldots, n_i \\
\mu_{ij} &= \beta_0 + \beta_1 x_{ij} \\
\log(\tau_i) &= \gamma_0 + \gamma_1 z_i + u_i \\
u_i &\sim N(0, \tau_u),\ \tau_u > 0 \\
\tau_u &\sim \text{Gamma}(1, 0.00005) \\
\beta_0, \beta_1 &\sim N(0, 0.001) \\
\gamma_0, \gamma_1 &\sim N(0, 0.001)
\end{aligned}$$

Here, p represents the number of groups and n_i the number of observations in group i. The values of the parameters used in the simulations are $\beta_0 = 1$, $\beta_1 = 0.25$, $\gamma_0 = 0$, $\gamma_1 = 5$ and $\tau_u = 1$. The total number of observations is 2500, which corresponds to $p = 5$ groups and $n_i = 500,\ i = 1, \ldots, p$. Furthermore, values of covariate x_{ij} have been simulated from a uniform distribution between 0 and 1, while values of covariate z_i have been obtained by sampling from a uniform distribution in the interval (-1, 1).

This model is a bit more complex because the log-precision depends on both fixed and random effects. Hence, conditioning on $\boldsymbol{\gamma}$ alone will not suffice to make this model a latent GMRF. It would be possible to condition on $\boldsymbol{\gamma}$ and $\mathbf{u} = (u_i, \ldots, u_p)$ but then the dimension of the parametric space may be difficult to handle by AMIS (in particular, when the value of the number of groups p is large). Furthermore, estimating the random effects u_i using importance sampling may be difficult, and we prefer INLA to perform this task.

Instead, conditioning will be on $\boldsymbol{\theta}_c = \boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$, which will split the main model into two independent submodels with response variables $\mathbf{y}$ and $\log(\boldsymbol{\tau})$, as illustrated in Figure 6.1(c). These two submodels can be fitted independently and the resulting log-marginal likelihood will be the sum of the corresponding values from the two submodels, which can be then used to compute the weights.

Note that, in this particular case, nodes $\tau_1, \ldots, \tau_p$ are not stochastic nodes as they are fully determined by $\boldsymbol{\gamma}$, z_i and u_i. For this reason, there is no prior for them. In order to ease the computations, and without loss of generality, we set $p(\tau_i) = 1,\ i = 1, \ldots, p$, which will not have any effect on the computation of the marginal likelihood.

It is worth mentioning that, among the three different examples provided in the simulation study, this one is an actual DHGLM as defined in Lee and Nelder (2006) because it includes random effects when modelling $\log(\tau_i)$. In order to explore convergence of the AMIS algorithm we have repeated the analysis using different sets of initial values for the parameters of the importance distribution and number of samples (see below). This will allow us to explore how the adaptive procedure in the AMIS algorithm behaves and to assess the resulting estimates precision. These scenarios are:

1. Initial step of 5000 iterations, 10 new adaptive steps with 1000 simulations each, vague initial parameters for the sampling distribution (AMIS-INLA1).

2. Initial step of 5000 iterations, 10 new adaptive steps with 1000 simulations each, parameters informed from the data for the sampling distribution (AMIS-INLA2).

3. Initial step of 1000 iterations, 10 new adaptive steps with 1000 simulations each, vague initial parameters for the sampling distribution (AMIS-INLA3).

4. Initial step of 1000 iterations, 10 new adaptive steps with 1000 simulations each, parameters informed from the data for the sampling distribution (AMIS-INLA4).

5. Initial step of 5000 iterations, 10 new adaptive steps with 1000 simulations each, vague initial parameters and large variance for the sampling distribution (AMIS-INLA5).

6. Initial step of 5000 iterations, 10 new adaptive steps with 5000 simulations each, parameters informed from the data for the sampling distribution (AMIS-INLA6).

In all the scenarios described above, the sampling distribution is a multivariate normal distribution for $(\log(\tau_1), \ldots, \log(\tau_p))$. Vague initial parameters refers to using a mean of 0 and a variance matrix that is diagonal with all entries equal to 5. Using a sampling distribution with parameters informed from the data refers to computing the sample variance of each group and computing the parameters of the sampling distribution from them. In particular, the mean is the log of the vector of sample variances and the variance matrix is diagonal with entries the variance of the log-sample variances divided by their corresponding values of n_i. If the scenario indicated that a larger variance for the sampling distribution has been used, these values are multiplied by 10. In

all cases these are initial values of the parameters of the sampling distribution and they will be updated at each adaptive step.

Table 6.4 summarizes the results of the estimation of the Gaussian model with the MCMC method. Similarly, Table 6.5 includes the results of the estimation of the Gaussian model with the AMIS-INLA method, where the different scenarios are considered. Estimation is good for all model parameters for most scenarios, with point estimates close to that of MCMC in most cases. However, estimates of τ_u do not seem to be good as AMIS with INLA tends to underestimate this parameter for scenario 3. The effective sample sizes of AMIS with INLA range from 5.12 (scenario 3, based on 11000 simulations) to 10444.68 (scenario 2, 15000 total simulations) and 51536.74 (scenario 6, based on a total of 55000 simulations). Hence, scenario 3 is likely to produce poor estimates due to its low effective sample size.

It is worth noting that the estimation of the posterior marginal of τ_u has been conducted by first averaging the posterior marginal of $\log(\tau_u)$ (the internal scale of this parameter in INLA) and then transforming the resulting marginal to obtain that of τ_u. The reason is that INLA estimates of the posterior marginal of τ_u were not reliable.

Furthermore, Figure 6.6 presents the estimates of the posterior marginal distributions for the parameters obtained with MCMC, as well as for the different settings of the AMIS-INLA algorithm. Here we can see that, for all the considered scenarios, similar posterior marginals were obtained, except for scenario 3, which considerably deviates from the other distributions. Let us recall that for this scenario, vague initial parameters were assumed for the sampling distribution and, in addition, for the initial step, only 1000 iterations were used. Therefore, for these specific settings, it seems that the algorithm is not able to properly approximate the marginal posterior distributions of the parameters.

Table 6.4: Summary of the estimates of the Gaussian model with different scale parameters used in the simulation study, obtained by fitting the model with MCMC.

		MCMC		
Parameter	True value	Mean	St. dev.	95 % CI
β_0	1	0.9884	0.0170	(0.9553, 1.0226)
β_1	0.25	0.2864	0.0288	(0.2284, 0.3421)
γ_0	0	-0.2926	0.3766	(-1.0764, 0.4184)
γ_1	5	3.8365	0.7832	(2.2486, 5.3023)
τ_u	1	1.9128	1.2140	(0.3247, 4.7935)

Table 6.5: Summary of the estimates of the Gaussian model with different scale parameters used in the simulation study, obtained by fitting the model with the AMIS algorithm and INLA, for the six different scenarios considered.

AMIS-INLA1

Parameter	True value	Mean	St. dev.	95 % CI
β_0	1	0.9877	0.0174	(0.9536,1.0216)
β_1	0.25	0.2874	0.0299	(0.2288,0.3458)
γ_0	0	-0.3628	0.4062	(-1.1867,0.4568)
γ_1	5	3.5299	0.7874	(1.9312,5.1173)
τ_u	1	2.1079	1.3224	(0.3525,5.4083)

AMIS-INLA2

Parameter	True value	Mean	St. dev.	95 % CI
β_0	1	0.9883	0.0163	(0.9563,1.0201)
β_1	0.25	0.2866	0.0280	(0.2317,0.3414)
γ_0	0	-0.2620	0.4270	(-1.1280,0.5994)
γ_0	5	3.9200	0.8279	(2.2390,5.5881)
τ_u	1	1.9268	1.2212	(0.3193,4.9866)

AMIS-INLA3

Parameter	True value	Mean	St. dev.	95 % CI
β_0	1	0.9778	0.0244	(0.9299,1.0257)
β_1	0.25	0.3028	0.0420	(0.2203,0.3850)
γ_0	0	-0.9351	0.8263	(-2.6105,0.7322)
γ_1	5	3.0919	1.5998	(-0.1586,6.3142)
τ_u	1	0.5086	0.3189	(0.0852,1.3045)

AMIS-INLA4

Parameter	True value	Mean	St. dev.	95 % CI
β_0	1	0.9882	0.0163	(0.9562,1.0202)
β_1	0.25	0.2866	0.0280	(0.2316,0.3415)
γ_0	0	-0.2625	0.4269	(-1.1282,0.5987)
γ_1	5	3.9173	0.8276	(2.2369,5.5849)
τ_u	1	1.9282	1.2221	(0.3195,4.9903)

AMIS-INLA5

Parameter	True value	Mean	St. dev.	95 % CI
β_0	1	0.9882	0.0163	(0.9562,1.0201)
β_1	0.25	0.2866	0.0280	(0.2316,0.3415)
γ_0	0	-0.2630	0.4272	(-1.1295,0.5987)
γ_1	5	3.9173	0.8282	(2.2356,5.5861)
τ_u	1	1.9254	1.2205	(0.3190,4.9835)

Table 6.5: Summary of the estimates of the Gaussian model with different scale parameters used in the simulation study, obtained by fitting the model with the AMIS algorithm and INLA, for the six different scenarios considered (Continued).

AMIS-INLA6

Parameter	True value	Mean	St. dev.	95 % CI
β_0	1	0.9882	0.0163	(0.9562,1.0201)
β_1	0.25	0.2866	0.0280	(0.2316,0.3415)
γ_0	0	-0.2625	0.4270	(-1.1286,0.5989)
γ_1	5	3.9170	0.8279	(2.2360,5.5852)
τ_u	1	1.9273	1.2220	(0.3193,4.9894)

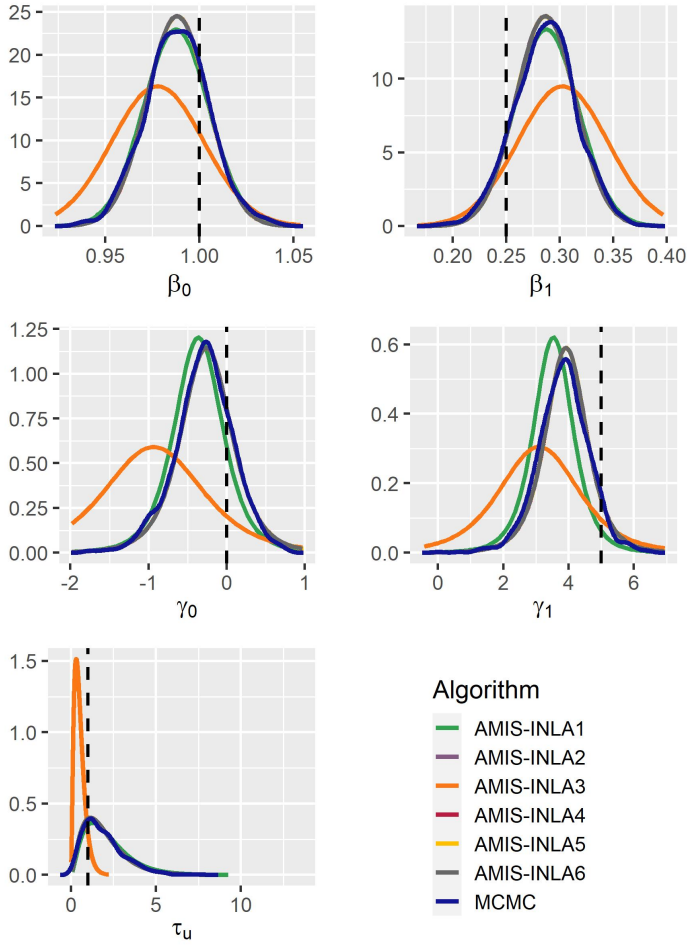


Figure 6.6: Estimates of the posterior marginals of the parameters obtained by fitting the Gaussian model with different scale parameters, using both the MCMC and AMIS-INLA methods, considering all scenarios for the AMIS-INLA algorithm setup.

In Figures 6.7, 6.8 and 6.9 we have included the graphical diagnostics for the parameters estimated in the models fitted in this section. With the exception of scenarios 1 and 3, for which vague initial sampling distributions were specified, model fitting seems to be very good.

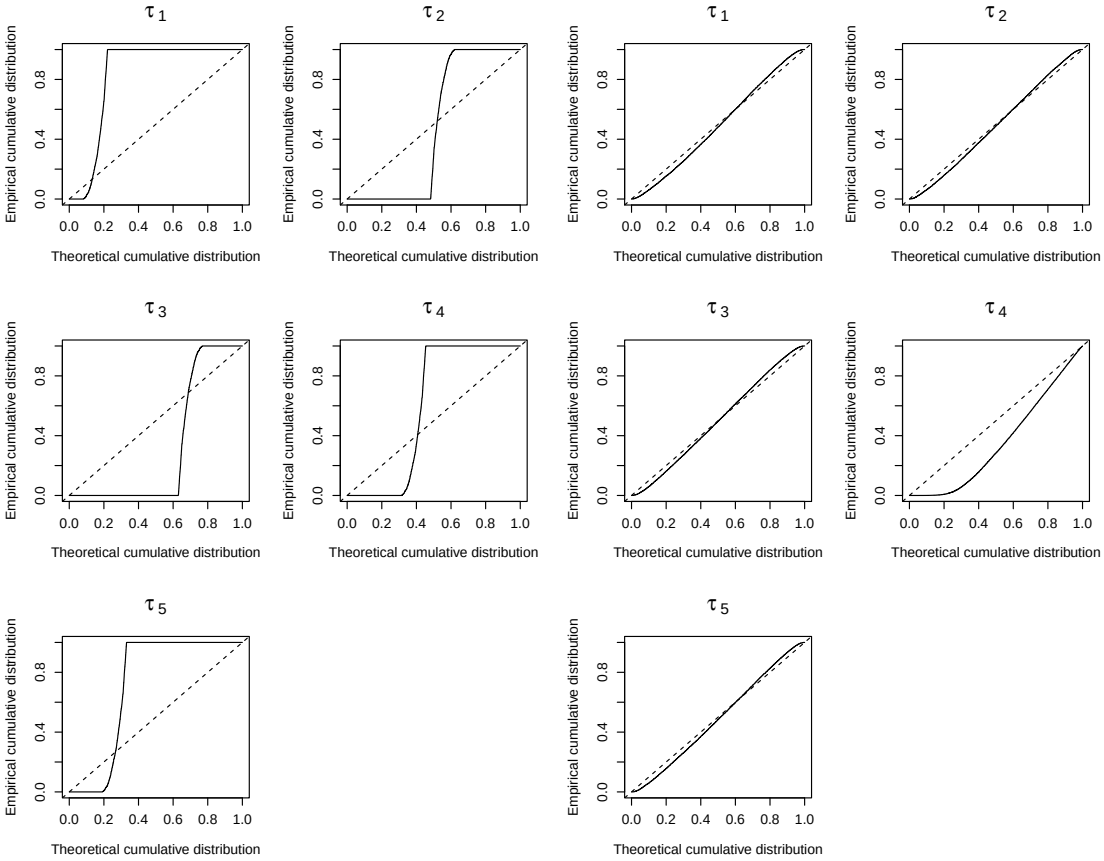


Figure 6.7: Graphical diagnostics for the Gaussian models in the simulation study under scenarios 1 (left panel) and 2 (right panel).

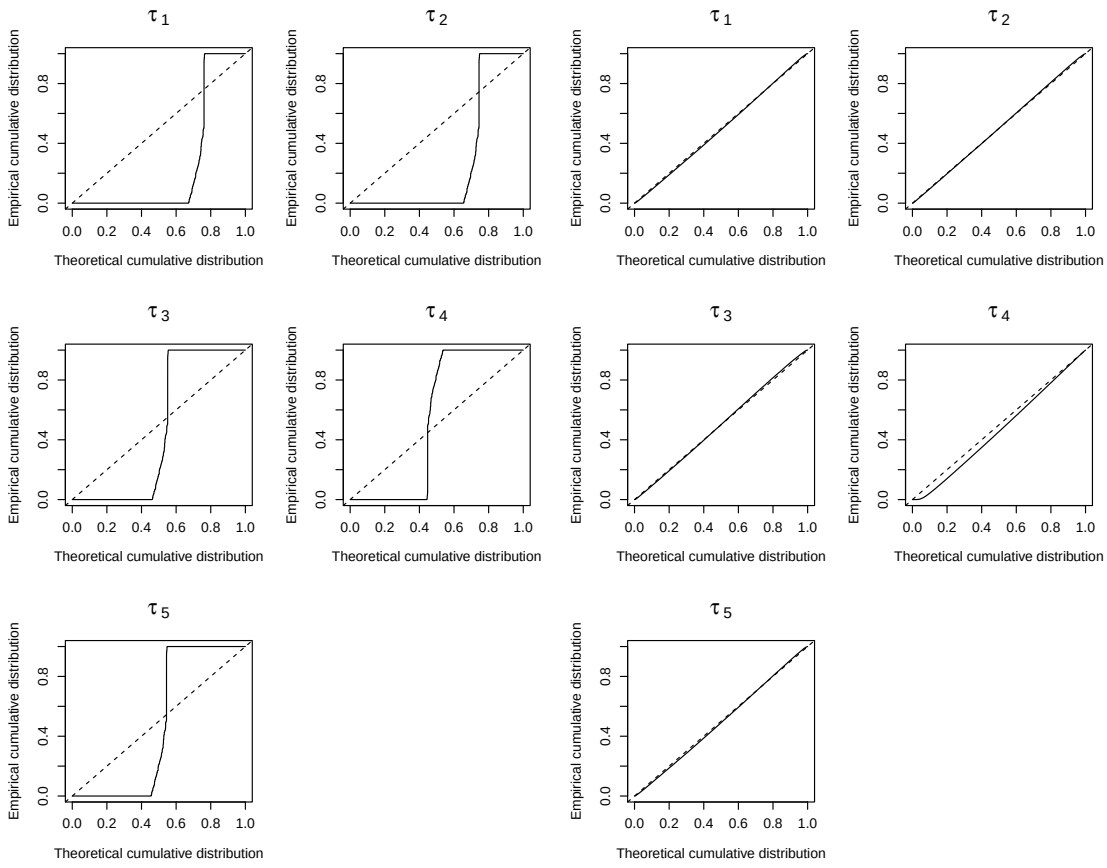


Figure 6.8: Graphical diagnostics for the Gaussian models in the simulation study under scenarios 3 (left panel) and 4 (right panel).

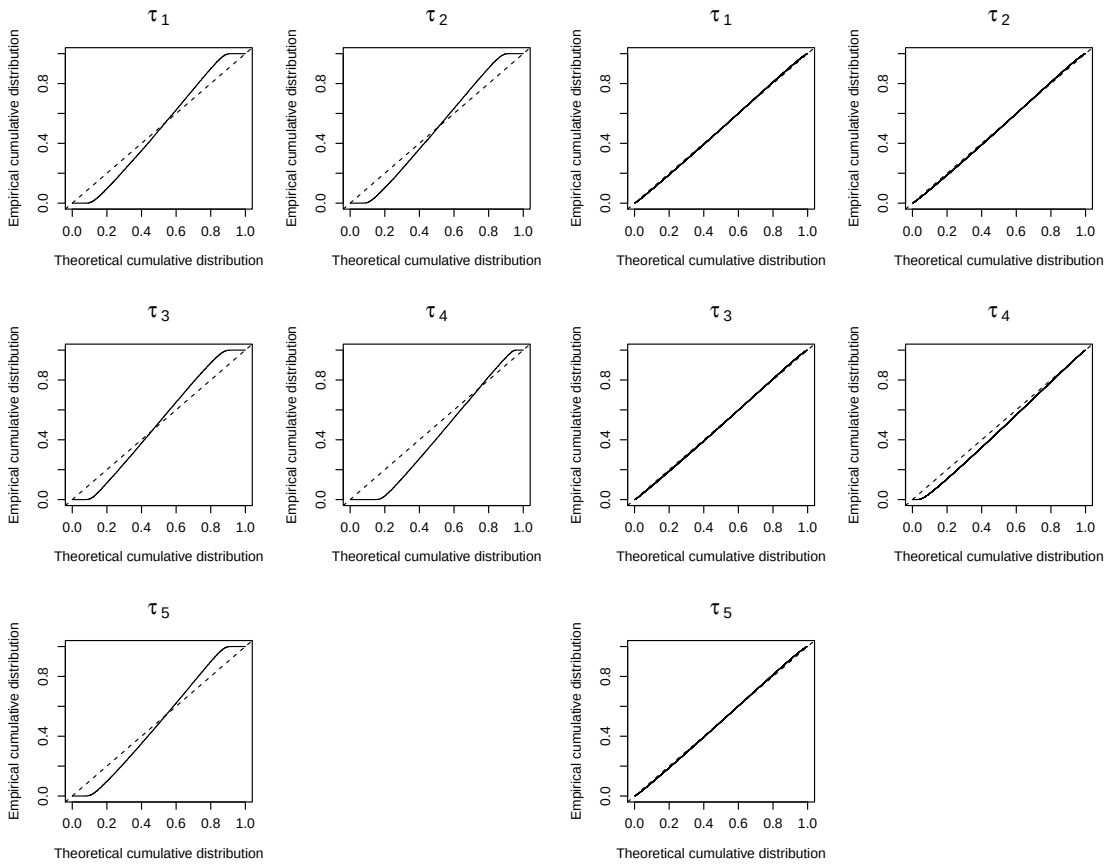


Figure 6.9: Graphical diagnostics for the Gaussian models in the simulation study under scenarios 5 (left panel) and 6 (right panel).

6.5.4 Summary of results

The simulation studies conducted above illustrate the use of AMIS with INLA to fit DHGLM. This approach will allow a flexible definition of the models using the R-INLA package as well as efficient model fitting. Given that AMIS can be run in parallel, DHGLM could be fitted in a short time provided a computer with a large number of CPUs is available (which is not uncommon these days).

Regarding the selection of the parameters in $\boldsymbol{\theta}_c$, we have provided new guidelines not discussed in Gómez-Rubio and Rue (2018) or Berild et al. (2022) by using the graphical representation of the models in Figure 6.1. By inspecting the graphical model, it is easier to find the parameters to condition on, so that the resulting model is a latent GMRF (see the Poisson and negative binomial models). Furthermore, for highly structured models, it is possible to split the model into more than one submodel (that are latent GMRF) by conditioning on a small sample of hyperparameters, as is the case of the Gaussian model with different scale parameters.

The parameters in $\boldsymbol{\theta}_c$ have been included in a red dotted box, which has been labelled AMIS as this is the method used to estimate the posterior distribution of these parameters. Similarly, the conditional latent GMRF has been included in a blue dotted line, which has been labelled as INLA because this is the method used to estimate the posterior marginals of the parameters in this conditional model.

In a nutshell, the parameters in $\boldsymbol{\theta}_c$ should be taken so that their dimension is as low as possible, preferable as part of coefficients of fixed effects or precisions of random effects, and so that they split the main model into one or more submodels that are easy to fitted with INLA. Choosing the random effects themselves as part of $\boldsymbol{\theta}_c$ should be avoided as it is difficult to sample efficiently using AMIS and their dimension is likely to increase with the size of the data.

6.6 Examples

In this section we illustrate model fitting of DHGLM with AMIS-INLA using two real data sets. In Section 6.6.1, we describe a Poisson model with random effects with a hierarchical structure on the precision and also a negative binomial model with a hierarchical structure on the size parameter to analyse infant mortality in Colombia. In Section 6.6.2, we fit a model with subject-level random slopes and precisions to participants in a sleep deprivation study.

6.6.1 Infant mortality in Colombia

The infant mortality data in Colombia that we analyse here was already presented and studied in Section 2.6, Chapter 2, a work also included in Morales-Otero and Núñez-Antón (2021). In addition, similar versions of this data set have been studied in previous research (see, for example, Quintero-Sarmiento, Cepeda-Cuervo and Núñez-Antón, 2012 or Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018).

The variables available in this data set are given for each of the $n = 32$ departments or regions of Colombia. Among them we can find the number of children under one year of age who died in year 2005 (**ND**), the total number of births in the same year (**NB**), an index that represents the percentage of people with their basic needs not satisfactorily attended for year 2005 (**IBN**) and the observed mortality rates, computed as the number of children under one year of age who died in 2005 per 1000 born alive (**Rates**).

As it was already seen in Section 2.6, this data presents overdispersion when fitting a Poisson regression model for the mortality rates, a phenomenon that arises when the real variance of the data is larger than the one specified in the model. Additionally, we have also found evidence of the presence of spatial autocorrelation present in the data. Therefore, these are issues that need to be taken into account if we wish to specify regression models for this data.

The first model considered is the generalized spatial conditional normal Poisson from Section 2.6.3, which is able to accommodate overdispersion and to explain spatial dependence. This model assumes that the variable representing the number of deaths

in each region (ND_i), conditioned on the set of values it takes in the neighbouring regions without including the i-th region itself ($\text{ND}_{\sim i}$), and on a set of normally distributed random effects $u_i \sim N(0, \tau_i)$, with $\tau_i > 0$, follows a Poisson distribution, so that $(\text{ND}_i \mid \text{ND}_{\sim i}, u_i) \sim \text{Poi}(\mu_i)$, for $i = 1, \ldots, n$.

This model allows the dispersion parameter to vary according to explanatory variables or any other terms by specifying a regression model for the variance of the random effect. It is also able to explain the spatial association which may be present in the data by including the spatial lag of the rates in the regression model for the mean or in the model for the dispersion as well (see, for example, Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018 or Morales-Otero and Núñez-Antón, 2021).

The connection with DHGLM appears here because we can model the log-precisions using a linear predictor on **IBN** so that $\log(\tau_i) = \gamma_0 + \gamma_1 \text{IBN}_i$, for $i = 1, \ldots, n$. It is worth mentioning that, in this particular case, the precisions are univocally determined by the linear predictor.

Following the example from Morales-Otero and Núñez-Antón (2021), we have specified the following model:

$$\begin{aligned}
(\text{ND}_i \mid \text{ND}_{\sim i}, u_i) &\sim \text{Poi}(\mu_i) \\
\log(\mu_i) &= \log(\text{NB}_i) + \beta + \rho \mathbf{W}_i \mathbf{Rates} + u_i \\
u_i &\sim N(0, \tau_i),\ \tau_i > 0 \\
\log(\tau_i) &= \gamma_0 + \gamma_1 \text{IBN}_i \\
\beta, \rho &\sim N(0, 0.001) \\
\gamma_0, \gamma_1 &\sim N(0, 0.001),
\end{aligned}$$

where $\mathbf{W}_i$ is the i-th row of a row-standardized spatial neighbourhood matrix $\mathbf{W}$. Adjacency here is defined so that two regions are neighbours if they share at least one point of their boundaries. Therefore, $\mathbf{W}_i\mathbf{Rates}$ is the spatial lag of the observed mortality rates, which in this case represents the average of **Rates** from the neighbours.

In the implementation of AMIS with INLA we have taken $\boldsymbol{\theta}_c = \boldsymbol{\gamma} = (\gamma_0, \gamma_1)$. The sampling distribution is a bivariate Gaussian with vector mean $(0, 0)$ and the variance matrix is a diagonal matrix with entries equal to 5. In this case, 5000 simulations were initially run, followed by 10 adaptive steps with 1000 simulations each.

Results of the estimation of this model are shown in Table 6.6 and Figure 6.10. As can be seen, AMIS-INLA and MCMC produce close results. The effective sample size of AMIS with INLA is 9263.002.

Table 6.6: Summary of the estimates of the generalized spatial conditional normal Poisson model with random effects and varying dispersion fitted to the infant mortality data in Colombia.

	AMIS		MCMC	
Parameter	Mean	St. dev.	Mean	St. dev.
β	-4.9124	0.2306	-4.8987	0.2310
ρ	0.0427	0.0094	0.0421	0.0095
γ_0	4.1951	0.6392	4.1893	0.6033
γ_1	-0.0423	0.0148	-0.0421	0.0140

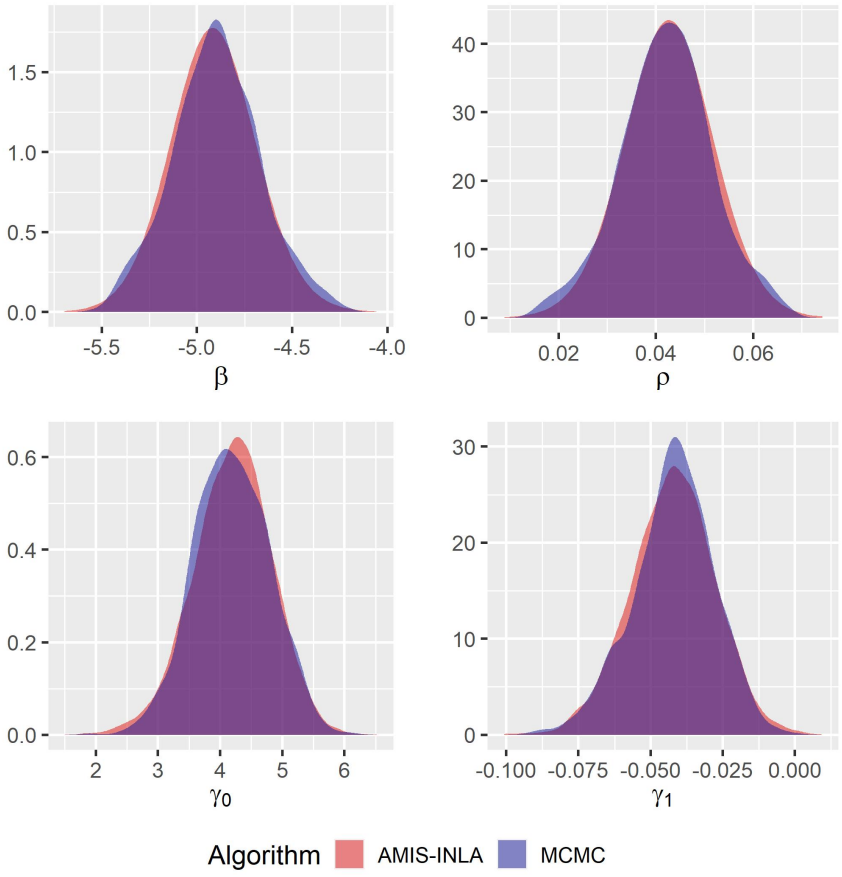


Figure 6.10: Posterior marginals of the estimated parameters obtained by fitting the generalized spatial conditional normal Poisson model to the infant mortality data in Colombia, using both the MCMC and AMIS-INLA methods.

The negative binomial model could be another option to consider in order to fit the infant mortality data described here. Therefore, we have specified the generalized spatial

conditional negative binomial model (Cepeda-Cuervo, Córdoba and Núñez-Antón, 2018), where it is assumed that $(\text{ND}_i \mid \text{ND}_{\sim i}) \sim \text{NB}(\mu_i, k_i)$, with μ_i being the conditional mean and k_i the size parameter of a negative binomial distribution. For this model, we can specify regression structures both for the mean and dispersion parameters, which can include the spatial lag of the rates and explanatory variables as well.

In particular, we have fitted the following model:

$$\begin{aligned}
(\text{ND}_i \mid \text{ND}_{\sim i}) &\sim \text{NB}(\mu_i, k_i) \\
\log(\mu_i) &= \log(\text{NB}_i) + \beta + \rho \mathbf{W}_i \mathbf{Rates} \\
\log(\text{k}_i) &= \gamma_0 + \gamma_1 \text{IBN}_i \\
\beta, \rho &\sim N(0, 0.001) \\
\gamma_0, \gamma_1 &\sim N(0, 0.001)
\end{aligned}$$

In order to fit this model with AMIS with INLA we have also taken $\boldsymbol{\theta}_c = \boldsymbol{\gamma} = (\gamma_0, \gamma_1)$. Conditional on $\boldsymbol{\theta}_c$, the resulting model is a negative binomial with different known sizes, which is easy to fit with INLA. Sampling has been done as with the Poisson distribution.

Table 6.7 and Figure 6.11 display the results of the estimation of this model, which show that AMIS with INLA provides very similar results to MCMC. The effective sample size of AMIS with INLA is 9717.207 now.

Table 6.7: Summary of the estimates of the generalized spatial conditional negative binomial model with varying dispersion fitted to the infant mortality data in Colombia.

	AMIS-INLA		MCMC	
Parameter	Mean	St. dev.	Mean	St. dev.
β	-4.8871	0.2341	-4.8933	0.2427
ρ	0.0425	0.0094	0.0423	0.0099
γ_0	4.2547	0.6235	4.2553	0.6191
γ_1	-0.0452	0.0142	-0.0454	0.0139

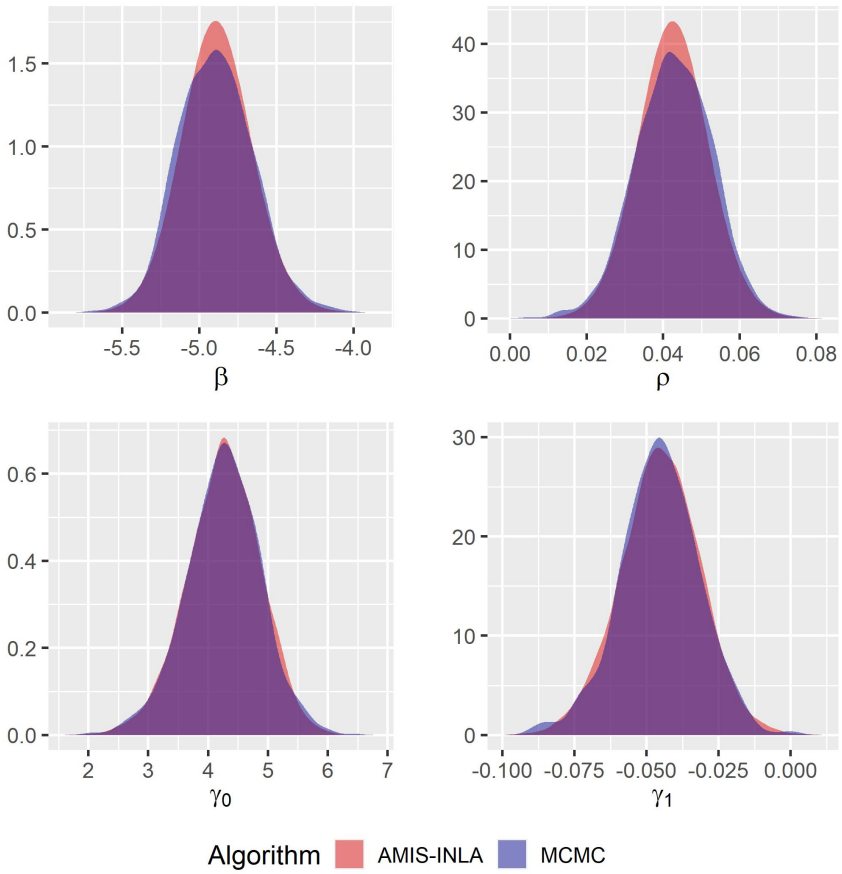


Figure 6.11: Posterior marginals of the estimated parameters obtained by fitting the generalized spatial conditional negative binomial model to the infant mortality data in Colombia, using both the MCMC and AMIS-INLA methods.

Figure 6.12 presents the graphical diagnostics for the parameters in the Poisson and negative binomial models fitted in this section. In general, model fitting is very similar for both models and, in addition, it is reasonably good.

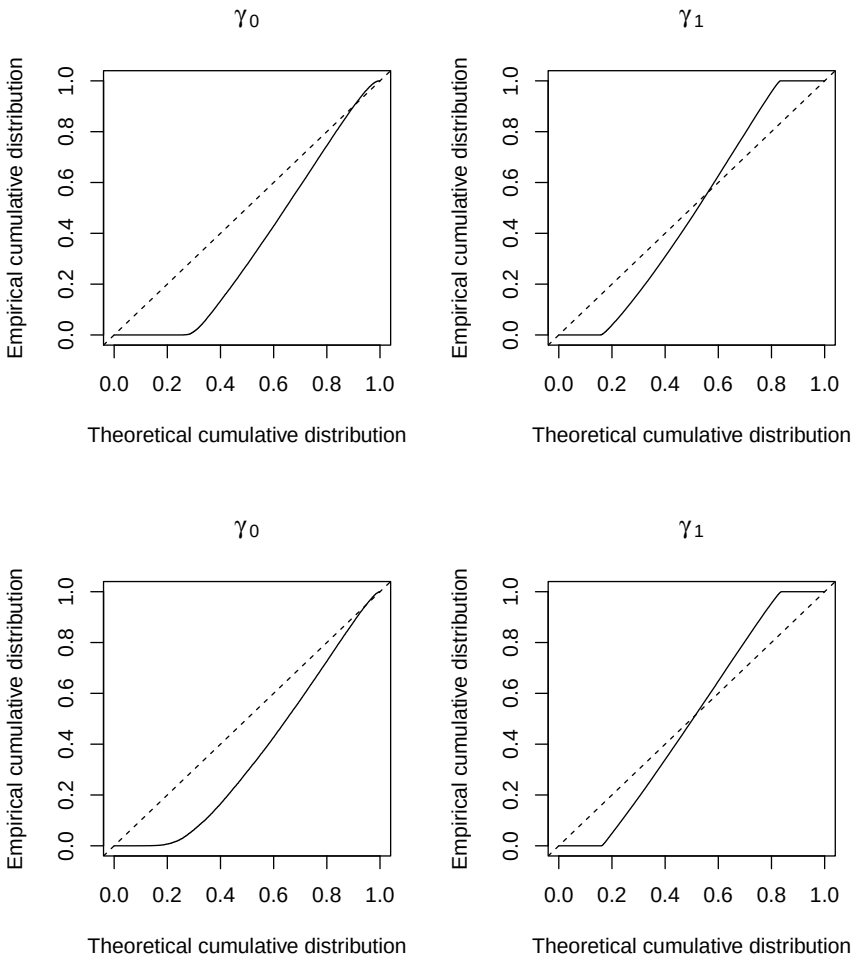


Figure 6.12: Graphical diagnostics for the Poisson model (top panel) and the negative binomial model (bottom panel) fitted to the infant mortality data in Colombia.

6.6.2 Sleep deprivation study

Belenky et al. (2003) conducted an experiment to measure the effect of sleep deprivation on reaction time on a number of subjects. A subset of this data set is included in the R package lme4 (Bates et al., 2015) and it includes observations for the most sleep-deprived group for the first 10 days of the study. This data set has been analysed by different authors (see, for example, Gómez-Rubio, 2020) using linear mixed-effects with random slopes as the number of days under sleep deprivation seems to have a different effect on the different subjects.

Figure 6.13 includes the subject-specific reaction times together with their corresponding linear regression lines. It also illustrates the fact that variability of the reaction

times among subjects is not uniform, with some subjects having a broader range of values than others. For this reason, we have fitted a model with random slopes per subject, where the within subject measurements precision is assumed to be different when using a DHGLM.

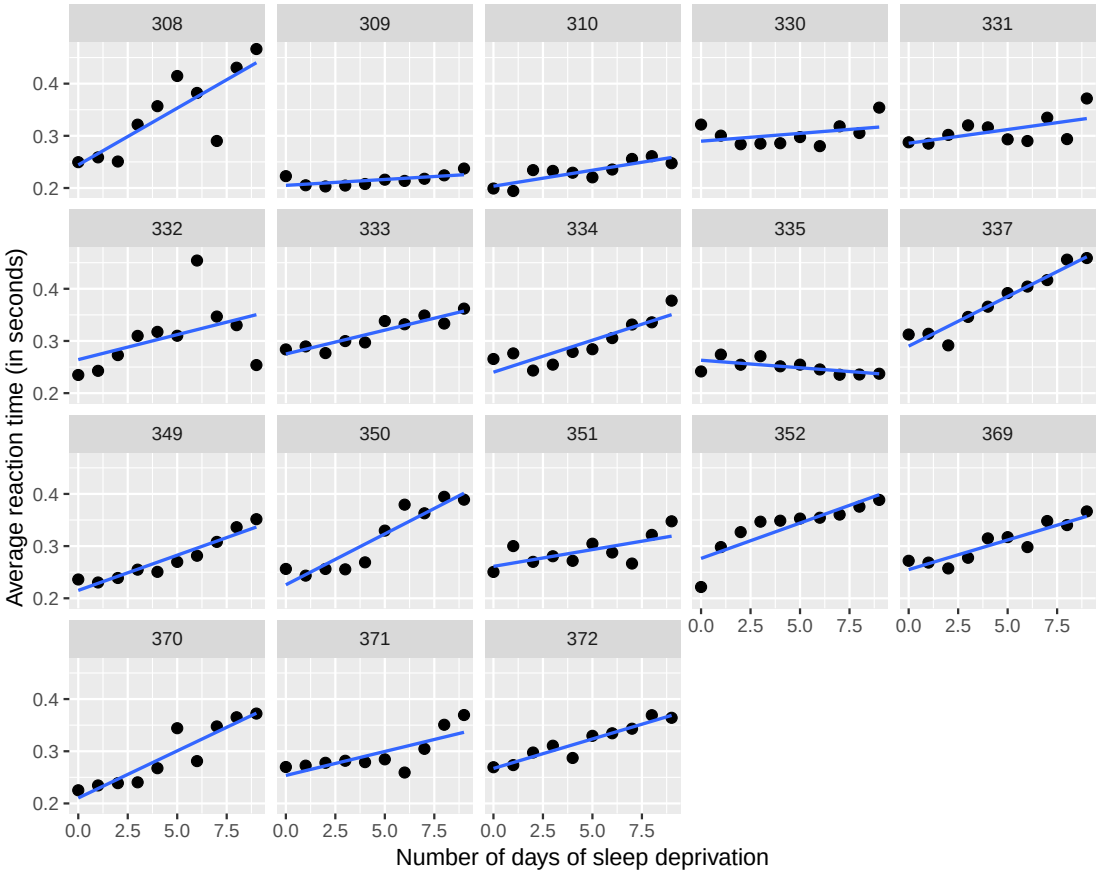


Figure 6.13: Effect of the number of days under sleep deprivation on the different subjects (based on code available in the lme4 package).

In particular, we have fitted the following model:

$$\begin{aligned}
Y_{ij} \mid u_i &\sim N(\mu_{ij}, \tau_i),\ \tau_i > 0\ ;\ i = 1, \ldots, p;\ j = 1, \ldots, n_i \\
\mu_{ij} &= \beta_0 + \beta_i \text{day}_{ij} \\
\log(\tau_i) &= \gamma + u_i \\
\beta_i &\sim N(0, \tau_\beta) \\
u_i &\sim N(0, \tau_u),\ \tau_u > 0 \\
\tau_\beta &\sim \text{Gamma}(1, 0.00005) \\
\tau_u &\sim \text{Gamma}(1, 0.00005) \\
\beta_0 &\sim N(0, 0.001) \\
\gamma &\sim N(0, 0.001)
\end{aligned} \tag{6.15}$$

Here, $p = 18$ is the number of subjects and $n_i = 10,\ i = 1, \ldots, p$ given that all subjects have the same number of measurements in the data set. Covariate day_{ij} is the number of the days since the beginning of the sleep deprivation experiment. Note that $\beta_i,\ i = 1, \ldots, p$, refers to random coefficients to allow for different per-subject slopes. It should be emphasized that this model is similar to the one in Section 6.5.3 and that it will be fitted in a similar way, i.e., by sampling from $(\log(\tau_1), \ldots, \log(\tau_p))$. Note that the dimension of the parametric space is 18, which may be large for algorithms such as IS and AMIS. A graphical representation of this model is shown in Figure 6.14.

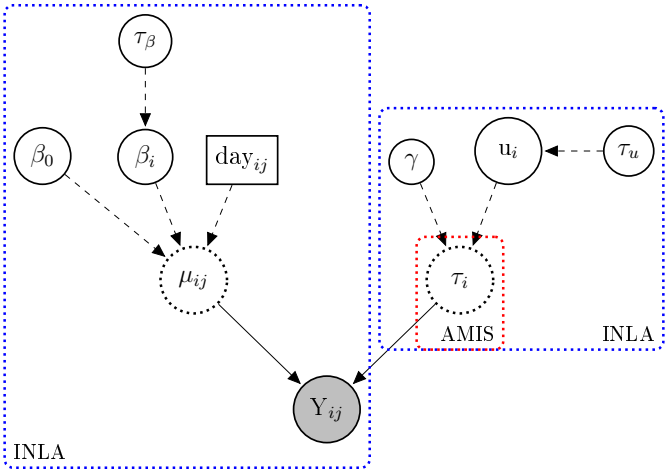


Figure 6.14: Graphical representation of the Gaussian model fitted in the sleep deprivation example.

In order to select the parameters of the importance distribution we have proposed different approaches. Initially, we assumed a multivariate normal distribution with zero mean and a diagonal precision matrix with entries equal to 5 along the diagonal. This provided a vague starting sampling distribution for the log-precisions that after a few adaptation steps may get close to the actual posterior distribution. Unfortunately, this provided very poor estimates and the results were discarded.

We noticed that importance sampling may not be efficient if the mean of the importance distribution is far from the posterior modes and also when its variance is too large. For this reason, we propose to use the data to obtain some rough estimates of the posterior mean and precisions based on S_i^2, the sample variance computed using measurements from subject i. Then, the mean of the importance distribution is $(\log(1/S_1^2), \ldots, \log(1/S_p^2))$ and the variance is diagonal with entries $0.05 \cdot (\log(1/S_1^2), \ldots, \log(1/S_p^2))$. In principle, this should provide a starting sampling distribution which is close to the posterior modes and with a variance in the scale of the posterior variances that allows for short jumps during the adaptive steps.

However, we noticed that we could obtain better initial parameters by performing permutations of the values of $(\log(1/S_1^2), \ldots, \log(1/S_p^2))$, fitting the conditional model and checking the values of the conditional marginal likelihood, so that the permutation

with the highest value is used to set the parameters of the initial sampling distribution. This simple prior step produced means of the sampling distribution that were very close to the posterior mode of $(\log(\tau_1), \ldots, \log(\tau_p))$. In particular, 500 random permutations were tested prior to running AMIS with INLA.

For all the models fitted in this example, AMIS with INLA has been run using an initial adaptive step based on 1000 simulations followed by 20 adaptive steps with 1000 simulations each. MCMC is based on 10000 burn-in simulations followed by 100000 simulations, of which only 1 in 100 has been kept, so that inference is based on 1000 samples.

Results of the estimation of this model are provided in Table 6.8 and the densities of the posterior estimations for the parameters are shown in Figure 6.15. The effective sample size of AMIS with INLA in this case is 2.015619, which is small but seems to provide good estimates of the marginals of the model parameters. It is worth mentioning that we have computed the effective sample size after each adaptive step and that it reached the value 81.14083 after 12 adaptation steps. AMIS with INLA could be stopped after a certain effective sample size has been achieved. It is worth noting that the estimates of $\log(\tau_i)$ did not change considerably in the last adaptive steps.

Table 6.8: Summary of the estimates of the Gaussian model with random slopes for each subject fitted to the sleep study data.

	AMIS		MCMC	
Parameter	Mean	St. dev.	Mean	St. dev.
β_0	0.2606	0.0034	0.2589	0.0042
τ_β	8240.2229	2742.75	8002.245	2898.704
γ	7.3170	0.2003	7.2612	0.2115
τ_u	2.1222	0.9348	2.6565	2.2836

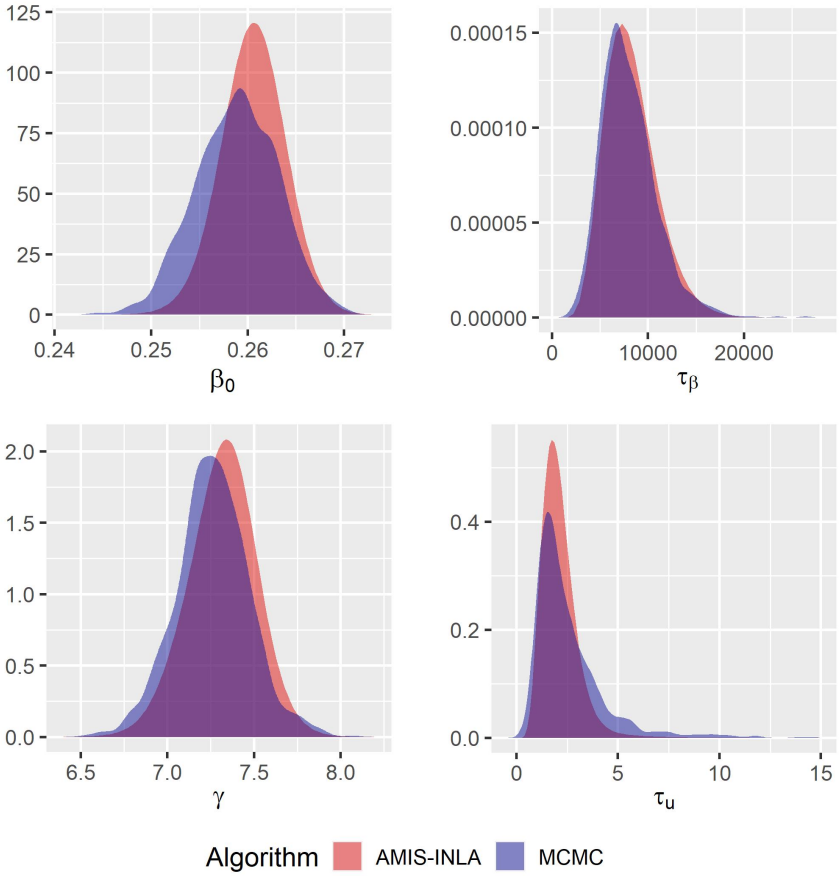


Figure 6.15: Posterior marginals of the estimated parameters obtained by fitting the Gaussian model with random slopes for each subject to the sleep study data, using both the MCMC and AMIS-INLA methods.

Figure 6.16 includes the graphical diagnostic for the parameters in the model fitted, which suggest that the estimates obtained are reliable.

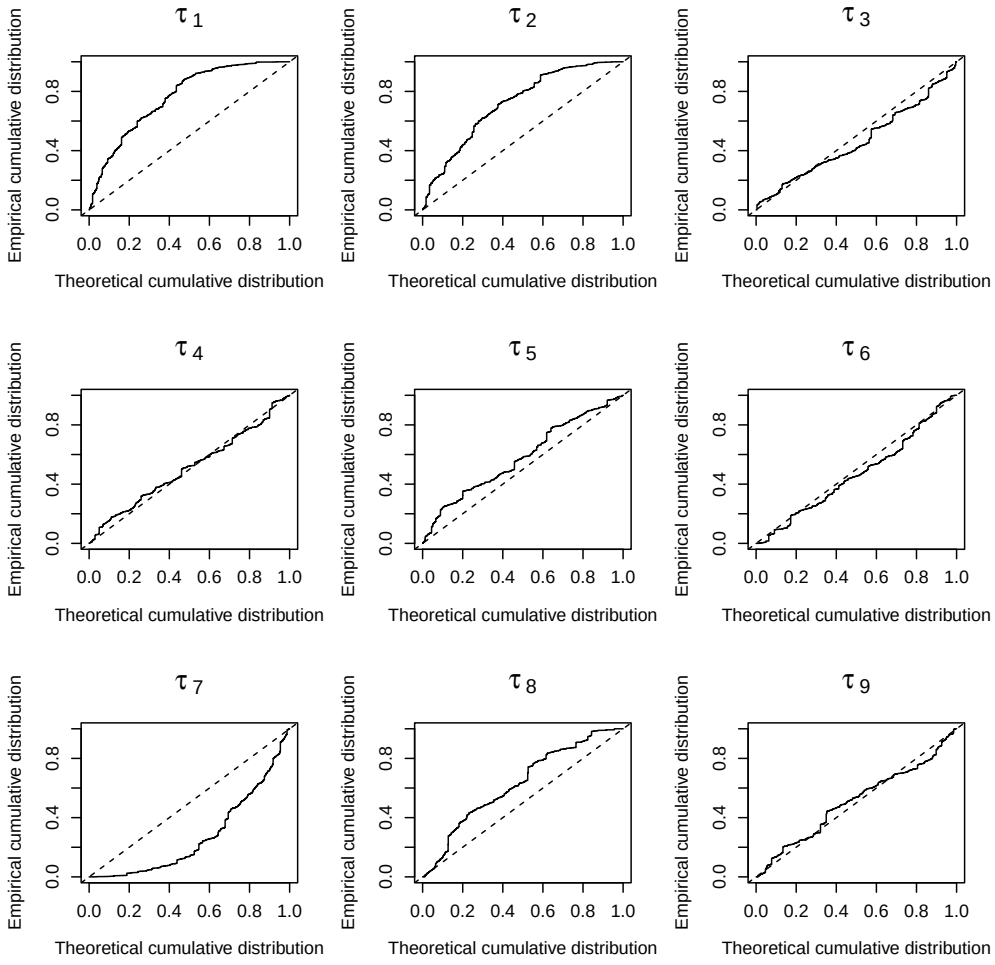


Figure 6.16: Graphical diagnostics for the model in the sleep deprivation study.

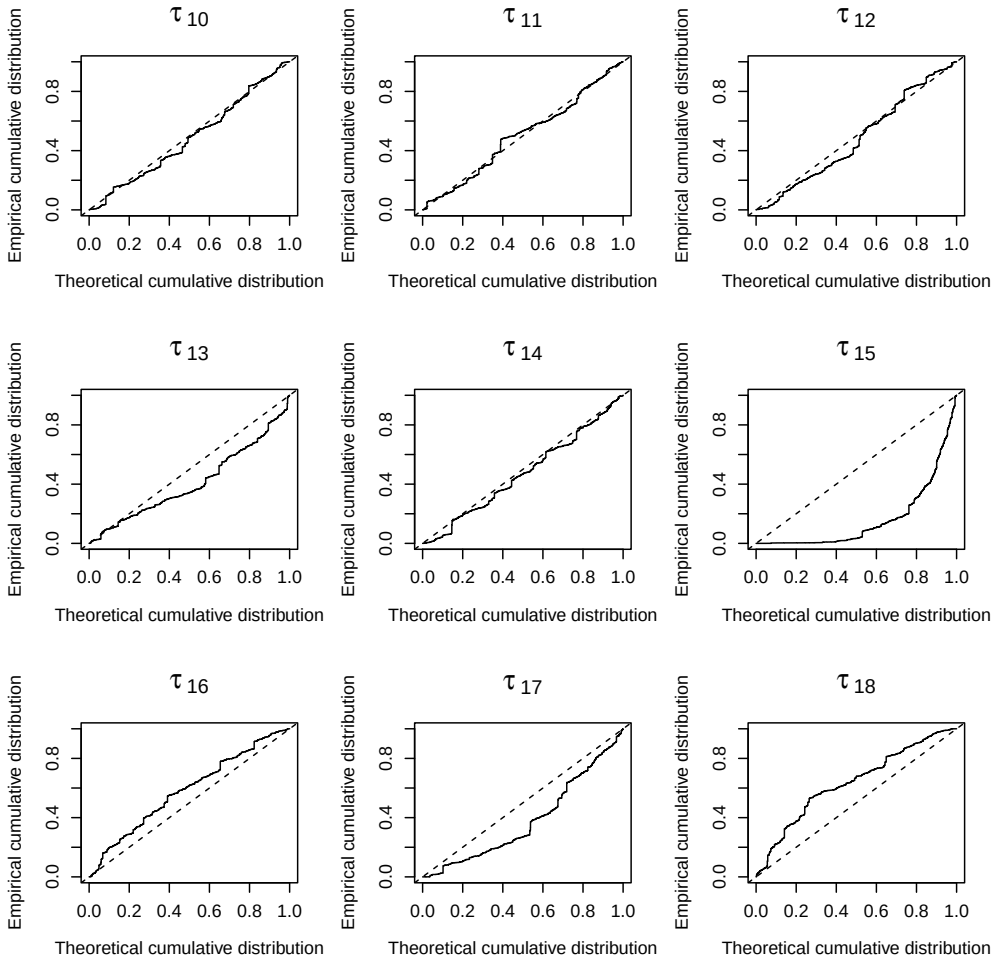


Figure 6.16: Graphical diagnostics for the model in the sleep deprivation study (Continued).

6.7 Computation times

In order to provide an overview of the computation times (in seconds) required to fit the different models discussed in the paper, Table 6.9 includes some approximate model fitting times. We would also like to mention that establishing a direct comparison between AMIS with INLA and MCMC is somehow difficult, mainly because of the different implementations used for these methods. More specifically, MCMC results have been obtained with JAGS (which is based on C++), whereas INLA with AMIS relies of calling the `inla()` function repeatedly from R, which introduces a considerable overhead.

In Table 6.9, elapsed times refer to the time required since starting the AMIS with

INLA or the MCMC sampling and until the end of the sampling process. Time per sample is the elapsed time divided by the total number of samples carried out. As stated in Section 6.5, simulations have been carried out on a Linux Ubuntu 18.2 cluster using 60 cores Intel(R) Xeon(R) CPU E5-2683 v4 @ 2.10GHz for INLA with AMIS, and using a single core for MCMC. Therefore, it is clear that INLA with MCMC times can still be reduced by increasing the number of cores used in the simulations.

Finally, we would like to mention that, although times per iterations are considerably smaller for MCMC, it is worth indicating that, while MCMC provides a sample from the joint posterior distribution of the parameters in the model, INLA with IS provides a sample from the parameters estimated with IS and the (conditional) posterior marginal distributions for the remaining parameters in the model. Hence, in our view, IS with INLA provides a more informative output per iteration than MCMC. This issue clearly illustrates the fact that time per iteration times cannot be really directly compared between the MCMC and IS with INLA approaches.

Table 6.9: Summary of the computation times (in seconds) required to fit the different models using INLA with AMIS and MCMC.

Simulation studies

	AMIS		MCMC	
Model	Total Elapsed	Time per sample	Total Elapsed	Time per sample
Poisson	1673.41	0.11	801.24	0.0073
Negative binomial	37814.47	2.52	593.14	0.0054
Gaussian (scenario 1)	3382.17	0.22	589.52	0.0054
Gaussian (scenario 2)	3303.17	0.22	589.52	0.0054
Gaussian (scenario 3)	2209.47	0.20	589.52	0.0054
Gaussian (scenario 4)	2366.28	0.22	589.52	0.0054
Gaussian (scenario 5)	4083.92	0.27	589.52	0.0054
Gaussian (scenario 6)	23854.80	0.43	589.52	0.0054

Examples

	AMIS		MCMC	
Model	Total Elapsed	Time per sample	Total Elapsed	Time per sample
Infant mortality (Poisson)	2323.63	0.15	29.84	0.0003
Infant mortality (neg. binomial)	4865.71	0.32	15.97	0.0001
Sleep study	8561.52	0.41	20.73	0.0002

6.8 Discussion

Double hierarchical models present a particular structure that models both the mean and scale parameter of different hierarchical models with likelihood within the exponential family. Hence, inference on these models can be difficult due to the different levels and effects in the model hierarchy. We have illustrated how the integrated nested Laplace approximation can be used to fit these models by using importance sampling and adaptive multiple importance sampling.

In practice, this allows INLA to integrate most of the latent effects and hyperparameters out so that a small subset of them is estimated using importance sampling. Given that IS can be easily parallelized, this provides an approach that is computationally competitive and computing times can be close to the ones provided by INLA.

We have illustrated model fitting of DHGLM by conducting three different simulation studies and the analysis of two real data sets. In all cases, conducting an adaptive multiple importance sampling provided good estimates of the model effects and hyperparameters that were similar to those obtained with Markov chain Monte Carlo methods.

For all the models presented here, we have provided graphical representations, where we illustrate how to split the DHGLM into submodels that can be fitted with INLA, so that the remainder parameters are fitted using adaptive multiple IS (AMIS). Moreover, we offer guidelines on how to choose the proposal distributions. In addition, for all cases we have also provided the effective sample sizes, computational times and graphical diagnostics to assess the convergence of the proposed method.

Among the examples that were presented here, we can find the fitting of the generalized overdispersion Poisson and negative binomial models to the infant mortality rates in Colombia, which were also illustrated in Section 2.6, Chapter 2, where the models were fitted using OpenBUGS and JAGS, based on the MCMC approach. Therefore, with the method proposed here, we now have an alternative to the MCMC algorithms that uses the INLA approach for fitting generalized overdispersion models.

Although we have discussed examples with Gaussian, Poisson and negative binomial data, the approach presented here can be applied to any of the distributions in the exponential family and, more generally, to other likelihood distributions that can be used together with the R-INLA software. Any model that can be expressed as a latent GMRF by conditioning on a (small) subset of latent effects or hyperparameters is susceptible to be fitted with IS/AMIS with INLA. In addition, we would like to mention that the proposed method has not been developed as a substitute of MCMC, but as a tool for INLA users

Finally, the R code used to develop the simulation study and the examples is available from https://github.com/becarioprecario/DHGLM-INLA. In this repository, the data for the infant mortality in Colombia in Section 6.6, and due to confidentiality constrains, have been replaced by a simulated data set.

Chapter 7

General discussion

The starting point of this thesis has been the study and application of Bayesian generalized spatial conditional overdispersion models proposed by Cepeda-Cuervo, Córdoba and Núñez-Antón (2018). These are generalized linear models which are able to account for the possibly existing spatial correlation in count data by including a spatial lag of the response variable in the regression structure for the mean. In addition, the authors assumed that the overdispersion could be partially caused by the spatial dependence. The remaining overdispersion that could still be present is captured by means of the inclusion of random effects in the model (i.e., in the spatial conditional normal Poisson and binomial models), and by assuming mixture distributions for the response (i.e., in the spatial conditional negative binomial and the beta binomial models). Furthermore, the generalized versions of these models also offer the possibility of modelling the dispersion parameter, allowing it to vary according to some covariates and/or spatial terms.

We have demonstrated their usefulness by applying them to real data examples and by comparing them with other models widely used in the literature. Therefore, we believe that they represent an excellent option for fitting spatial count data, especially for performing inference about the type and strength of the spatial dependence present in the data, when also taking into account and capturing the overdispersion.

These models are flexible enough so that they can be extended in a number of ways. In this sense, we have developed some proposals, which, in our view, are quite valuable and represent a considerable advance that widens the class of data that can be fitted with them. More concretely, we have proposed models for fitting spatio-temporal count data, semiparametric models for capturing non linear relationships and also, models that allow us to perform inference on the geometric mean of the values of the response variable in neighbouring locations.

In this thesis, we have used a Bayesian approach, due to the flexibility it offers in model fitting and the fact that it allows us to perform inference for the parameters in the models in a simple and direct manner. Here, we could highlight the advantages of the MCMC methods, where the implementation of almost any model is mostly straightforward. However, for highly parametrized models, computational times can be very large and convergence problems can arise given the potential autocorrelation that could

exist among the chains of the parameters. An alternative to MCMC could be given by INLA, which provides fast Bayesian inference, but we are bound to the models that are already specified in this package, as it is difficult to incorporate new ones. In general, all the work that has been done in this thesis was implemented in the R software, using a series of packages for the analysis and modelling of spatial and spatio-temporal data. The assessment of the fitted models has been performed by means of posterior predictive checks and comparison among them was carried out with the use of information criteria values such as the DIC and WAIC.

In Chapter 2, we focused on the spatial conditional models for Poisson and binomial count data, applying them to the infant mortality and to the mother's postnatal period screening test in Colombia data sets, respectively. We evaluated the fitted models with a number of posterior predictive checks, finding that they provided a reasonable accuracy with regard to parameter estimation. In addition, comparisons were performed with the BYM and BYM2 models, obtaining similar estimates. Nevertheless, with the spatial conditional model, we were able to obtain information about the type and strength of the spatial autocorrelation that was present in the data. Here, models were fitted in OpenBUGS and in JAGS, following the MCMC approach and in INLA, obtaining very similar estimates and revealing the much smaller computation times INLA needed for model fitting, as compared to the former approach. However, we came across the issue that the generalized models, for which regression structures are specified both for the mean and dispersion parameters, could not be implemented in INLA. Therefore, in Chapter 6 we developed a method that allowed us to accomplish this.

In Chapter 3, we proposed a semiparametric extension of the generalized spatial conditional overdispersion models in Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), which is able to account for possible non linear relationships between variables that are included in the study and the linear predictor. In particular, for the smoothing of such variables, we have specified P-splines in their mixed model representation. We present the application of these proposals to the infant mortality rates and to the mother's postnatal screening period in Colombia data sets, where we found evidence of a non linear relationship between the mortality rates and the variable representing the amount of resources provided by the government for academic achievement or education.

In Chapter 4, we proposed a model to fit spatio-temporal count data, which was a direct extension of the spatial conditional model of Cepeda-Cuervo, Córdoba and Núñez-Antón (2018), where we included the spatial lag of the response variable for each time unit in the linear predictor. Additionally, another extension of this model was also proposed, where we specified a random coefficient for the spatial term, which allowed us to investigate the temporal variation of the spatial dependence. We illustrated the models for Poisson and binomially distributed responses, being able to characterize the spatio-temporal behaviour of the respiratory hospital admissions in Glasgow and the low birth weight in Georgia. Here, it could be useful to mention that these proposals could also be specified for any other count data distribution.

In Chapter 5, we have seen how these models can be extremely useful within an epidemiological context, when we modelled the COVID-19 incidence data in the munici-

palities of Flanders, and were able to capture the spatial dependence among the regions. Here, these models allowed us to easily investigate different weights matrices that could be explaining the spatial underlying process, revealing the importance of conducting tests to further explore this issue. Furthermore, we proposed an extension where we included the logarithm of the incidence rates instead of the rates, which allowed us to perform inference based on the similarity of the incidence rate in one region with the geometric mean of these rates in the neighbouring regions.

As mentioned before, an alternative method to MCMC for fitting generalized spatial conditional overdispersion models in INLA was proposed in Chapter 6. This was achieved by combining the adaptive multiple importance sampling (AMIS) algorithm and INLA and it can be used to fit any model that belongs to the class of double hierarchical generalized linear models (DHGLM). We illustrated the proposed method with simulation studies and applications to real data examples, and compare the results obtained with those from the MCMC approach, finding that our proposal is able to provide good estimates.

The modelling of biostatistical data can offer important insights which may lead to a better understanding of the different data being analysed. In order to be able to identify the socio-economic or demographic characteristics of the population that may have a higher impact on the phenomenon under study, the information obtained from such studies can be extremely useful to governments or responsible authorities. In this way, authorities would be able to optimize the investment of resources and dictate more effective policies. In addition, modelling of infectious diseases allows us to characterize and forecast their spread and to identify the risk factors of the population, so that better decisions based on reliable statistical evidence can be made by the authorities. This fact has become especially evident after the recent COVID-19 outbreak, showing us the importance of disease mapping and epidemiology in the context of public health policy making (see Chapter 5).

For example, in the case of the analysis of the infant mortality rates in Colombia, performed in Chapter 2, the results obtained suggested that, if the index of unsatisfied basic needs is reduced in some departments, their infant mortality rates may also be reduced. In addition, if the government increases the amount of the resources provided for academic achievement and education per household for some departments, it may result in lower mortality rates for these regions. In the same way, the results obtained for the study of mother's postnatal period screening test in Colombia, suggest that reducing the index of unsatisfied basic needs perhaps may increase the probability of a mother going through such a medical test. Furthermore, as it was shown in Chapter 3, these relationships may not necessarily have a linear pattern, but it may be given by smooth functions or curves, so it would be advisable to check for this possibility in these types of studies.

Throughout this thesis, we have seen how spatial correlation is such an important issue in epidemiologic and public health data. One cannot neglect the major impact that the geographical component might have on the analysis being performed. In this sense, it is important to take into account the information that regions which are close together

in space can provide. Moreover, estimates for small areas, with lower population can be smoothed by considering the neighbours information. In general, a regression model which does not take into account this information may not produce reliable estimations. In this sense, it is strongly recommended to perform tests in order to find the spatial structure which best accommodates the spatial underlying process of the specific area under study (see Chapter 5). Additionally, being able to study the behaviour in time of the spatial dependence is also of interest in many applications (see Chapter 4).

In today's world, the analysis and modelling of such spatial and spatio-temporal data has become a challenging task requiring faster and more efficient methods. In our modest view, it is very important to keep methods as simple as possible and, in addition, to have models at our disposal that are flexible enough so as to easily adapt them to the specific needs of the application under consideration. For these and all the reasons that have been presented in this thesis, we would certainly advice the use of our proposed spatial conditional models, over the BYM or BYM2 models. In addition, one of the problems researchers most commonly come across when fitting regression models for these types of data are computational challenges (see Chapter 6). In order to overcome these issues, and although MCMC-based software packages offer high flexibility, we would recommend the use of INLA, which provides faster computation times. Even though there are models that are not currently implemented in this software package, there are alternatives for fitting these models in INLA, such as the ones we proposed in Chapter 6.

For the models we have presented here further research and extensions are also possible in a number of directions. For example, in the case of the spatio-temporal models presented in Chapter 4, we could assume other alternative specifications for the random effects and for the varying coefficient of the spatial lag, such as autoregressive processes of higher order. For these random effects, we could specify a regression structure for their variance parameters, which would be extensions of the generalized overdispersion models. In addition, multivariate extensions for the spatial conditional overdispersion models that allowed to model more than one dependent variable could be also proposed. These questions are indeed expected to be the focus of our future research.

References

Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). John Wiley & Sons, Inc.: Hoboken, USA.

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, **27**(**3**), 247–267.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D. and Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *WIREs Computational Statistics*, **10**(**6**), e1443.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(**1**), 1–48.

Belenky, G., Wesensten, N.J., Thorne, D.R., Thomas, M.L., Sing, H.C., Redmond, D.P., Russo, M.B. and Balkin, T.J. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. *Journal of Sleep Research*, **12**(**1**), 1–12.

Berild, M.O., Martino, S., Gómez-Rubio, V. and Rue, H. (2022). Importance sampling with the integrated nested Laplace approximation. *Journal of Computational and Graphical Statistics*, 1–13. In press.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, **14**(**21-22**), 2433–2443.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society - Series B*, **36**(**2**), 192–236.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**(**1**), 1–20.

Best, N., Richardson, S. and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, **14**(**1**), 35–59.

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

Blangiardo, M., Cameletti, M., Baio, G. and Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, **4**, 33–49.

Bonner, S., Kim, H.-N., Westneat, D., Mutzel, A., Wright, J. and Schofield, M. (2021). Dalmatian: A package for fitting double hierarchical linear models in R via JAGS and nimble. *Journal of Statistical Software*, **100**(**10**), 1–25.

Breslow, N.E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*, **33**(**1**), 38–44.

Brooks, S., Gelman, A., Jones, G.L. and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Press: Boca Raton, USA.

Burnham, K.P. and Anderson, D.R. (2002). *Model Selection and Multimodel Inference* (2nd ed.). Springer: New York, USA.

Carroll, R., Lawson, A.B., Faes, C., Kirby, R.S., Aregay, M. and Watjou, K. (2015). Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spatial and Spatio-temporal Epidemiology*, **14-15**, 45–54.

Carroll, R., Lawson, A.B., Faes, C., Kirby, R.S., Aregay, M. and Watjou, K. (2016). Spatio-temporal Bayesian model selection for disease mapping. *Environmetrics*, **27**(**8**), 466–478.

Case, A., Hines, J.R.J. and Rosen, H.S. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the States. *Journal of Public Economics*, **52**(**3**), 285–307.

Cepeda-Cuervo, E., Córdoba, M. and Núñez-Antón, V. (2018). Conditional overdispersed models: Application to count area data. *Statistical Methods in Medical Research*, **27**(**10**), 2964–2988.

Congdon, P. and Southall, H. (2005). Trends in inequality in infant mortality in the north of England, 1921–1973, and their association with urban and social structure. *Journal of the Royal Statistical Society - Series A*, **168**(**4**), 679–700.

Corneut, J.-M., Marin, J.-M., Mira, A. and Robert, C.P. (2012). Adaptive Multiple Importance Sampling. *Scandinavian Journal of Statistics*, **39**(**4**), 798–812.

Cowell, R.G., Dawid, A.P., Lauritzen, S.L. and Spiegelhalter, D.J. (1999). *Probabilistic Networks and Expert Systems*. Springer: New York, USA.

Crainiceanu, C.M., Ruppert, D. and Wand, M.P. (2005). Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software*, **14**(**14**), 1–24.

Cressie, N.A.C. and Wikle, C.K. (2011). *Statistics for Spatio-Temporal Data*. Wiley: Hoboken, USA.

Currie, I.D. and Durbán, M. (2002). Flexible smoothing with P-splines: A unified approach. *Statistical Modelling*, **2**(**4**), 333–349.

Cutland, C.L., Lackritz, E.M., Mallett-Moore, T., Bardají, A., Chandrasekaran, R., Lahariya, C., Nisar, M.I., Tapia, M.D., Pathirana, J., Kochhar, S. and Muñoz, F.M. (2017). Low birth weight: Case definition & guidelines for data collection, analysis, and presentation of maternal immunization safety data. *Vaccine*, **35**(**48, Part A**), 6492–6500.

D'Angelo, N., Abbruzzo, A. and Adelfio, G. (2021). Spatio-temporal spread pattern of COVID-19 in Italy. *Mathematics (Special issue on Statistical Models in the Era of Big Data)*, **9**(**19**), 2454.

De Boor, C. (1978). *A Practical Guide to Splines*. Springer: New York, USA.

Dean, C.B. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association*, **87**(**418**), 451–457.

Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R. and Beard, J. (2007). Evaluating the effect of neighbourhood weight matrices on smoothing properties of conditional autoregressive (CAR) models. *International Journal of Health Geographics*, **6**(**54**).

Eberly, L.E. and Carlin, B.P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine*, **19**(**17-18**), 2279–2294.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, **81**(**395**), 709–721.

Eilers, P.H. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**(**2**), 89–121.

Eilers, P.H. and Marx, B.D. (2021). *Practical Smoothing: The Joys of P-splines*. Cambridge University Press: Cambridge, UK.

Eilers, P.H., Marx, B.D. and Durbán, M. (2015). Twenty years of P-splines. *SORT*, **39**(**2**), 149–186.

Ejigu, B.A. and Wencheko, E. (2020). Introducing covariate dependent weighting matrices in fitting autoregressive models and measuring spatio-environmental autocorrelation. *Spatial Statistics*, **38**, 100454.

Ensoy-Musoro, C., Hens, N., Molenberghs, G. and Faes, C. (2022). *Spatio-temporal model to investigate the effect of mobility using mobile network data*. Preprint submitted to Spatial Statistics.

Franco-Villoria, M., Ventrucci, M. and Rue, H. (2019). A unified view on Bayesian varying coefficient models. *Electronic Journal of Statistics*, **13**(**2**), 5334–5359.

Gamerman, D. and Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2nd ed.). Chapman & Hall/CRC Press: Boca Raton, USA.

Geary, R.C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, **5**(**3**), 115–146.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**(**3**), 515–534.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**(**4**), 457–472.

Gelman, A., Hwang, J. and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, **24**(**6**), 997–1016.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D., Vehtari, A. and Rubin, D.B. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman and Hall/CRC: Boca Raton, USA.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Getis, A. (2008). A history of the concept of spatial autocorrelation: A geographer's perspective. *Geographical Analysis*, **40**(**3**), 297–309.

Gómez-Rubio, V. (2020). *Bayesian Inference with INLA*. Chapman and Hall/CRC Press: Boca Raton, USA.

Gómez-Rubio, V. and Rue, H. (2018). Markov chain Monte Carlo with the integrated nested Laplace approximation. *Statistics and Computing*, **28**(**5**), 1033–1051.

Gómez-Rubio, V., Palmí-Perales, F., López-Abente, G., Ramis-Prieto, R. and Fernández-Navarro, P. (2019). Bayesian joint spatio-temporal analysis of multiple diseases. *SORT*, **43**(**1**), 51–74.

Green, P.J. and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman & Hall: London, UK.

Griffiths, D.A. (1973). Maximum likelihood estimation for the Beta-Binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics*, **29**(**4**), 144–148.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society - Series B*, **55**(**4**), 757–779.

Held, L., Höhle, M. and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, **5**(**3**), 187–199.

Hilbe, J.M. (2011). *Negative Binomial Regression* (2nd ed.). Cambridge University Press: New York, USA.

Hinde, J. (1982). Compound Poisson regression models. In *GLIM 82: Proceedings of the International Conference on Generalised Linear Models* (R. Gilchrist, ed.). Springer: New York, USA, 109-121.

Hinde, J. and Demétrio, C.G.B. (1998). Overdispersion: Models and estimation. *Computational Statistics & Data Analysis*, **27**(**2**), 151–170.

Jackson, M.C. and Sellers, K.F. (2008). Simulating discrete spatially correlated Poisson data on a lattice. *International Journal of Pure and Applied Mathematics*, **46**(**1**), 137–154.

Johnson, D.P., Ravi, N. and Braneon, C.V. (2021). Spatiotemporal associations between social vulnerability, environmental measurements, and COVID-19 in the conterminous United States. *GeoHealth*, **5**(**8**), e2021GH000423.

Kaiser, M. and Cressie, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statistics & Probability Letters*, **35**(**4**), 423–432.

Kazembe, L. (2009). Modelling individual fertility levels in malawian women: A spatial semiparametric regression model. *Statistical Methods and Applications*, **18**, 237–255.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, **19**(**17-18**), 2555–2567.

Knorr-Held, L. and Richardson, S. (2003). A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Applied Statistics*, **52**(**2**), 169–183.

Konstantinoudis, G., Cameletti, M., Gómez-Rubio, V., Gómez, I.L., Baio, M.P.G., Larrauri, A., Riou, J., Egger, M., Vineis, P. and Blangiardo, M. (2022). Regional excess mortality during the 2020 COVID-19 pandemic in five European countries. *Nature Communications*, **13**(**482**).

Lambert, D., Brown, J. and Florax, R. (2010). A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application. *Regional Science and Urban Economics*, **40**(**4**), 241–252.

Lawson, A.B. (2008). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology* (3rd ed.). Chapman & Hall/CRC Press: Boca Raton, USA.

Lázaro, E., Armero, C. and Gómez-Rubio, V. (2020). Approximate Bayesian inference for mixture cure models. *TEST*, **29**, 750–767.

Lee, D. (2013). CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software*, **55**(**13**), 1–24.

Lee, D., Rushworth, A. and Napier, G. (2018). Spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST package. *Journal of Statistical Software*, **84**(**9**), 1–39.

Lee, Y. and Nelder, J.A. (2006). Double hierarchical generalized linear models (with discussion). *Applied Statistics*, **55**(**2**), 139–185.

Lee, Y. and Noh, M. (2012). Modelling random effect variance with double hierarchical generalized linear models. *Statistical Modelling*, **12**(**6**), 487–502.

Leroux, B.G., Lei, X. and Breslow, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (M.E. Halloran and D. Berry, eds.). Springer: New York, USA, 179-191.

Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, **63**(**19**), 1–25.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, **28**(**25**), 3049–3067.

Margolin, B.H., Kaplan, N. and Zeiger, E. (1981). Statistical analysis of the Ames Salmonella/microsome test. *Proceedings of the National Academy of Sciences of the United States of America*, **78**(**6**), 3779–3783.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (2nd ed.). Chapman & Hall/CRC: London, UK.

Moraga, P. (2018). Small area disease risk estimation and visualization using R. *The R Journal*, **10**(**1**), 495–506.

Morales-Otero, M. and Núñez-Antón, V. (2021). Comparing Bayesian spatial conditional overdispersion and the Besag-York-Mollié models: Application to infant mortality rates. *Mathematics (Special issue on Spatial Statistics with its Applications)*, **9**(**3**), 282.

Morales-Otero, M., Gómez-Rubio, V. and Núñez-Antón, V. (2022). Fitting double hierarchical models with the integrated nested Laplace approximation. *Statistics and Computing*, **32**(**4**), 62.

Moran, P.A.P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society - Series B*, **10**(**2**), 243–251.

Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S.J., Gelman, A. and DiMaggio, C. (2019). Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in Stan. *Spatial and Spatio-temporal Epidemiology*, **31**, 100301.

Natalia, Y.A., Faes, C., Neyens, T. and Molenberghs, G. (2022). The COVID-19 wave in Belgium during the Fall of 2020 and its association with higher education. *PLOS ONE*, **17**(**2**), e0264516.

Onkelinx, T. (2019). The inlatools package. Retrieved August 9, 2022, from https://inlatools.netlify.app/

Plummer, M. (2021). *Rjags: Bayesian graphical models using mcmc*. R package version 4-12.

Quintero-Sarmiento, A., Cepeda-Cuervo, E. and Núñez-Antón, V. (2012). Estimating infant mortality in Colombia: Some overdispersion modelling approaches. *Journal of Applied Statistics*, **39**(**5**), 1011–1036.

Riebler, A., Sørbye, S.H., Simpson, D.P. and Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, **25**, 1145–1165.

Rönnegård, L., Felleki, M., Fikse, F., Mulder, H.A. and Strandberg, E. (2010). Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *Genetics Selection Evolution*, **42**(**1**), 8.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society - Series B*, **71**(**2**), 319–392.

Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P. and Lindgren, F.K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, **4**(**1**), 395–421.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, **11**(**4**), 735–757.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression*. Cambridge University Press: Cambridge, UK.

Simpson, D., Rue, H., Riebler, A., Martins, T.G. and Sørbye, S.H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, **32**(**1**), 1–28.

Smith, P.L. (1979). Splines as a useful and convenient statistical tool. *The American Statistician*, **33**(**2**), 57–62.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society - Series B*, **64**(**4**), 583–639.

Ugarte, M.D., Goicoa, T., Etxeberria, J., Militino, A.F. and Pollán, M. (2010). Age-specific spatio-temporal patterns of female breast cancer mortality in Spain (1975-2005). *Annals of Epidemiology*, **20**(**12**), 906–916.

Vranckx, M., Neyens, T. and Faes, C. (2019). Comparison of different software implementations for spatial disease mapping. *Spatial and Spatio-temporal Epidemiology*, **31**, 100302.

Wall, M.M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, **121**(**2**), 311–324.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, **11**(**116**), 3571–3594.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, **41**(**3-4**), 434–449.

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics*, **31**(**4**), 949–952.

Williams, D.A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics*, **31**(**2**), 637–648.

Zeger, S. and Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, **44**(**4**), 1019–1031.