PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA

BAYESIAN ROBUST MODELS WITH APPLICATIONS TO SMALL AREA ESTIMATION

by Francisco J. Torres Avilés

Submitted in partial fulfillment of the requirements for the degree of Doctor in Statistics at Facultad de Matemáticas Pontificia Universidad Católica de Chile.

Directed by Ph.D. Gloria Icaza Noguera
Co-directed by Ph.D. Reinaldo Arellano-Valle

July,2008
Santiago, Chile

BAYESIAN ROBUST MODELS WITH APPLICATIONS TO SMALL AREA ESTIMATION

By

Francisco J. Torres Avilés

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR IN STATISTICS
AT
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SANTIAGO - CHILE
JULY 2008

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTMENT OF STATISTICS

The undersigned hereby certify that they have read and recommend to the Faculty of Mathematics for acceptance a thesis entitled "**Bayesian Robust Models with Applications to Small Area Estimation**" by **Francisco J. Torres Avilés** in partial fulfillment of the requirements for the degree of **Doctor in Statistics**.

Dated: July 2008

Research Supervisors: ______________________
Gloria Icaza Noguera

Supervisor ______________________
Reinaldo Arellano Valle

Examining Committee: ______________________
Alicia L. Carriquiry

____
Wilfredo Palma Manríquez

____
Ignacio Vidal García

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Date: **July 2008**

Author: **Francisco J. Torres Avilés**

Title: **Bayesian Robust Models with Applications to Small Area Estimation**

Department:

Degree: **D.S.** Convocation: **July** Year: **2008**

Signature of Author

Table of Contents

List of Tables

List of Figures

Abstract

This work takes up methods for Bayesian inference in generalized linear mixed models with applications to small-area estimation. A previous work (Datta and Lahiri, 1995) focused on Bayesian estimation with a prior scale mixture distribution for the error component in a normal linear model, to smooth small area means when one or more outliers are present in the data. Following this idea, an appropriate scale mixture of normals (Andrews and Mallows, 1974, Fernández and Steel, 2000) for the spatial random effects distribution is proposed in the context of the Markov random field theory, which is applied to the usual spatial intrinsically autoregressive random effect. Conditions are stablished in order to guarantee the posterior distribution existence when the random field is observed directly. Given a joint observed random field, a simulation study is performed to illustrate the use of hierarchical algorithms. Inference over the variability parameter is obtained, showing that the best estimators are related to a particular scale mixture of normal random field.

Based on the work of Ghosh et al. (1998), theoretical conditions are presented to guarantee the posterior distribution propriety, when a generalized linear mixed model with a spatial component is assumed. Due to the equivalence between the normal and the scale mixture of normal models, specifically with Student-t and Slash distributions, it is possible to obtain hierarchical representations, therefore, Markov Chain Monte Carlo sampler methods are used to perform the computations.

Lung, trachea and bronchi cancer relative risk and childhood diabetes incidence in Chilean communes are estimated to illustrate the proposed methods. Inference over unknown parameters are discussed. Results are presented using appropriate thematic maps.

As part of the work in progress, theoretical aspects to measure Bayesian learning are explored, taking into account that in the spatial hierarchical model considered in this work, only the sum of two sets of random effects are identified by the data. Specific expressions for the L_1 distance were obtained. Other considerations are discussed as part of the future work, considering extensions to be developed in different directions.

Acknowledgements

I wish to thank to God who gave me the energy to finish this stage. This work would not be possible without the initial support and great encouragement of my friend Pilar Iglesias, under whose supervision I choose this topic and began to develop this thesis. Gloria Icaza, my advisor in the final stages of the work, who have assisted me in numerous ways, which includes several discussions related to epidemiology aspects of the study and her critical point of view respect to the written work. I also appreciate the helpful comments and critics that Reinaldo Arellano Valle made to my work when it was necessary.

Thanks to the examining committee for their valuable contributions and considerations made to the submitted work. Finally, I also want to thank to the Chilean National Commission of Science and Technology (CONICYT) and its financial support through the doctoral fellowship, which has supported me during the last years of research, and the award of one travel grant, which allow me to present advances in the 9th EBEB in Maresias, Sao Paulo, Brazil.

Chapter 1

Introduction

1.1 Spatial epidemiology issues

Spatial epidemiology concerns to the analysis of the spatial and spatio-temporal distribution of a specific disease. This discipline has become an area of research with high development in the last years. The continuous computational development and technical advances in statistical methodologies and geographic information system have helped to obtain satisfactory results to solve problems in different research areas. Different formats of epidemiological data naturally give rise to different statistical methods. In general, the spatial analysis data can be classified as follows:

- Geostatistics or point referenced data, when the main goal is prediction of different georeferenced points or extrapolation of areas, since available samples of measurements are from a spatially continuous phenomenon of interest.
- Point pattern analysis data, when identification of clusters in space is the focus.
- Areal or lattice data, when the focus is to explain the geographical variation of an event (economical or epidemiological), in administrative separated small areas.

This work presents a methodological review and extensions of usual areal data models with applications in epidemiology.

Large scale disease mapping in spatial epidemiology, shows disease spatial variation with primarily descriptive purposes [88]. From a public health perspective, it is an important tool because it allows the implementation of policies in areas where high risks are detected.

There are well known problems with mapping raw and standardized rates for rare diseases and/or small areas, since sampling variability tends to dominate the subsequent maps [16]. This implies that the statistical analysis aims to provide a map free of distortion, such that, precise estimates would be obtained for each small area.

The problem is treated with discrete data that include the number of individuals y_i, defined under two types of design, the number of cases of a disease that are present in a particular population at a given time (prevalence), or the number of newly diagnosed cases during a specific time period (incidence). In epidemiology, the application of methods to adjust raw rates, called "methods of standardization", are necessary. These methods aim to provide comparable rates between areas, when different sex and age population structures are present. These standardized rates measure the risk of having a disease within each area. There are two methods of rate standardization, direct and indirect. Indirect standardization allows the specific estimation of rates through what in literature is called Standardized Morbidity/Mortality Ratio (SMR) [19]. In either of these two methods, every observed y_i is an aggregation within each i area, that is, the number of cases of interest are considered for the analysis, which means that the individual variation of cases will be lost.

Successful studies in Europe ([7], [17], [72], [83], among others) with applications to cancer and diabetes data, was a motivation to apply these statistical methods for borrowing strength in small area disease, which could reduce the variability not explained by the model, incorporating a probabilistic structure that represents the relationship between the areas of study.

This thesis work was motivated by real data, in the sense that results obtained from usual convolution models [68] were too smooth in order to be representative for the true risks. The latter was the reason to fit spatial robust models, which assume heavier tailed distributions in the spatial random effects. Applications are oriented to obtain incidence rates for insulin dependent diabetes mellitus (IDDM) in Chilean Metropolitan Region and relative risks of female trachea, bronchi and lung cancer mortality in the northern area of the country.

1.2 Background

Some useful definitions about the modeling stage will be exposed in this section. A brief introduction and some general aspects will be discussed in the following order: general treatment of the generalized linear mixed models, definition of gaussian Markov random fields and the usual construction of spatial Poisson models.

1.2.1 The generalized linear mixed model

McCullagh and Nelder [66] define a wide class of regression models, when the assumption of normality in the response variable is not appropriate. They define the class of Generalized Linear Models (GLM) as follows: consider m independent random variables $y_1, \ldots, y_m$ following a distribution that belong to the exponential family, i.e., with the joint distribution given by

$$f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi}) = \prod_{i=1}^{m} \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i)) + \rho(\phi_i; y_i)\}, \tag{1.2.1}$$

where $\mathbf{y} = (y_1, \ldots, y_m)'$,, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$ is a vector of unknown canonical parameters, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)'$ is a vector of known scale parameters and ρ is a known function that does not depend on the unknown parameters.

The main specification in a GLM is the relation between a linear predictor and an strictly increasing function h, called link function. In a GLM, linear relationship present the following form,

$$h(\theta_i) = \mathbf{x}_i'\boldsymbol{\beta},$$

where $\mathbf{x}_i$ is a $p \times 1$ vector of covariates, β is a vector of unknown structural parameters associated to each component of $\mathbf{x}_i$. Generally, β is considered as a vector of fixed effects. Allowing for one or more random effects, say ϵ_i, GLM are extended to what is called Generalized Linear Mixed Model (GLMM). If an additional parameter is considered into the link function, a GLMM will consider the following additive structure,

$$h(\theta_i) = \mathbf{x}_i'\beta + \epsilon_i, \quad i = 1, \ldots, m,$$

where ϵ_i is called random effect. Pioneer GLMM literature is referred to works developed by Breslow and Clayton [18] and Clayton [27].

1.2.2 Gaussian Markov random fields

Let $\pi(\mathbf{u})$ be a probability distribution associated to a region R. Denote by u_i a random variable measured on area $i \in R$ and, $\mathbf{u}_{-i}$ a subset which contains random values on all areas other than area i. Consider the conditional distribution $\pi(u_i|\mathbf{u}_{-i})$ associated to area i given observed values $\mathbf{u}_{-i}$. Viewed through its conditional distribution at each area, $\pi(\mathbf{u})$ is termed Markov random field (MRF) [10]. Brook's lemma ([6],[9]) is applied to construct the joint distribution of the random field. A brief discussion will be exposed in section two, in the context of the proposed models.

Pioneer applications in the context of this class of processes were related to image reconstruction ([10], [15]) and agricultural field experiments ([9], [12]).

Nowadays, Gaussian MRF's are the most commonly models used for disease mapping [76]. If the observations of any two areas are connected by a common boundary, and they are assumed conditionally independent normally distributed on all other observations, then the joint distribution $\pi(\mathbf{u})$ is called Gaussian MRF (GMRF) [74]. A formal way to define a GMRF is,

$$\pi(\mathbf{u}) \propto \exp\left\{-\frac{1}{2\sigma^2}\mathbf{u}'D_w\mathbf{u}\right\}, \tag{1.2.2}$$

where $\mathbf{u} \in \mathbb{R}^m$ with m denoting the number of areas at the region R, and D_w is a $m \times m$ symmetric matrix. The entries $w_{ij} = (D_w)_{ij}$ represent the spatial connection between areal units i and j. D_w is usually referred to as proximity or adjacency matrix. In this work, D_w will be concerned with areas whose conditional distributions depend only on the values of areas in the immediate neighborhood of the area i.

Definition (1.2.2) can be rewritten as

$$\pi(\mathbf{u}) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i\neq j} w_{ij}(\mathbf{u}_i - \mathbf{u}_j)^2\right\}, \tag{1.2.3}$$

which is generally referred to as pairwise difference condition. This specification is often called as intrinsically autoregressive (IAR) model [76]. For further details of MRF, see for example, Besag ([9], [10]), Besag et al. [15] and the books of Banerjee et al. [6] and Cressie [30]. In this work, the usual GMRF model (1.2.2) will be considered instead of the conditional IAR specification (1.2.3).

1.2.3 Spatial Poisson regression models

The Poisson model has been extensively studied since 1837 when Dennis Poisson introduced a probability rule based on the incidence of death by coz of mules to soldiers in the French army, and this discrete distribution has been used in various contexts. Griffith and Haining [48] summarizes applications of the Poisson model in the context of geographical analysis. They addressed this distribution from an applied point of view, starting from its genesis with a historical summary to its current use in geographic patterns.

Consider the number of individuals who suffer a non contagious disease in a certain area of interest, follows a Binomial distribution with parameters n_i and p_i, where n_i is the size of the population at risk and p_i the probability of having the disease. In epidemiology, there are rare diseases within a large population at risk. This argument leads to assume that the number of cases can be modelled as a Poisson distribution with mean parameter $n_i p_i$, when the population at risk is large ($n_i \to \infty$) and the probability that represents the occurrence of the disease tends to zero ($p_i \to 0$).

Overdispersion problems ($Var(y) > E(y)$) under the assumption that the number of cases (y) follow a Poisson distribution, have led to consider a model with random effects using a regression model with the following specification,

$$y_i \sim Poisson(n_i p_i)$$
$$log(p_i) = \mathbf{x}_i'\beta + \epsilon_i,$$

where ϵ_i's are the random effects. Bayesian formulation can be fitted using the developed GLMM theory [27]. Common linear extensions in disease mapping have been developed, allowing for the sum of two independent components for each area, that is, if $\epsilon_i = v_i + u_{i,}$, $i = 1, \ldots, m$, the link function can be expressed as,

$$h(p_i) = \mathbf{x}_i'\beta + v_i + u_i, \tag{1.2.4}$$

where $\mathbf{v} = (v_1, \ldots, v_m)'$ are iid normally distributed random effects, $\mathbf{u} = (u_1, \ldots, u_m)'$ are spatially structured random effects and h represent the link function, which is represented by the usual natural log for this particular case. From a Bayesian perspective, it is common to represent the spatial effect by (1.2.2), which is frequently used as a prior distribution.

The structure in 1.2.4 let to capture variability which is not explained neither by the model nor by the unstructured random effects $\mathbf{v}$. Taking this idea into account, Clayton and Kaldor [28], Besag et al. ([13], [14], [15]), Mollié [68], Wakefield and Elliot [88], Pascutto et al. [72], among others, proposed methodologies based on empirical and fully Bayesian theory to solve the rates estimation problem considering different spatial structures. Best et al. [17] summarizes advances in the area, including parametric and semi parametric models.

Empirical and fully Bayesian methods have worked fairly well ([28], [65], [68]). One difference between them is that when a Bayesian hierarchical model is considered, uncertainty can be represented through a prior information, which allows to make inference (e.g. credibility intervals) over all the unknown parameters, including those that measures variability. Following this argument, this work will primarily focus in applying Bayesian techniques to analyze data from the epidemiological area.

1.3 Model assessment

Three common ways to measure model assessment are taken into account. The first two are oriented to penalize the observed deviance. Deviance information criterion (DIC) [84] and Bayesian (or Schwarz) information criterion (BIC) [82] will be used. These methodologies make an adjustment to the model log likelihood or deviance, taking into account the number of parameters in the model, analogue to Akaike's information criterion ([1], [2]).

In practice, deviance is defined as $D(\boldsymbol{\theta}) = -2\log(f(\mathbf{y}|\boldsymbol{\theta}))$, where $f(\mathbf{y}|\boldsymbol{\theta})$ is the probability density function of $\mathbf{y}$, the observed random vector, given an unknown parameter vector $\boldsymbol{\theta}$. In Bayesian context, samples from MCMC chains are available, so it is possible to estimate the deviance associated to the model through

$$\bar{D}(\boldsymbol{\theta}) = -2\frac{1}{M}\sum_{l=1}^{M}\log(f(\mathbf{y}|\boldsymbol{\theta}_l)),$$

where M is the MCMC sample size. The effective number of parameters is still in discussion, the approximation proposed by Spiegelhalter et al. [84] will be used through the difference,

$$p_e = \bar{D}(\boldsymbol{\theta}) - D(\bar{\boldsymbol{\theta}}).$$

where $\bar{\boldsymbol{\theta}}$ is the Bayesian point estimation of $\boldsymbol{\theta}$. So, the penalized measures for model assessment, DIC and a modified BIC [29] are defined as,

$$DIC = D(\bar{\boldsymbol{\theta}}) + 2p_e$$

and

$$BIC = D(\bar{\boldsymbol{\theta}}) + p_e\log(m).$$

Small values for both measures allow to choose the best model fit.

Following the work developed by Laud and Ibrahim [61], and generalized by Gelfand and Ghosh [40], a third model choice criterion is applied. This criterion is based on a predictive check of the model and measures the discrepancy between the observed data and predicted observations, taking into account quadratic loss measures. Following this idea, the model with the best adjustment is that one with the smallest C^2, defined as

$$C^2 = \sum_{i=1}^{m} \sigma^2_{i_{pred}} + \mathrm{E}((y_{i_{obs}} - y_{i_{pred}})^2 | y_{i_{obs}}, M_0),$$

where $\sigma^2_{i_{pred}} = Var(y_{i_{pred}} | y_{i_{obs}}, M_0)$, $y_{i_{obs}}$ is the i-th random variable and $y_{i_{pred}}$ is the i-th future replication under the proposed model M_0.

1.4 Applications and exploratory analysis

Chile is located at the southern part of South America. Until 2006, the administrative division was given by 13 regions, which were subdivided into 342 districts, including the islands. Each district is called commune, and is the smallest geographical scale for which data are available.

Using standard modeling techniques, such as the convolution model studied by Molliè ([68], [67]), it was possible to obtain results related to stomach cancer and trachea, bronchi and lung cancer mortality in our country, which were published in Icaza et al. [51]. Spatio-temporal analysis for IDDM is presented in Torres et al. [86], including a complementary disease clustering analysis for pattern data [35]. Explorative analysis is developed in the next subsections.

1.4.1 Insulin dependent diabetes mellitus incidence rates

This application focuses on studying the spatial distribution of IDDM, in the Metropolitan Region of Santiago de Chile. This region is divided into 52 communes, where 34 are highly urbanized (Gran Santiago) and 18 are considered rural areas.

IDDM data correspond to the registered cases aged up to 15 years old, during the 2000-2006 period, in 52 communes of the region.

According to 2002 census, approximately 29% of the total population living in the region are under 15 years old; this population is considered as disease susceptible population or at

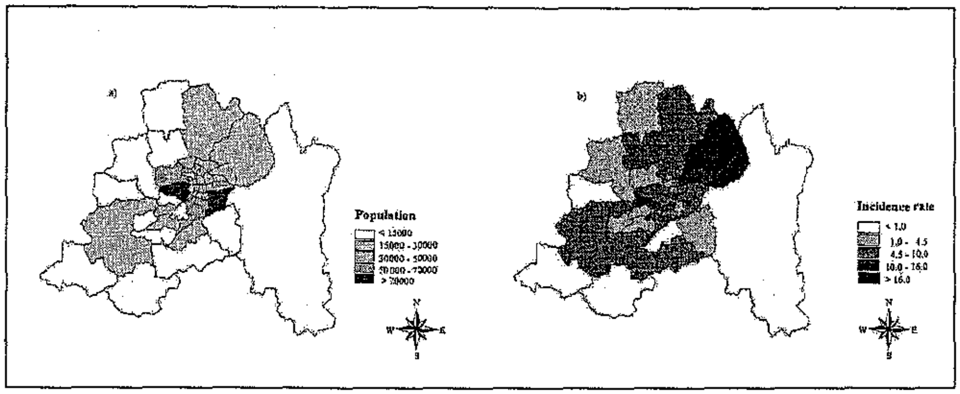

Figure 1.1: (a) Population at risk distribution; (b) Raw IDDM incidence rates

25% lower (Rural communes)			25% higher (Urban communes)		
Comuna	Cases	Incidence Rate	Comuna	Cases	Incidence Rate
San José de Maipo	0	0	El Bosque	18	5.675
María Pinto	0	0	Santiago	19	8.678
San Pedro	0	0	Vitacura	20	17.967
Alhue	0	0	Ñuñoa	22	10.815
Til Til	1	3.273	Providencia	23	20.110
Pirque	1	3.003	Lo Barnechea	24	16.252
Buin	1	0.829	La Reina	25	16.521
Curacaví	2	4.351	Peñalolén	26	6.307
El Monte	2	3.838	San Bernardo	26	5.346
Calera de Tango	3	8.011	Las Condes	43	12.761
Talagante	4	3.320	La Florida	45	7.353
Isla de Maipo	4	8.134	Maipú	65	7.231
Peñaflor	4	3.126	Puente Alto	74	7.331

Table 1.1: Lower and higher IDDM rates descriptive statistics in Metropolitan Region

risk. Figure 1.1 shows, a) the spatial distribution of population at risk and, b) raw IDDM incidence rates associated with each area.

From figure 1.1 it is possible to appreciate that communes with highest incidence of IDDM do not match with those with highest population at risk.

In figure 1.2 it is possible to appreciate the crude rates distribution according to rurality. Something evident is that urban communes as La Reina, Providencia, Lo Barnechea and

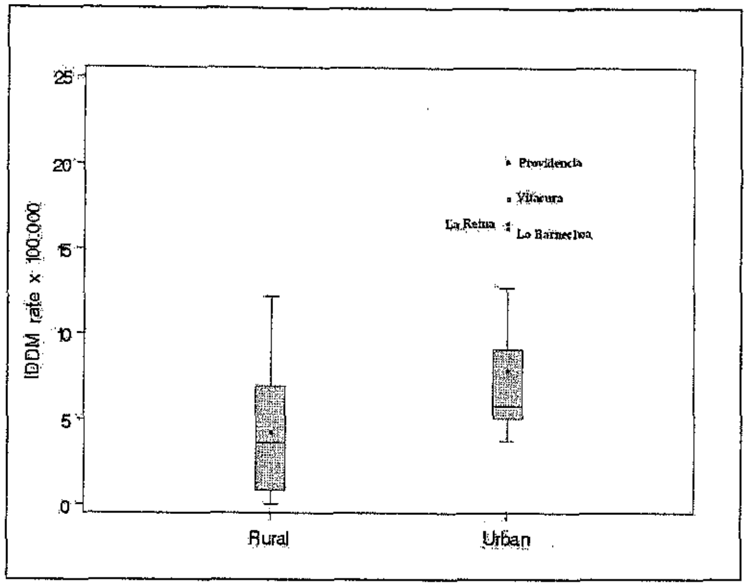

Figure 1.2: IDDM incidence rates boxplot

Vitacura present highest rates in comparison with the other urban areas. In addition, these communes belong to the group with highest socioeconomic levels in Metropolitan Region. This effect can also be appreciated in table 1.1, where 25% of the communes with the highest raw rates of the region, includes them. That is, even when most cases are related to the communes with higher population, such as La Florida, Puente Alto and Maipú, the highest incidence rates are associated to communes with high socioeconomic levels.

Figure 1.3 shows an usual behavior of rates, when they are related to population size. That is, incidence rates related to low population have greater variability.

A study of the spatial distribution of IDDM in Gran Santiago area [86] was developed based on usual Bayesian statistical methods, modeling the number of cases annually between the years 2000-2005. This study showed that communes with high incidence are associated with high socioeconomic levels.

This problem is of interest because rural areas rates tend to be influenced by their urban neighbors. The specification of a robust random effect model will help to achieve better

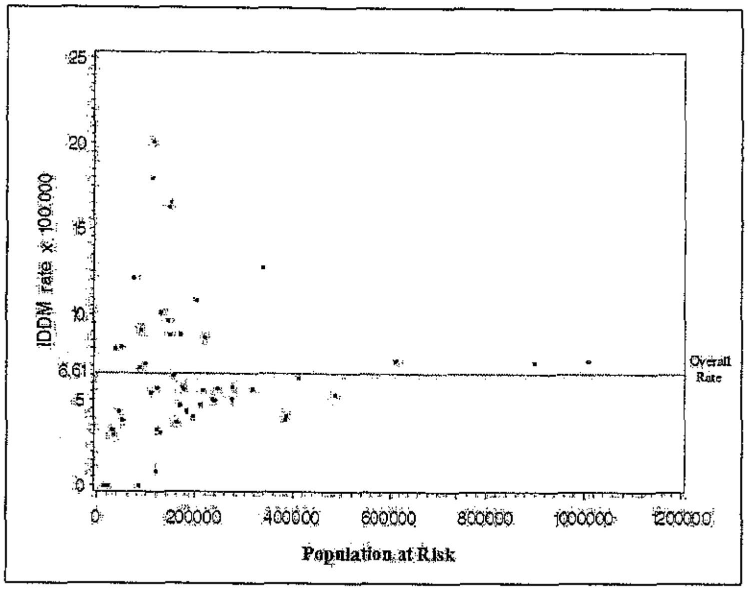

Figure 1.3: IDDM rates vs population at risk

epidemiological results.

1.4.2 Female trachea, bronchi and lung cancer mortality.

Cardiovascular mortality statistics from 1997 to 2004 published by Ministry of Health of Chile and population from 2002 census are used by Icaza and Núñez [50] to calculate sex and age standardized mortality rates. GLMM ([18], [93]) were used, under the assumptions of different slopes for the random effects and independence among communes. The results were published as the Chilean cardiovascular mortality atlas. Some other works related to this research problem are the master theses developed by Díaz [34] and Orellana [71]. Recently, Icaza et al. [51] show the estimation of mortality relative risks using the spatial convolution model discussed in the literature ([17], [68]) for stomach cancer and bronchi, trachea and lung cancer in Chile.

In this thesis work, the application is aimed at estimating the risk in female population deaths related to cancer associated with organs of the respiratory system, lung, trachea or

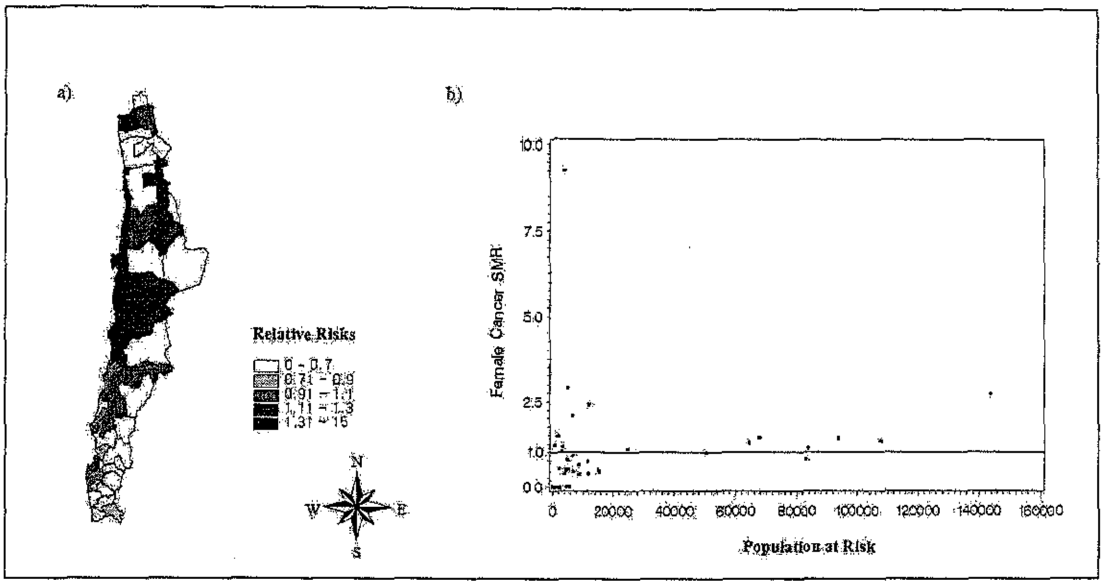

Figure 1.4: Female trachea, bronchi and lung cancer standardized mortality ratio

bronchi. Even when the estimation procedure is applied to the complete country, only 43 communes results, which belong to the five regions of northern Chile, will be shown.

Figure 1.4 show the variability associated with the mortality risk of lung, trachea and bronchi cancer in the north area of the country. It is evident the association between the population size and the cancer SMR variability. This effect produce the erratic SMR results which are displayed in the aside map.

SMR distribution considered by region is shown in figure 1.5. Evident differences can be seen from the plots. Region of Antofagasta (II) shows an excess in the variability of SMR in contrast to the other regions. Even more, in this region the median is around 2 an present an extremely high risk in the commune of Mejillones (SMR=9.25).

Icaza et al. [51] presented the distribution of cancer mortality risks, based on the aged standardized mortality ratios for the whole country, with no distinction of gender. The problem arises when the phenomenon is stratified by sex, delivering too smooth estimates and some communes with zero observed cases present high risks.

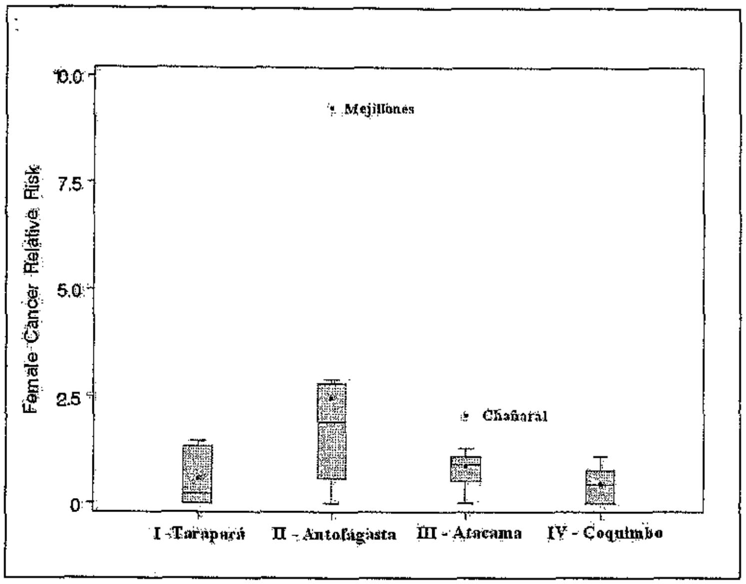

Figure 1.5: Female lung, trachea and bronchi cancer standardized mortality ratio

It is expected that application of robust models to this problem will help to obtain better estimates of relative risks, making them more appropriate for interpretations in epidemiology.

1.5 Thesis goals and structure.

The specific goals of this work are:

- to propose Bayesian robust models for the spatial random effects ([68], [67]), defining a more general family of symmetric Markov random fields based on the stochastic representation of scale of mixture of normal distributions,
- to construct of Bayesian hierarchical models, which includes likelihood functions derived from the generalized linear model theory and robust Markov random fields for the spatial random effects in different stages,
- to apply MCMC techniques and related algorithms in order to make inference possible, in the context of the proposed model,

- to discuss and propose a methodology to evaluate Bayesian learning, as an alternative to identifiability study,
- to implement of MCMC algorithms on simulated and real data.

This thesis is structured as follows.

Chapter 2 present extensions of usual Markov random fields, which are frequently used for spatial modelling aspects. Several theoretical aspects and conditions are established under the assumption of joint multivariate scale mixture of normal distributions.

In chapter 3 a generalized Bayesian hierarchical model is proposed. Required integrability conditions are presented under the standard assumptions and the extension presented in chapter 2. Bayes procedure is implemented via MCMC integration techniques, specifically through the Gibbs sampler algorithm.

Chapter 4 contains the analysis of two real data sets previously described. Results considering standard and proposed models are obtained.

Exploratory Bayesian learning is measured based on two discrepancy measures (q-divergence), L_1 discrepancy distance and Kullback Leibler divergence. Methodology and discussions are exposed in chapter 5.

Finally, comments and conclusions are discussed in chapter 6, as well as, future work and remained opened directions.

Chapter 2

Scale mixture of normal distributions and related Markov random fields.

2.1 Introduction

Scale mixture of normal (SMN) distributions have been proposed as extensions of the normal model. They have been defined as a subclass of the elliptical distributions family by Fang and Anderson [37]. This subfamily present similar properties to the normal distribution, with the exception that their behavior let capture unusual patterns present in the data. Kano [55] and Gupta and Varga [49] studied conditions in order to guarantee that SMN distribution exists. Specifically, Kano [55] established that a mixture distribution is well defined if a known and positive random variable with a cumulative density function P_{ψ}, called mixing term, exists. Gómez - Sánchez - Manzano et al. [47] proved and gave conditions in order to this subclass could belong as sequence terms of elliptical distributions.

Robust linear models has been studied since West [91], based on the stochastic representation of SMN distributions, proposed a Bayesian regression linear models to detect outliers.

Lange et al. ([59], [60]) present theory and computational issues to obtain maximum likelihood estimations through EM algorithm.

Several Bayesian works have used this stochastic representation to make inferences ([43], [38], [39]). Student-t distribution has also been characterized through SMN distributions ([43], [38], [59]). In a Bayesian context, Fernandez and Steel [39] showed the Bayesian inference validity and the existence of posterior moments, when usual non informative prior distributions are available. Specifically, Geweke [43] and Fernandez and Steel [38] discuss the advantage and pitfalls of modeling data when a Student-t linear regression model is considered.

From a MRF context, this class of mixtures can be found in papers developed from a geological point of view, where prediction is the main focus. Student-t distributed MRF was treated by Roislien and Omre [77], using a frequentist approach. Lyu and Simoncelli [64] made the extension of GMRF theory to what they called Gaussian Scale Mixture Fields, in image reconstruction modeling.

In this work, SMN theory is applied to extend the GMRF model ([9], [13]) when epidemiological data is available. Then, spatial random effect density is assumed following a robust distribution, in the sense of heavier tailed behavior in comparison to normal case.

2.2 Scale mixtures of the normal distribution

According to Andrews and Mallows [3], West([91] [92]), Fang and Anderson [37] and Fernandez and Steel [39], a SMN distribution is generated if the variable of interest, u, can be represented as

$$u = \mu + \psi^{-1/2} z, \tag{2.2.1}$$

where μ is a location parameter, and z and ψ are independent random variables, with z following an standard normal distribution and ψ having a c.d.f P_ψ such that $P_\psi(0) = 0$.

Therefore, this characterization of heavier tailed distributions compare to the normal model, can be defined in a multivariate way as follows:

Definition 2.2.1. Let $\psi_1, \ldots, \psi_m$ be a non observable sequence of independent and identically distributed (iid) sequence random variables, following a (known) probability distribution with cumulative density function P_ψ on $(0, \infty)$, that is, $P_\psi(0) = 0$. A random vector $\mathbf{u} = (u_1, \ldots, u_m)'$ follows a multivariate SMN distribution, if its density can be specified as

$$\pi(\mathbf{u}|\boldsymbol{\mu}, \sigma_u^2) = \int_{(0,\infty)^m} \frac{|\Psi|^{1/2}}{(2\pi\sigma_u^2)^{m/2}} \exp\left\{-\frac{1}{2\sigma_u^2}(\mathbf{u}-\boldsymbol{\mu})'\Psi^{-1}(\mathbf{u}-\boldsymbol{\mu})\right\} dP_\Psi, \tag{2.2.2}$$

where $\Psi = diag(\psi_1, \ldots, \psi_m)$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ is a location vector, σ_u^2 is a dispersion parameter, and P_Ψ represents the cumulative density function of a product probability measure on $\mathbb{R}^+$.

Notice that if the scale random factors $\psi_1, \ldots, \psi_m$ are all different, then from the above definition it follows that the components $u_1, \ldots, u_m$ of the random vector $\mathbf{u}$ are independent, with $u_i \sim SMN(\mu_i, \sigma_u^2)$. While if $\psi_1 = \psi_2 = \ldots = \psi_m = \psi$, then the u_i's will be uncorrelated $SMN(\mu_i, \sigma_u^2)$ random variables, but not (necessary) independent. Moreover, under this last assumption, a more general dependence structure between the u_i's can be introduced by replacing the scalar dispersion parameter σ_u^2 by a $m \times m$ dispersion matrix Σ.

Definition 2.2.2. A random vector $\mathbf{u} = (u_1, \ldots, u_m)'$ follows a dependent multivariate SMN distribution, with location parameter $\boldsymbol{\mu}$ and dispersion matrix Σ, if its density is defined by

$$\pi(\mathbf{u}|\boldsymbol{\mu}, \Sigma) = \int_{(0,\infty)} \frac{\psi^{m/2}}{(2\pi)^{m/2}} |\Sigma|^{-1/2} \exp\left\{-\frac{\psi}{2}(\mathbf{u}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{u}-\boldsymbol{\mu})\right\} dP_\psi. \tag{2.2.3}$$

where ψ is a mixing random factor with cumulative density function P_ψ defined on $\mathbb{R}^+$.

Some properties, related to moments, marginal and conditional distributions based on the elliptical family of distributions are derived by Fang and colleagues ([36], [37]). These authors construct these class of distributions through the characterization of a less complex symmetric family, called spherical, through $\mathbb{E}(e^{i\mathbf{t}'\mathbf{x}}) = \phi(\mathbf{t}'\mathbf{t})$, which is called the characteristic generator function. From this fact, marginal and conditional distributions can be obtained with their respective moments, previous application of a linear transformation.

Kano [55] gives conditions in order to construct SMN distributions with a *preferable consistency property*, in the sense that, the marginal distribution belongs to the same family as the joint distribution. He established that elliptical distributions with the consistency property must be a SMN distribution.

As the multivariate normal distribution, each multivariate elliptical distribution depends on a location vector $\boldsymbol{\mu}$ and a dispersion matrix Σ. A similar fact occurs with each conditional elliptical distribution. In that follows, we consider the partition given by

$$\begin{pmatrix} u_i \\ \mathbf{u}_{-i} \end{pmatrix}, \quad \begin{pmatrix} \mu_i \\ \boldsymbol{\mu}_{-i} \end{pmatrix}, \quad \begin{pmatrix} \sigma_i^2 & \Sigma_{i,-i} \\ \Sigma_{-i,i} & \Sigma_{-i} \end{pmatrix},$$

where $-i$ represent the set of indexes that excludes the i-th component. Following Kano [55], each conditional SMN distribution belongs to the same family of the original SMN distribution, and they are symmetric location-scale distributions. In particular, for each SMN family of distributions, the location and scale parameters of the conditional distribution of u_i given $\mathbf{u}_{-i}$ are the form of

$$\begin{aligned} \mu_{i|-i} &= \mu_i + \Sigma_{i,-i}\Sigma_{-i}^{-1}(\mathbf{u}_{-i} - \boldsymbol{\mu}_{-i}) \\ \sigma_{i|-i}^2 &= \sigma_i^2 - \Sigma_{i,-i}\Sigma_{-i}^{-1}\Sigma_{-i,i}. \end{aligned} \tag{2.2.4}$$

This representation will be useful to determine SMN conditional distributions, which are important for the posterior analysis in the MRF context.

2.3 Scale mixture of normal random fields

Besag [9] present a pioneer work in the context of the MRF theory, with applications to regular lattice systems, when spatial heterogeneity is considered. This is a key reference to state the statistical modeling in this area.

Recent results were obtained by Kaiser and Cressie [54] for the construction of MRF when conditional distributions are available. Other results are exposed in Lee et al. [63], extending the idea of the construction of a MRF for distributions that belong to exponential family, relaxing assumptions like pairwise dependence [9].

In this work, an extension of the usual multivariate GMRF will be developed, by assuming a multivariate SMN distribution. Pairwise dependence will be assumed, which is natural when the MRF is represented by a Gaussian kernel. Pairwise dependence will be consistent with a SMN kernel, due to the relationship between the Gaussian and SMN distributions.

As mentioned in chapter 1, a multivariate Gaussian distribution derived from a MRF, is

$$\pi(\mathbf{u}|\sigma_u^2) \propto \exp\left\{-\frac{1}{2\sigma_u^2}\mathbf{u}'D_w\mathbf{u}\right\}, \tag{2.3.1}$$

where $\mathbf{u} \in \mathbb{R}^m$, σ_u^2 represents the scale parameter of the MRF and D_w is a $m \times m$ proximity matrix, with diagonal elements w_{i+} representing the number of neighbors of the i-th component, and off-diagonal elements w_{ij} taking values -1 if the elements i and j share boundary and 0 in other case, i.e.,

$$w_{ij} = \begin{cases} w_{i+} & i = j \\ -1 & i \neq j; i \sim j \\ 0 & \text{otherwise.} \end{cases} \tag{2.3.2}$$

The notation $i \sim j$ is used for communes i and j which are contiguous, and they are neighbors if both communes share a common border. A basic discussion and treatment of several proximity matrices can be found in Banerjee et al. [6].

The next definition will provide an extension of (2.3.1) to the SMN random field (SMN RF).

Definition 2.3.1. A spatial random vector $\mathbf{u} = (u_1, \ldots, u_m)'$ follows a SMN RF, if the kernel density is specified as

$$\pi(\mathbf{u}|\sigma_u^2) \propto \int_0^\infty \psi^{m/2} \exp\left\{-\frac{\psi}{2\sigma_u^2}\mathbf{u}'D_w\mathbf{u}\right\} dP_\psi, \tag{2.3.3}$$

where ψ is a positive random variable with a known c.d.f. P_ψ, σ_u^2 is a dispersion parameter and D_w denotes a proximity matrix. A SMN RF with scale parameter σ_u^2 will be denoted as $SMNRF(\mathbf{0}, \sigma_u^2 D_w^{-1}, \nu)$, where ν is an additional parameter (or set of parameters) which controls the tails behavior.

For the Gaussian case, it is known that specification of D_w 2.3.2 makes (2.3.1) improper [6], since the matrix D_w is singular, so that D_w^{-1} does not exist, hence

$$\int_{\mathbb{R}^m} \pi(\mathbf{u}|\sigma_u^2)\mathbf{du} \propto \int_{\mathbb{R}^m} \exp\left\{-\frac{1}{2\sigma_u^2}\mathbf{u}'D_w\mathbf{u}\right\}\mathbf{du} = \infty.$$

The last equation implies that a density function is available, but not integrable. This result is the IAR model property, and it is usually relegated to the prior distribution elicitation. If additional assumptions are not considered, the improper condition will imply that if a multivariate SMN RF is assumed with kernel 2.3.1, then consistent property [55] fails. Therefore, integration theory can not be applied.

As the same way of the joint distribution of the GMRF treated in the spatial literature, for every SMN RF, the joint distribution will be also improper. In fact, this distribution will be proper only if the associated dispersion matrix is definite positive. Hence, some additional restrictions should be imposed to obtain a proper joint distributions, as is discussed in Banerjee et al. [6] and Assunção et al. [5]. The following proposition establishes conditions to makes proper the SMN RF associated joint distribution.

Proposition 2.3.1. *Suppose that a set of spatial indexed random variables, represented by the vector* $\mathbf{u} = (u_1, \ldots, u_m)'$, *is available. Consider the SMN RF in (2.3.1) as the distribution of* $\mathbf{u}$. *Additionally, let suppose that* P_ψ *is a known positive cumulative density function. If* $\sum_{i=1}^m u_i = 0$ *and* $\mathbb{E}(\psi^{1/2}) < \infty$, *then (2.3.3) is proper.*

Proof. As was showed by Assunção et al. [5], the $\sum_{i=1}^m u_i = 0$ constraint makes the Gaussian kernel 2.3.1 proper. The proof of this result involves the usual spectral decomposition of D_w, that is, $D_w = \mathbf{P}'\mathbf{A}\mathbf{P}$, where the matrices $\mathbf{A}$ and $\mathbf{P}$ represent the eigenvalues and the eigenvectors of D_W, respectively. Thus, if the transformation $\mathbf{x} = \mathbf{Pu}$ is applied, is known [5] that,

$$\int_{\mathbb{R}^m} \exp\left\{-\frac{1}{2\sigma_u^2}\mathbf{u}'D_w\mathbf{u}\right\} \mathbf{du} \propto \int_{\mathbb{R}^{m-1}} \exp\left\{-\frac{1}{2\sigma_u^2}\mathbf{x}'\mathbf{A}\mathbf{x}\right\} \mathbf{dx} < \infty. \tag{2.3.4}$$

Hence, under the $\sum_{i=1}^m u_i = 0$ constraint, we have in (2.3.4) by applying the Fubini's theorem and the change variable $\mathbf{y} = \psi^{1/2}\mathbf{x}$ that

$$\begin{aligned}\int_{\mathbb{R}^m} \pi(\mathbf{u}|\sigma_u^2)d\mathbf{u} &= \int_0^\infty \psi^{m/2} \int_{\mathbb{R}^m} \exp\left\{-\frac{\psi}{2\sigma_u^2}\mathbf{u}'D_w\mathbf{u}\right\} d\mathbf{u}dP_\psi \\ &\propto \int_0^\infty \psi^{m/2} \int_{\mathbb{R}^{m-1}} \exp\left\{-\frac{\psi}{2\sigma_u^2}\mathbf{x}'\mathbf{A}\mathbf{x}\right\} d\mathbf{x}dP_\psi \\ &\propto \int_0^\infty \psi^{1/2} d\psi < \infty.\end{aligned}$$

□

It is of interest to find the full conditional distributions. They are necessary to make posterior inferences when MCMC methods are applied. In this sense, we consider the next result.

Corollary 2.3.2. *Under the conditions of the Proposition 2.3.1, the full conditional distributions arising from a SMN RF are given by*

$$u_i|\mathbf{u}_{-i}, \sigma_u^2 \sim SMN(\mu_{i|-i}, \sigma_{i|-i}^2, \nu), i \neq j, \tag{2.3.5}$$

where,

$$\mu_{i|-i} = \frac{\sum_{j\neq i} \omega_{ij} u_j}{\sum_{j\neq i} \omega_{ij}} \quad \text{and} \quad \sigma_{i|-i}^2 = \frac{\sigma_u^2}{\sum_{j\neq i} \omega_{ij}}. \tag{2.3.6}$$

It is possible to obtain the distribution of a SMN RF directly. Following Besag [9], for each fixed P_ψ, and applying the Brook's lemma ,$\pi(\mathbf{u}, \psi)$ can be obtained by using the conditional-marginal decomposition; that is, the distribution of a SMN RF can be obtained as follows: Let $\mathbf{u}$ and $\mathbf{v}$ be two arbitrary realizations in $\Omega = \{\mathbf{u} : P(\mathbf{u}) > 0\} = \{\mathbf{v} :$

$P(\mathbf{v}) > 0\}$. Then the distribution of the SMN-RF under consideration, which seems to be reasonably well defined, can be obtained as

$$\frac{\pi(\mathbf{u},\psi)}{\pi(\mathbf{v},\psi)} = \prod_{i=1}^{n} \frac{\pi(u_i|v_1, v_2, \ldots, v_{i-1}, u_{i+1}, \ldots, u_m, \psi)}{\pi(v_i|u_1, u_2, \ldots, u_{i-1}, v_{i+1}, \ldots, v_m, \psi)}. \tag{2.3.7}$$

2.4 Some specific scale mixture of normal random fields

Specific choices for P_ψ in (2.3.3) leads to different scale mixture probability distributions. Student-t and slash MRF's will be treated in this work, which can be characterized using earlier results. Both distributions can be obtained by using stochastic representations, which depends on the selected mixing distribution P_ψ.

2.4.1 Hierarchical specification

The SMN RF can be represented hierarchically in term of stages: The SMN RF model is developed by introducing, at the first stage of the hierarchy, a GMRF. At the second stage a mixing distribution for the scale perturbation must be specified. Specifically:

1. The Student-t MRF:

$$\text{i)} \quad \mathbf{u}|\sigma_u^2, \psi \sim Normal\left(\mathbf{0}, \sigma_u^2 \psi^{-1} D_w^{-1}\right) \tag{2.4.1}$$

$$\text{ii)} \quad \psi \sim Gamma(\nu/2, \nu/2). \tag{2.4.2}$$

   In this case, the Student-t MRF with ν degrees of freedom follows, which will be denote by $\mathbf{u}|\sigma_u^2 \sim t(\mathbf{0}, \sigma^2 D_w^{-1}, \nu)$.

2. The slash MRF:

$$\text{i)} \quad \mathbf{u}|\sigma_u^2, \psi \sim Normal\left(\mathbf{0}, \sigma_u^2 \psi^{-1} D_w^{-1}\right), \tag{2.4.3}$$

$$\text{ii)} \quad \psi \sim Beta(\nu/2, 1). \tag{2.4.4}$$

   In this case, the slash MRF, denoted by $\mathbf{u}|\sigma_u^2 \sim Slash(\mathbf{0}, \sigma_u^2 D_w^{-1}, \nu)$, is obtained.

Here, D_w is an adjacency matrix defined earlier in (2.3.1), $\sigma_u^2 > 0$ is a scale parameter and $\nu > 0$ is a shape parameter. These hierarchical structures will be useful to implement the MCMC method.

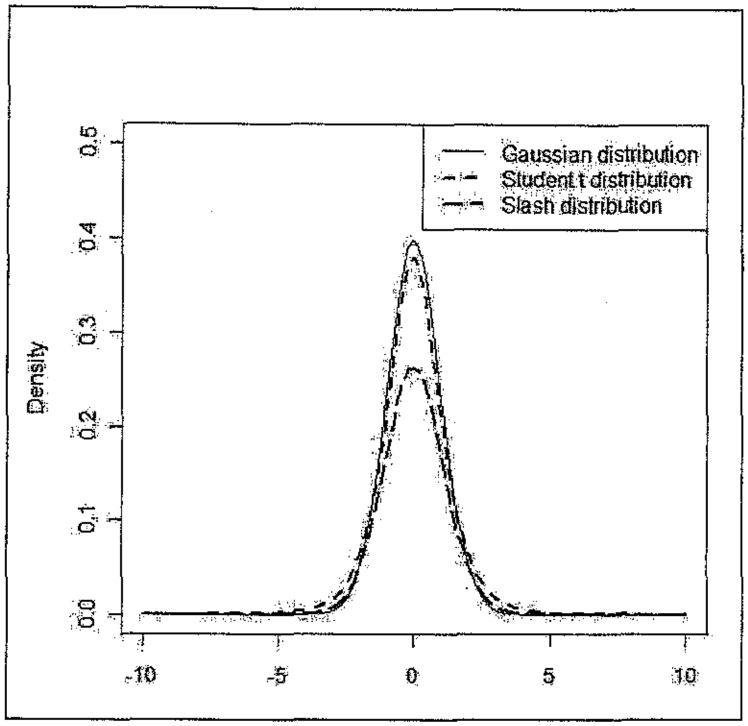

Figure 2.1: Specific standard scale mixture of normal distributions (Gaussian, Student-t (5) and Slash(5))

It is important to mention that the distribution of both of the above random fields have the finite condition exposed in Proposition 2.3.1. Graphical comparisons between standard normal, Student-t (with $\nu = 5$ d.f.) and slash (with $\nu = 5$ d.f.) distributions can be visualized in figure 2.1. Notice that even when the slash distribution present heavier tails than the normal distribution, its behavior is lighter than the Student-t distribution.

Prior distribution for σ_u^2 is required in order to assume a valid Bayesian model. Usually, a proper non informative type of prior distribution is considered for this parameter. For this work, a special inverse gamma distribution can be used, in other words, by considering

$$\sigma_u^{-2} \sim Gamma(a, b), \quad (a,\, b > 0), \tag{2.4.5}$$

with $a, b \to 0$.

2.4.2 Full posterior distributions

For the Gibbs Sampling scheme, full posterior distributions are necessary. They will depend on the distribution assumed for the random field.

If (2.4.1)-(2.4.2) are assumed, which is equivalent to the Student-t MRF formulation, and the prior in (2.4.5) is considered for σ_u^2, then the full conditional distributions presented in Algorithm I are obtained for MCMC implementation. By the other side, if (2.4.3)-(2.4.4) are assumed, that is a Slash RF modeling, taking into account the prior (2.4.5) for σ_u^2, then the full conditional distributions are shown in Algorithm II.

Algorithm I. [The Student-t RF]

1. $\sigma_u^{-2}|\psi, \mathbf{u}, \nu \sim Gamma\left(a + \frac{m}{2}, \frac{\psi}{2}(\mathbf{u}'D_w\mathbf{u}) + b\right)$,
2. $\psi|\sigma_u^2, \mathbf{u}, \nu \sim Gamma\left(\frac{m+\nu}{2}, \frac{1}{2\sigma_u^2}(\mathbf{u}'D_w\mathbf{u}) + \frac{\nu}{2}\right)$.

Algorithm II. [The slash RF]

1. $\sigma_u^{-2}|\psi, \mathbf{u}, \nu \sim Gamma\left(a + \frac{m}{2}, \frac{\psi}{2}(\mathbf{u}'D_w\mathbf{u}) + b\right)$,
2. $\psi|\sigma_u^2, \mathbf{u}, \nu \sim Gamma\left(\frac{m+\nu}{2}, \frac{1}{2\sigma_u^2}(\mathbf{u}'D_w\mathbf{u})\right) \mathbf{1}_{[0,1]}(\psi)$.

Each algorithm must be iterated until convergence is reached. Notice that the Algorithm II present a truncated gamma distribution in [0, 1] interval. To draw from this distribution, the Damien and Walker [31] algorithm can be performed.

2.5 Simulation study

This section presents a simulation study, which allows to assess the behavior and validity of the different proposed models. It also allows to compare their behavior when the data are drawn under controlled parameter values.

Rue and Held [79] present several algorithms to draw GMRF. In this study, a dependent multivariate GMRF is drawn, constrained to sum zero, to be consistent with the theory

Real RF	df	Estimated RF	Dispersion parameter: σ^2			
			$\sigma_u^2 = 1$			
			Mean	Std. Dev.	95% HDP CI	MC Error
Std-t	10	Normal	1.178	0.554	(0.475,2.947)	0.028
		Std-t	1.182	0.556	(0.455,2.883)	0.029
		Slash	0.980	0.456	(0.397,2.446)	0.000
	50	Normal	1.109	0.373	(0.656,2.813)	0.016
		Std-t	1.109	0.373	(0.653,2.801)	0.015
		Slash	1.066	0.358	(0.629,2.690)	0.006
	100	Normal	1.083	0.251	(0.663,1.768)	0.013
		Std-t	1.082	0.250	(0.675,1.756)	0.013
		Slash	1.061	0.246	(0.653,1.735)	0.008
Slash	10	Normal	1.349	0.468	(0.8270,3.048)	0.117
		Std-t	1.354	0.469	(0.8353,3.011)	0.120
		Slash	1.127	0.395	(0.6883,2.615)	0.020
	50	Normal	1.070	0.193	(0.7092,1.517)	0.012
		Std-t	1.071	0.194	(0.7100,1.522)	0.012
		Slash	1.028	0.186	(0.6858,1.456)	0.002
	100	Normal	0.992	0.179	(0.6596,1.337)	0.000
		Std-t	0.991	0.179	(0.6518,1.327)	0.000
		Slash	0.972	0.174	(0.6535,1.300)	0.002
			$\sigma_u^2 = 25$			
Std-t	10	Normal	30.260	16.123	(12.361,111.698)	0.844
		Std-t	30.217	16.111	(12.240,110.991)	0.832
		Slash	25.256	13.472	(10.221,92.786)	0.002
	50	Normal	26.032	4.551	(15.273,34.252)	0.116
		Std-t	25.981	4.564	(15.212,34.174)	0.105
		Slash	25.036	4.390	(14.622,32.912)	0.001
	100	Normal	25.872	7.083	(15.934,48.969)	0.054
		Std-t	25.875	7.040	(16.143,48.737)	0.055
		Slash	25.351	6.907	(15.748,47.767)	0.009
Slash	10	Normal	32.763	13.688	(18.062,117.652)	2.068
		Std-t	32.790	13.594	(18.212,114.883)	2.094
		Slash	27.308	11.370	(14.884,97.563)	0.234
	50	Normal	26.023	4.555	(15.168,34.303)	0.115
		Std-t	26.009	4.523	(15.139,34.125)	0.112
		Slash	25.009	4.367	(14.553,32.964)	0.000
	100	Normal	26.352	4.526	(18.560,37.004)	0.199
		Std-t	26.332	4.471	(18.915,36.420)	0.197
		Slash	25.813	4.376	(18.442,35.819)	0.075
			$\sigma_u^2 = 100$			
Std-t	10	Normal	118.519	64.830	(56.323,416.717)	2.619
		Std-t	118.481	64.305	(55.887,406.197)	2.628
		Slash	98.942	54.123	(47.890,342.590)	0.011
	50	Normal	104.108	20.670	(66.039,155.863)	0.408
		Std-t	104.088	20.827	(66.434,157.633)	0.401
		Slash	100.051	19.756	(63.725,148.841)	0.000
	100	Normal	103.896	26.629	(59.750,168.915)	0.286
		Std-t	103.850	26.782	(63.240,172.941)	0.278
		Slash	101.874	26.152	(61.082,169.528)	0.067
Slash	10	Normal	122.152	35.826	(80.874,274.819)	6.355
		Std-t	121.998	35.611	(82.821,268.380)	6.305
		Slash	101.728	29.962	(67.263,228.890)	0.050
	50	Normal	104.191	20.813	(67.799,157.736)	0.421
		Std-t	104.078	20.624	(67.017,155.590)	0.403
		Slash	100.168	19.979	(64.152,150.602)	0.001
	100	Normal	104.847	19.024	(68.017,142.582)	0.614
		Std-t	104.752	19.008	(67.113,141.934)	0.591
		Slash	102.717	18.593	(66.083,139.261)	0.200

Table 2.1: Dispersion parameter estimation: Posterior mean, standard deviation, 95% HPD credibility intervals for simulations and Monte Carlo error

exposed in this chapter. Several samples were simulated, specifically 100 samples of size $m = 52$ for each model were obtained. The spatial adjacency matrix is based on the Chilean Metropolitan region. Each sample coming from one of the studied models will be called replication. Estimations for ψ and σ_u^2 were obtained in order to have a wide picture of the different values that these parameters can take.

Here, σ_u^2 is the parameter of interest, because it express the variability of the random field. In the next chapter, its estimate will be useful to obtain the percentage of variability explained by the spatial component of the model.

Scenarios for MRF simulations includes:

- Student-t MRF with $\nu = 10, 50, 100$,
- Slash MRF with $\nu = 10, 50, 100$.

An important aspect is to verify the ability of the model to estimate σ_u^2. Thus, three different values for this parameter are taking into account for all models, $\sigma_u^2 = 1, 25, 100$. The degrees of freedom parameter ν is assumed to be known. This aspect will be discussed in the later chapters, under an epidemiological context.

High posterior density (HDP) credibility interval lengths needs to be considered, that is, narrower intervals will be and evidence of a better precision on estimation process, and distributions associated to these intervals are potential candidates to be chosen. Posterior mean and standard deviation will measure the point estimation for parameters of interest.

Table 2.1 shows inferences related to the unknown parameter σ_u^2. Specifically, it gives the posterior mean, the standard deviation, the 95% HDP credibility interval and the Monte Carlo (MC) error obtained from the simulated samples. These results show that most of the credibility intervals associated to the true model present better estimations, improving its behavior when SMN RF tends to a GMRF. When the Slash distribution is the target distribution, estimation for σ_u^2 present a better behavior for all σ_u^2, showing better model fit in comparison with other proposed MRF.

Small degrees of freedom must be treated carefully, because it is possible to find similar cases to Cauchy distribution, where first and second moments are not defined; the slash distribution present a similar behavior, when this fact occurs.

Chapter 3

Robust small area modeling

3.1 Introduction

Bayesian spatial models has become increasingly popular for epidemiologists and statisticians over the last two decades. A pioneer work in this direction was developed by Clayton and Kaldor [28] who proposed empirical Bayes approach with application to lip cancer data in Scotland. MCMC methods yield to an explosive increment of the use of Bayesian analysis in different areas of application. In particular, in the context of spatial epidemiology, several works can be mentioned; some of the most relevant are commented in the next lines. In Ghosh et al. [44], conditions to demonstrate Bayesian GLM integrability are formalized. Integrability aspects are important since improper priors [8] are used to represent lack of knowledge over unknown parameters. Best et al. [16], investigated several spatial prior distributions, based on MRF theory, and present discussions related to methods for model comparison and diagnostics. Pascutto et al. [72] examined some structural and functional assumptions of these models and illustrate its sensitivity through the presentation of results related to informal sensitivity analysis for prior distributions choices. They also explored the effect that cause outlying areas, assuming a Student-t distribution for the non structured effect. Gelfand et al. [57], developed nonparametric methodologies, based on Dirichlet processes as prior distributions with its respective computational implementation.

The works of Best et al. [17] and Waller [89] contain comprehensive reviews. Most common models used in the area can be found in the books of Banerjee et al. [6] and Lawson [62].

This chapter shows the development of robust spatial Bayesian models to detect unusual rates or relative risks at a particular area. Robust models will be obtained using the SMN RF, formalized in chapter 2.

3.2 Spatial generalized linear mixed models

Assuming that a region of interest is divided into m independent areas, then the probabilistic representation of a GLMM [27] will be structured assuming (1.2.1), with general link function represented by (1.2.4). Depending on the focus of the problem and the available data, estimation of rates or relative risks, $\boldsymbol{\theta}$, is of interest. Ghosh et al. [44] established conditions for a GLMM, including the case when the spatial effect is included in the model, proving the existence of a proper posterior distribution under some restrictions, such as, elicitation of proper prior distributions for dispersion parameters and spatial effects restricted to sum zero. Chen et al. [25] developed a more specific methodology based on Bayesian GLMM, and characterized sufficient and necessary conditions in order to make Bayesian inference possible, identifying precise conditions that guarantee the posterior distribution property.

Let $\mathbf{y} = (y_1, \ldots, y_m)'$, a set of m random variables indexed to a specific region. A general formulation when a SMN distribution is assumed for spatial random effects, includes the following elements:

1. A general model can be specified as in (1.2.1), that is,

$$f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi}) = \prod_{i=1}^{m} \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i)) + \rho(\phi_i; y_i)\},$$

   where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$ is the vector of canonical parameters, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)'$ is a vector of known scale parameters and ρ is a known function that does not depend on the unknown parameters.

2. An associated link function represented by (1.2.4) can be considered, that is

$$h(\theta_i)|\mathbf{x}_i, \beta, u_i \overset{ind}{\sim} Normal(\mathbf{x}_i'\beta + u_i, \sigma^2),$$

where $\mathbf{x}_i$ is a $p \times 1$ vector of covariates associated to a $p \times 1$ vector of fixed effects β, and u_i's are spatially structured random effects. σ^2 measures the non-structured variability.

3. Let $\mathbf{u}$ follow a SMN RF, in the sense of (2.3.3), the spatial behavior is represented by,

$$\mathbf{u}|\sigma_u^2 \sim SMN\left(\mathbf{0}, \sigma_u^2 D_w^{-1}, \nu\right), \tag{3.2.1}$$

where σ_u^2 is the associated dispersion parameter, D_w is the adjacency matrix defined in (2.3.2) and ν are the degrees of freedom.

Equivalent to (3.2.1), an stochastic representation can be specified as follows,

a. $\mathbf{u}|\psi, \sigma_u^2 \sim Normal\left(\mathbf{0}, \psi^{-1}\sigma_u^2 D_w^{-1}\right),$ (3.2.2)

b. $\psi \sim P_\psi$, where P_ψ is an cumulative density function such as $P_\psi(0) = 0$,

where ψ is considered as a hidden parameter to reproduce SMN distributions.

In the literature, it is recurrent to find that the spatial random effect in (3.2.1), is influenced by a predefined neighborhood represented by the adjacency matrix D_w, controlling the local variability. Hence, the spatial random effect mean and dispersion are smoothed by the information given by its neighbors. The robust construction exposed here is attractive due to the existence of unusual zeros and/or the complex geographic structure of the country in which the application is of interest.

Instead of working with the usual assumption of normality for the random effect, (3.2.1) allows the extension to robust models for the random effects, using the Student-t distribution treated by Geweke [43] and the slash distribution proposed by Lange and Sinscheimer [60]. From a Bayesian point of view, constructions of (3.2.1) through the representation in

(3.2.2) seems to be a natural step of the analysis, allowing for the immediate Gibbs implementation.

Properties and conditions for the existence of posterior moments when SMN densities (2.2.2) are considered, were developed by Fernandez and Steel [39] in a general context, through the different SMN representations, with econometric and financial applications. In the spatial framework, Laplace and double exponential distributions [11] has been proposed as parametric robust alternatives. Datta and Lahiri [32] focused on Bayesian estimation with a prior scale mixture distribution, for the error component in a normal linear model, to smooth small area means when one or more outliers are present in the data. If the focus is nonparametric small area analysis Knorr-Held and Raßer [56], Cangnon and Clayton ([21], [22]) and Gelfand et al. [57], developed techniques and models to estimate relative risks.

As a final step of the modeling, prior distributions are required for the unknown parameters to complete the hierarchical model. Usual non informative prior distributions are represented by

$$\begin{aligned} &i. \quad \boldsymbol{\beta} \propto constant \\ &ii. \quad \sigma^{-2} \sim Gamma(a/2, b/2) \\ &iii. \quad \sigma_u^{-2} \sim Gamma(c/2, d/2), \end{aligned} \tag{3.2.3}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ and $a, b, c, d > 0$. σ^2 and σ_u^2 represents dispersion parameters included in the model. σ_u^2 is the local dispersion parameter related to a specific spatial structure. Other useful measure in spatial models is the computation of the proportion or percentage of spatial aggregation explained by the model, which is usually estimated by the ratio,

$$\frac{s_u^2}{s_u^2 + \sigma^2},$$

where s_u^2 is the empirical variance, which can be obtained from the estimation of $\boldsymbol{u}$ for each MCMC iteration. The interpretation is related to obtain the relative contribution given by the spatial aggregation effect.

Under the model (1.2.1), link function (1.2.4) and prior assumption (3.2.3), theorems 1 and 2 of the work developed by Ghosh et al. [44], give the conditions to obtain a proper posterior distribution for $\boldsymbol{\theta}|\mathbf{y}$ when $P(\psi = 1) = 1$. Following both theorems, it is possible to find a generalization towards the SMN case. The next proposition gives conditions when non-structured random effects are assumed.

Proposition 3.2.1. *Under the following assumptions:*

i. *The support of* $\theta_i \in (\underline{\theta}_i, \overline{\theta}_i)$, *for some* $-\infty < \underline{\theta}_i < \overline{\theta}_i < \infty$.

ii. $m - p + a > 0$

iii. $b > 0,\ d > 0,\ m + c > 0$

iv. u_i*'s* $\overset{iid}{\sim} Normal(0, \psi_i^{-1}\sigma_u^2),\ i = 1, \ldots, m.$

v. $\psi_i \overset{iid}{\sim} P_\psi, i = 1, \ldots, m,$ *such that* $P_\psi(0) = 0.$

If

$$\int_{\underline{\theta}_i}^{\overline{\theta}_i} \exp\{\phi_i^{-1}(y_i\theta - g(\theta))\}h'(\theta)d\theta < \infty,$$

$\forall i = 1, \ldots, m,$ *then* $\pi(\boldsymbol{\theta}|\mathbf{y})$ *is proper.*

Proof. The full joint posterior distribution present the following structure,

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma^2, \psi|\mathbf{y}) \propto \quad & \textstyle\prod_{i=1}^m \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\} \\ & \times \textstyle\prod_{i=1}^m \exp\{-(1/2\sigma^2)(h(\theta_i) - \mathbf{x}_i'\boldsymbol{\beta} - u_i)^2\}h'(\theta_i) \\ & \times \textstyle\prod_{i=1}^m \exp\{-(\psi_i/2\sigma_u^2)u_i^2\}\psi_i^{1/2}(\sigma^2\sigma_u^2)^{-1/2} \\ & \times \exp\{-a/2\sigma_u^2\}(\sigma^2)^{-(b/2+1)}\exp\{-c/2\sigma^2\}(\sigma_u^2)^{-(d/2+1)}P_\psi',\end{aligned}$$

where P_ψ' represent the density function of ψ. Integrating with respect to $\boldsymbol{\beta}$, σ^2 and σ_u^2, the kernel reduce to,

$$\begin{aligned}\pi(\boldsymbol{\theta}, \mathbf{u}, \psi|\mathbf{y}) \propto \quad & \textstyle\prod_{i=1}^m \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\}h'(\theta_i) \\ & \times \textstyle\prod_{i=1}^m \psi_i^{1/2}(a + \psi_i u_i^2)^{1/2(m+b)}P_\psi'.\end{aligned}$$

Notice that this last result is the product of m Student-t kernels. This fact let to integrate over $\mathbf{u} \in \mathbb{R}^m$, producing the following result,

$$\pi(\boldsymbol{\theta}, \psi|\mathbf{y}) \leq \quad C \prod_{i=1}^{m} \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\} h'(\theta_i) \times \prod_{i=1}^{m} P'_{\psi},$$

where C is a constant that does not depend on $\boldsymbol{\theta}$ or any of the parameters previously integrated. Finally, integration over ψ leads to the desirable result. □

In a spatial framework the treatment is different. The random effects $\mathbf{u}$ must verify the MRF properties, which implies to specify the same $\psi_i, \forall i = 1, \ldots, m$. The next proposition extend the previous results, when the spatial random effect follow a SMN RF.

Proposition 3.2.2. *Under the following assumptions:*

i. *The support of* $\theta_i \in (\underline{\theta}_i, \overline{\theta}_i)$, *for some* $-\infty < \underline{\theta}_i < \overline{\theta}_i < \infty$.

ii. $m - p + a - 1 > 0$

iii. $b > 0$, $d > 0$, $m + c > 0$

iv. u_i*'s distributed as (2.3.3) and constrained to sum zero.*

v. $\psi_1 = \psi_2 = \ldots = \psi_m = \psi \sim P_\psi$, *with* $\mathbb{E}(\psi^{1/2}) < \infty$.

If the condition of integrability in the proposition 3.2.1 is verified then $\pi(\boldsymbol{\theta}|\mathbf{y})$ *is proper.*

Proof. The full joint posterior distribution is specified as follows,

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma^2, \psi|\mathbf{y}) \propto \quad & \prod_{i=1}^{m} \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\} h'(\theta_i) \\ & \times \prod_{i=1}^{m} \exp\{-(1/2\sigma^2)(h(\theta_i) - \mathbf{x}_i'\boldsymbol{\beta} - u_i)^2\} \\ & \times \exp\{-(\psi/2\sigma_u^2)\mathbf{u}' D_w \mathbf{u}\} \\ & \times \psi^{m/2}(\sigma^2\sigma_u^2)^{-m/2} \exp\{-a/2\sigma_u^2\} \\ & \times (\sigma^2)^{-(b/2+1)} \exp\{-c/2\sigma^2\}(\sigma_u^2)^{-(d/2+1)} P'_\psi,\end{aligned}$$

where P'_ψ represent the density function of ψ. Integrating with respect to $\boldsymbol{\beta}$, σ^2 and σ_u^2, the following joint distribution is obtained,

$$\pi(\boldsymbol{\theta}, \mathbf{u}, \psi|\mathbf{y}) \propto \quad \prod_{i=1}^{m} \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\} h'(\theta_i) \times \psi^{m/2}(a + \psi\mathbf{u}' D_w \mathbf{u})^{1/2(m+b)} P'_\psi.$$

Notice that under sum zero constraint condition, the last function in the above relation correspond to a $m-1$ Student-t kernel. Thus, integration on $\mathbf{u}$ is performed over $\mathbb{R}^{m-1}$, giving the following result,

$$\begin{aligned}\pi(\boldsymbol{\theta},\psi|\mathbf{y}) \leq \quad & C\prod_{i=1}^{m}\exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\}h'(\theta_i)\\ & \times\psi^{1/2}P'_{\psi},\end{aligned}$$

with C, constant which not include any of the parameters previously mentioned. As $\int_{Rec(\psi)}\psi^{1/2}dP_{\psi} < \infty$, then the result is obtained. □

3.3 Markov chain Monte Carlo schemes

When a hierarchical model is considered, the MCMC implementation requires the specification of full conditionals distributions. In this case, the implementation will depend on the choice of the hidden parameters. Analysis is treated separately, depending on random effects specification, given by Propositions 3.2.1 and 3.2.2.

The analytic problems that Bayesian models present has been widely discussed in the literature. MCMC methods has become the best solution to make inference. Gibbs sampling ([23], [42]) , Metropolis Hastings [26] and adaptive rejection [45] algorithms are the most discussed choices ([70], [69], [44], among others) in the context of GLM for small area analysis.

In order to obtain general expressions for the computational treatment of this model, the following matrix representations will be used for the implementation purposes,

$$\mathbf{h}(\boldsymbol{\theta}) = (h(\theta_1), \ldots, h(\theta_m))'$$

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$$

and $\mathbf{X} = [x_i], i = 1, \ldots, m; k = 1, \ldots, p$, defines an arbitrary design matrix.

3.3.1 Non-structured random effects

Consider the model (1.2.1), link function (3.2.1), stochastic representations given by (2.4.1)-(2.4.2) or (2.4.3)-(2.4.4) when $D_w = I_n$, and independent prior distributions for $\boldsymbol{\beta}$, σ^2, and σ_u^2 given by 3.2.3. Full conditional distributions, under the above assumptions and Proposition 2.3.1 when a non spatial structure is considered, is described by the algorithm III.

Algorithm III. [Non spatial dependence]

1. $\boldsymbol{\beta}|\mathbf{X}, \sigma^2, \mathbf{u} \sim Normal(\hat{\boldsymbol{\beta}}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ where,
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{h}(\boldsymbol{\theta}) - \mathbf{u})$$
2. $u_i|\boldsymbol{\theta}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2, \sigma_u^2, \psi_i \sim Normal(\mu_i^u, v_i^u)$ where,
$$\mu_i^u = (h(\theta_i) - \mathbf{x}_i'\boldsymbol{\beta})v_i^u \text{ and } v_i^u = \left(\frac{\psi_i}{\sigma_u^2} + \frac{1}{\sigma^2}\right)^{-1}$$
3. $\sigma^{-2}|\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{X}, \mathbf{u}, c, d \sim Gamma(a^*, b^*)$ where,
$$a^* = \frac{1}{2}[m + a] \text{ and } b^* = \frac{1}{2}\left[(\mathbf{h}(\boldsymbol{\theta}) - X'\boldsymbol{\beta} - \mathbf{u})'(\mathbf{h}(\boldsymbol{\theta}) - X'\boldsymbol{\beta} - \mathbf{u}) + b\right]$$
4. $\sigma_u^{-2}|\mathbf{u}, c, d, \boldsymbol{\psi} \sim Gamma(c^*, d^*)$ where,
$$c^* = \frac{m+c}{2} \text{ and } d^* = \frac{1}{2}\left(\sum_{i=1}^{m} \psi_i u_i^2 + d\right)$$
5. Two different scenarios can be obtained for the hidden parameters ψ_i, $i = 1, \ldots, m$, depending on the dependence structure initially adopted. Therefore, the full conditionals for this parameter can be expressed by one of the following specifications:
   - Independent random effects:

     5a. if the mixing distribution is $\psi_i \sim Gamma(\nu/2, \nu/2), i = 1, \ldots, m$, then
$$\psi_i|\mathbf{u}, \sigma_u^2, \nu \sim Gamma\left(\frac{1}{2}(\nu + 1), \frac{1}{2\sigma_u^2}\left(u_i^2 + \sigma_u^2\nu\right)\right),$$

5b. if the mixing distribution is $\psi_i \sim Beta(\nu/2, 1), i = 1, \ldots, m$, then

$$\psi_i|\mathbf{u}, \sigma_u^2, \nu \sim Gamma\left(\frac{1}{2}(\nu+1), \frac{1}{2\sigma_u^2}\left(u_i^2\right)\right)\mathbf{1}_{(0,1)}(\psi_i).$$

- Non correlated random effects:

5a.' if the mixing distribution is $\psi \sim Gamma(\nu/2, \nu/2), i = 1, \ldots, m$, then

$$\psi|\mathbf{u}, \sigma_u^2, \nu \sim Gamma\left(\frac{1}{2}(\nu+m), \frac{1}{2\sigma_u^2}\mathbf{u}'\mathbf{u}+\nu\right),$$

5b.' if the mixing distribution is $\psi \sim Beta(\nu/2, 1), i = 1, \ldots, m$, then

$$\psi|\mathbf{u}, \sigma_u^2, \nu \sim Gamma\left(\frac{1}{2}(\nu+m), \frac{1}{2\sigma_u^2}\mathbf{u}'\mathbf{u}\right)\mathbf{1}_{(0,1)}(\psi).$$

6. $\pi(\theta_i|\mathbf{y}, \boldsymbol{\beta}, \mathbf{X}, \sigma^2, \mathbf{u}) \propto h'(\theta_i)\exp\{\phi_i^{-1}(y_i\theta_i + \psi(\theta_i) - \frac{1}{2}(h(\theta_i) - \mathbf{x}_i'\boldsymbol{\beta} - u_i)^2)\}$

3.3.2 Spatially-structured random effects

The main interest is focussed in establishing a robust parametric spatial random effect. Consider the model (1.2.1) with associated link function (3.2.1), stochastic representation (3.2.2) and independent prior distributions given by (3.2.3). Based on these assumptions and recalling the Proposition 3.2.2, the full conditional distributions for this case differs from the last specifications on the distributions of $\mathbf{u}$, σ_u^2 and ψ, which structure is replaced. The next algorithm present the full conditional distributions for this particular model.

Algorithm IV. [Spatially correlated]

1. $\boldsymbol{\beta}|\mathbf{X}, \sigma^2, \mathbf{u} \sim Normal(\hat{\boldsymbol{\beta}}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ where,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{h}(\boldsymbol{\theta}) - \mathbf{u})$$

2. $u_i|\boldsymbol{\theta}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2, \sigma_u^2, \psi, \mathbf{u}_{-i} \sim Normal(\mu_i^u, v_i^u)$ where,

$$\mu_i^u = \left(\frac{(h(\theta_i) - \mathbf{x}_i'\boldsymbol{\beta})}{\sigma^2} + \frac{\psi}{\sigma_u^2}\mu_{i|-i}\right)v_i^u \text{ and } v_i^u = \left(\frac{\psi}{\sigma_{i|-i}^2} + \frac{1}{\sigma^2}\right)^{-1},$$

where, $\boldsymbol{\mu}_{i|-i}$ and $\sigma_{i|-i}^2$ are the location and dispersion parameter given in (2.3.6).

3. $\sigma^{-2}|\theta, \beta, \mathbf{X}, \mathbf{u}, c, d \sim Gamma(a^*, b^*)$ where,

$$a^* = \frac{1}{2}[m+a] \text{ and } b^* = \frac{1}{2}\left[(\mathbf{h}(\theta) - X'\beta - \mathbf{u})'(\mathbf{h}(\theta) - X'\beta - \mathbf{u}) + b\right]$$

4. $\sigma_u^{-2}|\mathbf{u}, \psi, c, d \sim Gamma(c^*, d^*)$ where,

$$c^* = \frac{m+c}{2} \text{ and } d^* = \frac{1}{2}\left(\psi(\mathbf{u}'D_w\mathbf{u}) + d\right).$$

5. In this case, mixing distribution is common $\forall i = 1, \ldots, m$, according to Proposition 3.2.2, thus,

5a.' if $\psi \sim Gamma(\nu/2, \nu/2)$, then

$$\psi|\mathbf{u}, \sigma_u^2, \nu \sim Gamma\left(\frac{1}{2}(\nu+m), \frac{1}{2\sigma_u^2}(\mathbf{u}'D_w\mathbf{u}) + \nu\right)$$

5b.' if $\psi \sim Beta(\nu/2, 1)$, then

$$\psi|\mathbf{u}, \sigma_u^2, \nu \sim Gamma\left(\frac{1}{2}(\nu+m), \frac{1}{2\sigma_u^2}(\mathbf{u}'D_w\mathbf{u})\right)\mathbf{1}_{(0,1)}(\psi)$$

6. $\pi(\theta_i|\mathbf{y}, \beta, \mathbf{X}, \sigma^2, \mathbf{u}) \propto h'(\theta_i)\exp\{\phi_i^{-1}(y_i\theta_i + \psi(\theta_i) - \frac{1}{2}(h(\theta_i) - \mathbf{x}_i'\beta - u_i)^2)\}$.

MCMC scheme maintains its routines as in the subsection 3.3.1. Nandram et al. [70] discussed computational details such as the construction of proposal densities for the Metropolis Hastings sampler, when full posterior distributions for the Gibbs sampler do not exhibit closed forms. Software as WINBUGS [85], through its Geobugs library, has facilitated the computational treatment of a great variety of hierarchical models, including the GLMM spatial models.

3.4 Simulation Study

Simulation study present similar characteristics as in chapter 2. Several additional stages will be considered to complete the spatial model. The steps of this study are reviewed in the next points.

1. Dispersion parameters σ^2 and σ_u^2 will be assumed known. Without lost of generality, the equal geographical influence assumption will be considered, when $\sigma^2 = \sigma_u^2 = 1$.

Real MRF	Fitted MRF	Proportion of spatial variability			
		Mean	Std. Dev.	95% HPD CI	MC Error
Slash(10)	Slash(5)	0.656	0.0328	(0.608,0.746)	0.160
	Slash(10)	0.622	0.0284	(0.541,0.671)	0.126
	Slash(50)	0.476	0.0232	(0.421,0.520)	0.034
	Normal	0.474	0.0214	(0.406,0.509)	0.034
Student-t(10)	Student-t(5)	0.659	0.0257	(0.574,0.703)	0.243
	Student-t(10)	0.623	0.0310	(0.553,0.678)	0.280
	Student-t(50)	0.490	0.0245	(0.454,0.562)	0.026
	Normal	0.475	0.0193	(0.444,0.531)	0.032

Table 3.1: Proportion of spatial variability estimations: Posterior mean, standard deviation, 95% HPD credibility intervals and Monte Carlo error

2. A zero mean Gaussian distribution is considered for the non structured random effect $\boldsymbol{v}$.

3. Three different SMN-RF will be assumed for the spatial random effect $\boldsymbol{u}$, normal, Student-t and slash. For the last two densities, the related degree of freedom is $\nu = 10$. Adjacency matrix associated to the Chilean Metropolitan region is used to perform the simulations.

4. Finally, given $\boldsymbol{\beta} = \mathbf{0}$, $\boldsymbol{u}$ and $\boldsymbol{v}$, simulated data will be generated assuming a Poisson distribution.

The proportion of spatial variability inference is of interest in spatial models. Most fitted models gave similar results for each simulated scenario. Therefore, just simulated results related to the slash and Student-t, both with ten degrees of freedom, fit are presented in table 3.1. From the table, while the degrees of freedom increase, it is possible to appreciate a monotonic decreasing behavior in the posterior mean of the proportion of spatial variability. Similar trends could be observed when a Student-t MRF or a slash MRF is assumed. This downward trend suggests that a SMN RF should be appropriate in order to give more importance to the spatial aggregation effect, instead of obtaining smoothed estimations.

The proposed models are illustrated by using two real data sets, in the context of epidemiological applications.

Chapter 4

Applications

The aim of this work is to apply robust spatial Bayesian model developed in Chapter 3 to detect unusual high relative risks or disease rates in Chilean communes. Using data associated to IDDM incidence rates in Metropolitan region and female lung, trachea and bronchi cancer SMR in the country's northern zone, an exploratory analysis was performed in Chapter 1.

Studies related to incidence rates for diseases like childhood diabetes and cancer have not been studied extensively in the spatial context in Chilean population. Results of robust spatial Bayesian modeling related to both disease are presented in the next sections.

4.1 Insulin dependent diabetes mellitus incidence, Metropolitan Region, Chile

Spatial behavior of IDDM has been studied through Bayesian perspective in diverse populations, such as the Sardinia Island, Italy ([7], [24], [83]), Sweden [81], Norway [52] or Finland [80]. In these countries, IDDM is of a particular importance, due to their incidence rates are higher than in the rest of the world and the trend has been increasing.

Torres et al. [86] show an aggregation of incident rates in space coordinates for urban

	Gaussian MRF	Student-t MRF	Slash MRF
β_0	-9.721 (0.004)	-9.760 (0.006)	-9.752 (0.002)
	(-9.844,-9.634)	(-9.876,-9.631)	(-9.841,-9.656)
σ^2	0.346 (0.013)	0.291 (0.016)	0.275 (0.014)
	(0.162,0.574)	(0.089,0.537)	(0.090,0.507)
σ_u^2	0.230 (0.016)	0.071 (0.001)	0.067 (0.001)
	(0.102,0.547)	(0.035,0.117)	(0.032,0.112)
% Spatial	0.441 (0.011)	0.537 (0.0114)	0.546 (0.012)
Variability	(0.242,0.649)	(0.332,0.749)	(0.338,0.749)
ν	-	10.475 (16.482)	7.346 (6.226)
	-	(3.958,18.277)	(3.038,12.389)

Table 4.1: Posterior mean, standard deviation and 95% HPD credibility intervals for unknown parameters when a Gaussian MRF, Student-t MRF and Slash MRF are assumed.

	DIC		BIC	Predictive (G&G)
Model	Dbar	pD		
Gaussian	846.778		1151.067	13240.408
	534.886	311.892		
Student-t	852.687		1160.315	13335.069
	537.371	315.315		
Slash	836.498		1136.405	13301.901
	529.097	307.401		

Table 4.2: IDDM model selection criteria, DIC, BIC and predictive check.

areas of Metropolitan region, using Bayesian methodology proposed by Molliè ([67], [68]).

Robust Bayesian models proposed in chapter 3 were applied to this problem. Posterior estimations are obtained from a single run of the Gibbs sampler, with a burn-in of 1,000 iterations followed by 10,000 further cycles. Convergence have been checked through trace and autocorrelation plots. Inference over unknown parameters are displayed in table 4.1, when GMRF, Student-t MRF and Slash MRF are assumed to control spatial variability. It is possible to appreciate that similar values are estimated for β_0 and σ^2, under the three MRF

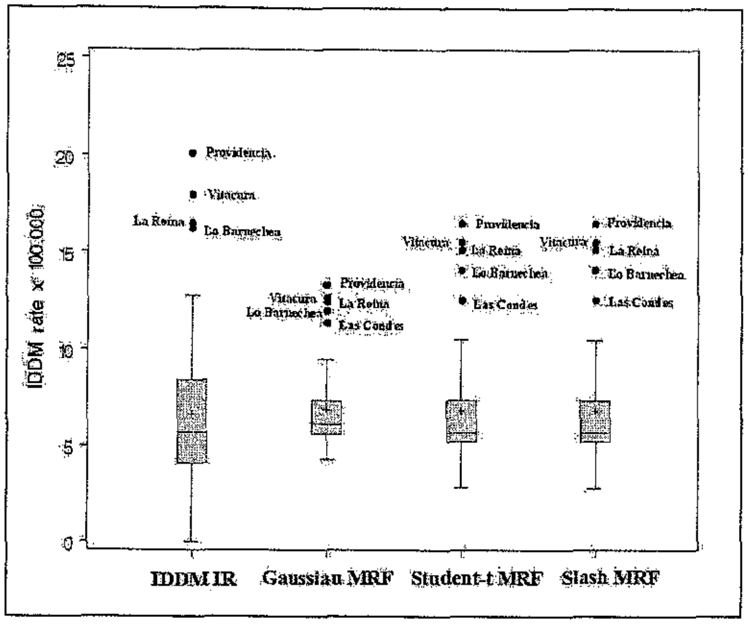

Figure 4.1: IDDM incidence rate (IR) variability: Raw estimates, Mollié's convolution model (Gaussian MRF), Student-t convolution model (Student-t MRF) and Slash convolution model (Slash MRF).

models. In contrast, σ_u^2 present different values, depending on the distribution assumed for the MRF. The robust model (Slash MRF) increase the degree of spatial aggregation from 44.1 % to 54.6 %, that is, the excess of spatial variability presented in this data seems mostly due to clustering effect. As it is mentioned in Banerjee et al. ([6], p. 166), differences could exist in this quantity, when other prior distributions are considered. By the other hand, even when degrees of freedom were not included as an additional stage on modeling step, inference over this parameter is considered. Notice that estimated degrees of freedom are small, which implies that the excess of variability is better captured by one of the SMN RF model.

In figure 4.1 it is possible to appreciate that fully Bayesian estimates of IDDM incidence rates show less variation than raw incidence rate. The three Bayesian variation plots seems to have a similar behavior, due to the presence of several communes with high incidence rates, which are considered as outliers. The main difference, in comparison to the raw IDDM

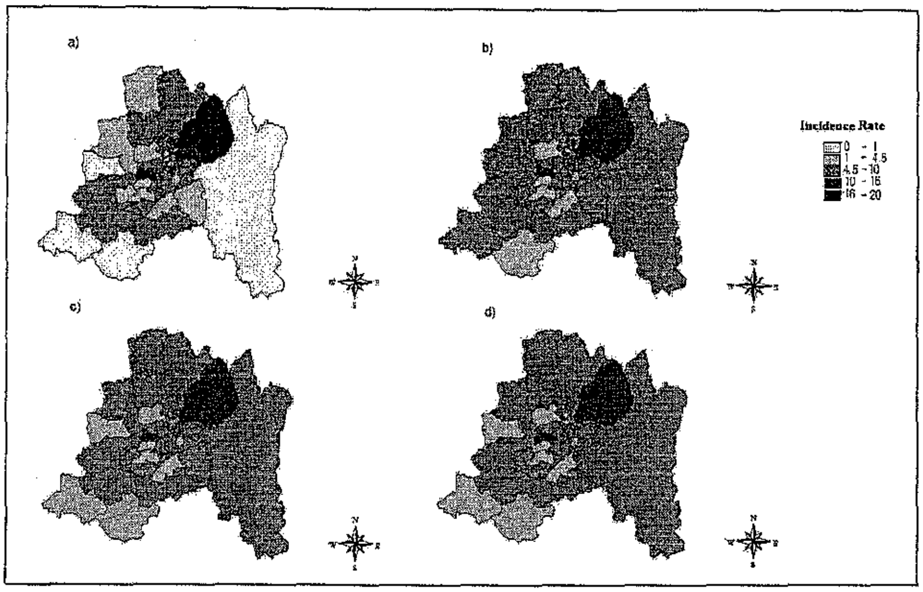

Figure 4.2: IDDM incidence rate: a) Raw incidence rate. b) Mollié's convolution model (Gaussian MRF). c) Student-t convolution model (Student-t MRF). d) Slash convolution model (Slash MRF).

incidence rate, is that commune Las Condes is added in that extreme group by the three models. The Normal MRF assumption lead to estimate smoother rates, however, Student-t and Slash MRF's present slight differences of variability. That difference let to control the excess of smoothness, i.e., robust shrinkage gives a more adequate estimate of the pattern of underlying risk of disease than that provided by the Mollié's convolution estimates.

Model selection criteria results are presented in table 4.2. According to goodness of fit criteria mentioned in chapter 1, small values implies better adjustment. Therefore, spatial model that includes Slash random effects with 7 d.f. is a strong candidate to model geographical dependence.

High incidence estimates remains on communes with high socioeconomic level, such as Vitacura and Providencia, for the four maps. Even more, on those communes this excess is estimated assuming any of the proposed Bayesian models. Spatial distribution of IDDM incidence rates Bayesian estimates can be observed in figure 4.2. Slight differences can be observed between results when Slash MRF (d) and Student-t MRF (c) models are assumed.

	Gaussian MRF	Student-t MRF	Slash MRF
β_0	-0.348 (0.001) (-0.409,-0.300)	-0.372 (0.001) (-0.441,-0.313)	-0.391 (0.001) (-0.425,-0.331)
σ^2	0.092 (0.0003) (0.060,0.129)	0.087 (0.0004) (0.054,0.128)	0.085 (0.0003) (0.055,0.128)
σ_u^2	0.197 (0.001) (0.153,0.238)	0.203 (0.001) (0.150,0.253)	0.203 (0.001) (0.153,0.244)
% Spatial Variability	0.770 (0.001) (0.708,0.841)	0.788 (0.001) (0.740,0.848)	0.788 (0.002) (0.715,0.863)
ν	- -	26.406 (116.944) (15.742,53.499)	32.049 (87.516) (15.585,50.462)

Table 4.3: Posterior mean, standard deviation and 95% HPD credibility intervals for unknown parameters when a Gaussian MRF, Student-t MRF and Slash MRF are assumed.

4.2 Female trachea, bronchi and lung cancer mortality, Chilean northern regions

Several applications applied Bayesian methods to estimate relative risks in small-areas to cancer mortality data can be found in the literature, for example, Ghosh et al. [44], Giuducci et al. [46], Pacutto et al. [72], and Molliè ([68], [67]).

In this work, the application is related to the estimation of female lung, bronchi and trachea cancer mortality relative risks in the northern of Chile. The problem arises when Molliè's model estimates for women cancer mortality risks were too smooth and high on communes where zero cases were observed. Posterior summaries were based on a single sample of the Gibbs sampler, with 10,000 iterations after discarding 1,000-iteration (burn-in). Convergence have been checked via an informal assessment of trace and autocorrelation plots.

SMR and model estimations variability are shown in figure 4.3. Estimation variability is reduced when any of the Bayesian models is considered.

	DIC		BIC	Predictive (G&G)
Model	Dbar	pD		
Gaussian	4821.381		8187.896	381675.00
	3064.272	1757.108		
Student-t	4805.212		8152.110	381671.59
	3058.344	1746.869		
Slash	4792.151		8125.845	381950.00
	3052.174	1739.977		

Table 4.4: Cancer mortality model selection criteria, DIC, BIC and predictive check.

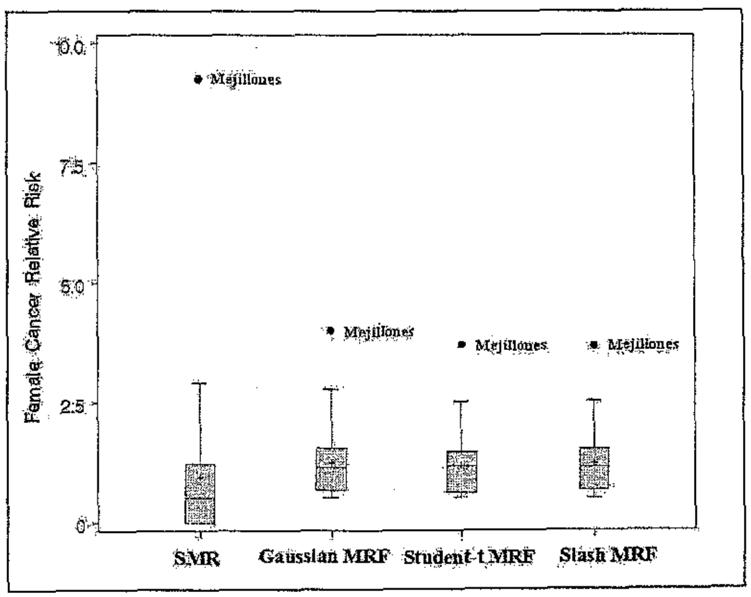

Figure 4.3: SMR Rate variability: Raw rate, Mollié's convolution model (Gaussian MRF), Student-t convolution model (Student-t MRF) and Slash convolution model (Slash MRF).

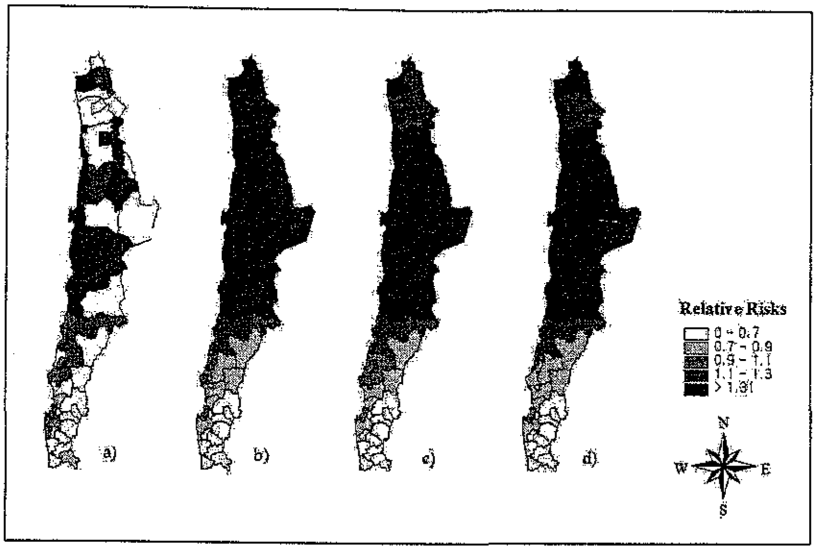

Figure 4.4: Female lung, trachea and bronchi cancer SMR: a) Standardized mortality ratio (SMR). b) Mollié's convolution model. c) Student-t convolution model (Student-t MRF). d) Slash convolution model (Slash MRF).

Figure 4.4 displays the cancer mortality relative risk estimation using three different models, with Mollié's convolution model (b), Student-t MRF (c) and Slash MRF (d) as spatial random effects. Models were tested and the best fit was selected among the three different proposed spatial structures. Table 4.4 selected the Slash spatial random effect with approximately 32 degrees of freedom. The degrees of freedom were estimated assuming a proper non informative prior; see table 4.3 to check the estimates. One important result is referred to the 79% estimated proportion of spatial variability. Notice that this proportion is almost the same for the three proposed models. This could be related to the estimated degrees of freedom.

Although the selected Slash MRF model presented better rates adjustment, notice that the predictive criteria points out to select the model with Student-t MRF spatial random effect. Therefore, from figure 4.4(d) it is possible to appreciate that the first and darkest area in the extreme north, the most populated commune (Arica) in that region, presents the highest rates in comparison with its closer neighbors. It was not possible to reduce the effect produced by the larger areas in the next darkest zones; these ones corresponds to Tarapacá and Antofagasta regions, which are placed in the Atacama Desert. The over-smoothing effect lead to flat true variations in risk, even by the selected model.

Chapter 5

Identifiability issues

5.1 About identifiability and Bayesian learning

Let consider the model (1.2.1), general link function (1.2.4) and prior distributions (2.3.3) and (3.2.3). Taking into account that in the spatial hierarchical model considered in this work, only the sum of two sets of random effects $(\boldsymbol{u} + \boldsymbol{v})$ are identified by the data, a Bayesian identification study of one of the parameters is proposed.

Identifiability has been an important issue to consider when frequentist modeling has been applied. This concept measures the ability of data to estimate parameters of interest. Bunke and Bunke's [20] definition essentially says that two distinct values of an identifiable parametric function should always lead to different likelihoods of the observation. Rothenberg [78] give general conditions when identification is the point of interest, which are based on Fisher's information matrix. An important result for exponential family of distributions is given in this paper, related to the existence of a non singular information matrix for the likelihood in discussion.

Kadane [53] treats identification and defines frecuentist point of view as follows.

Definition 5.1.1. A parameter space Θ is identified if and only if, given two parameters θ, $\theta' \in \Theta$, such that,

$$P_\theta(A) = P_{\theta'}(A), \quad A \in \mathcal{B}(\mathbb{R}),$$

then, $\theta = \theta'$. The converse is also true.

A definition for identification of functions, as a generalization of the notion of identification in definition 5.1.1 is presented. This last extension is useful to characterize identification in sufficient statistics theory. On the other hand, Bayesian results are based on expressions that represent optimal statistical decision when the experiment is observed in two stages, pre and post experiment. Necessary conditions must be satisfied in order to reach Bayesian identifiability, such as, the existence of prior distributions over the parameter space to guarantee positiveness on the difference between prior and posterior expected utility functions.

Bayesian identifiability is also discussed and defined by Dawid [33], who defined it as the ability of posterior distribution to be updated by the available data, as a result of conditional independence. He recalls sufficiency definition, arguing that a sufficient parameter, which is directly related to sufficient statistics, is an identified parameter in the sense that no additional information is necessary, when a sample is observed. Formally, let $\mathbf{y} = (y_1, \ldots, y_n)$ be observed data and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ a vector of parameters, assuming the existence of a prior distribution for $\boldsymbol{\theta}$, $\boldsymbol{\theta}_2$ will be non-identified by observed data $\mathbf{y}$ if

$$\pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \mathbf{y}) = \pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1).$$

Under this intuitive definition, several parameters written in hierarchy can not be identified in the model. Even more, conditions based on conditional and prior independence over a parameter space, characterize Bayesian non-identifiability for some specific models. Non-identifiable parameters are frequently treated as auxiliary parameters in construction of stochastic representations. The latter is just one way to represent uncertainty over parameters of interest instead of making fixed assumptions over them.

Last decade, Poirier [75] and Gelfand and Sahu [41], provided general definitions, results and conditions to preserve identifiability in certain class of models. In some cases, reparameterizations can alleviate this kind of problems. They agree with the fact that identifiability is not an important issue to consider in Bayesian theory, because unbiased estimation is not of main interest; even more, Poirier [75] argues that a Bayesian analysis is always possible when prior distributions present proper or integrability properties. By the other hand and under the same Poirier's ideas, Gelfand and Sahu [41] discussed and establish conditions when generalized linear models are considered to make inferences, due to the evident identifiability problems that this class present.

Ghosh et al. [44] treated an identifiable version of the convolution spatial model [68], presenting conditions and results of integrability when non informative prior distributions are assumed. Now, instead of determining the existence of identifiability, measure the Bayesian learning in the sense of Xie and Carlin [94] is of interest. They exposed two ways of measuring it, instead of justify lack of identifiability. In this context, Kullback Leibler (KL) divergence and precision measurements are proposed in order to determine the cost of introducing data information after a prior belief is specified. One result in the spatial scenario is presented, in the form of the convolution model, assuming normality over the areas.

Two measurements are treated in this chapter, comparing the amount of information lost if a prior distribution for some specific set of parameters (non-identifiable) is considered, exploring results when generalized linear models are assumed and presenting specific results for a spatial discrete model.

5.1.1 Information measures

Initially, two procedures based on Vidal et al. [87] and Xie and Carlin's [94] are considered. Two different alternatives to measure Bayesian learning are specified under assumptions of non normality; therefore, some necessary elements will be exposed in the next lines.

Definition 5.1.2. Bayesian non information are related to the following fact interpretations.

- $\pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \mathbf{y}) = \pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1)$ when $\mathbf{y}$ is conditionally uninformative for $\boldsymbol{\theta}_2$, given $\boldsymbol{\theta}_1$, and,

- $\pi(\boldsymbol{\theta}_2|\mathbf{y}) = \pi(\boldsymbol{\theta}_2)$ when $\mathbf{y}$ is marginally uninformative for $\boldsymbol{\theta}_2$.

Let re-define the following expressions as,

$$\begin{array}{ll} \text{Prior distribution} & p_0 = \pi(\boldsymbol{\theta}_2), \\ \text{Posterior distribution} & p_1 = \pi(\boldsymbol{\theta}_2|\mathbf{y}), \\ \text{Full Posterior distribution} & p_2 = \pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1). \end{array} \tag{5.1.1}$$

Two usual information measures to make the analysis are L_1 distance ([73], [90]) and KL divergence [58]. These quantities will allow the quantification of gained knowledge, when data $\mathbf{y}$ was observed. So, Bayesian learning process will depend on the results when full conditional, posterior and prior distributions for the "non-identifiable parameter" are compared. Both discrepancy measures will depend on the structure of distributions earlier mentioned.

In this work the study will be applied to the spatial model discussed in chapter three. Given data $\mathbf{y}$, associated likelihood function and prior distributions are represented by $L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \mathbf{y})$ and $\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ respectively. Under the latter specification, it is possible to obtain complete posterior for $\boldsymbol{\theta}_2$, $\pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \mathbf{y})$, and its posterior distribution $\pi(\boldsymbol{\theta}_2|\mathbf{y})$.

Two variations in the model (1.2.1), link function (3.2.1), prior distributions (3.2.3) and SMN RF (3.2.2) must be taken into account,

1. Link function (3.2.1) re-parametrization, has the form,

$$\begin{aligned} &\boldsymbol{\eta} = \boldsymbol{u} + \boldsymbol{v}, \text{ where } \boldsymbol{\eta}|\boldsymbol{v}, \psi, \sigma_u^2 \overset{ind}{\sim} N\left(\boldsymbol{v}, \frac{\sigma_u^2}{\psi} D_w^{-1}\right) \\ &\boldsymbol{v}|\sigma^2, \boldsymbol{\beta} \overset{ind}{\sim} N(\mathbf{x}'\boldsymbol{\beta}, \sigma^2 \mathbf{I}). \end{aligned} \tag{5.1.2}$$

   Therefore, in this work $\boldsymbol{v}$ is the set of non identifiable parameters.

2. Without loss of generality, fixed dispersion parameters σ^2, σ_u^2 and structural parameter vector $\boldsymbol{\beta} = \mathbf{0}$ are assumed.

3. A positive P_ψ c.d.f. is assumed for ψ.

After some algebraic manipulations and under re-parameterized spatial model recently described it is possible to obtain the following proposition.

Proposition 5.1.1. *If model (1.2.1), hierarchical parametrization (5.1.2) and independent prior distributions given by (3.2.3) are assumed, then* $p_0 = \pi(\boldsymbol{v})$, $p_1 = \pi(\boldsymbol{v}|\boldsymbol{y})$ *and* $p_2 = \pi(\boldsymbol{v}|\boldsymbol{\eta})$ *have the following forms,*

$$\begin{aligned} p_0 &= N(\mathbf{0}, \sigma^2\mathbf{I}) & (5.1.3) \\ p_1 &= \int_{Rec\boldsymbol{\eta}} \exp\{(\boldsymbol{\phi}\otimes\boldsymbol{y})'\boldsymbol{\eta} - \boldsymbol{\phi}' g(\boldsymbol{\eta})\} h'(\boldsymbol{\eta}) \int_0^\infty N_{\boldsymbol{\eta}}\left(\boldsymbol{v}, \frac{\sigma_u^2}{\psi} D_w^{-1}\right) dP_\psi d\boldsymbol{\eta} \\ &\quad \times N_{\boldsymbol{v}}(\mathbf{0}, \sigma^2\mathbf{I}), & (5.1.4) \\ p_2 &= \int_0^\infty N(\boldsymbol{\mu}_v, \Sigma_v) dP_\psi = SMN(\boldsymbol{\mu}_v, \Sigma_v, \nu), & (5.1.5) \end{aligned}$$

with

$$\boldsymbol{\mu}_v = \boldsymbol{\eta}'\left[\tfrac{\psi}{\sigma_u^2} D_w\right]\Sigma_v \qquad and \qquad \Sigma_v = \left[\tfrac{\psi}{\sigma_u^2} D_w + \tfrac{1}{\sigma^2}\mathbf{I}\right]^{-1},$$

where, $\boldsymbol{\phi} = (\phi_1^{-1}, \ldots, \phi_n^{-1})$ *are known dispersion parameters,* $g(\boldsymbol{\eta}) = (g(\eta_1), \ldots, g(\eta_n))$ *is a set of functions for the linear link and the operator* $\boldsymbol{a}\otimes\boldsymbol{b}$ *indicates the elementwise product between* $m \times 1$ *vectors* $\boldsymbol{a}$ *and* $\boldsymbol{b}$.

These distributions will be useful to obtain the required learning measurements. Mathematical expressions (5.1.4) and (5.1.5) can be approximated using Monte Carlo integration techniques.

Remark 5.1.1. Under definition 5.1.2, it is possible to show that random vector $\boldsymbol{y}$ is conditionally uninformative for $\boldsymbol{v}$ given $\boldsymbol{\eta}$; however, according to expression (5.1.4), $\boldsymbol{y}$ is marginally informative for $\boldsymbol{v}$.

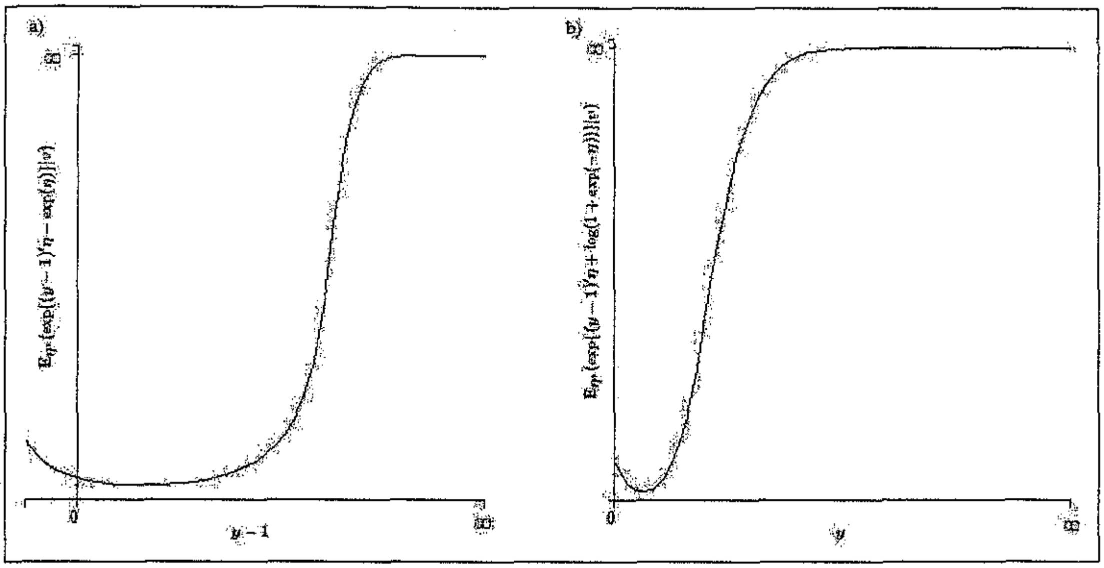

Figure 5.1: Expectation (5.1.4) vs $\boldsymbol{y}$ for (a) Poisson model and (b) Bernoulli model.

Remark 5.1.2. One more condition must satisfy (5.1.4) related to the existence of $\mathbb{E}((\boldsymbol{\phi} \otimes \boldsymbol{y})'\boldsymbol{\eta} - \boldsymbol{\phi}' g(\boldsymbol{\eta})\} h'(\boldsymbol{\eta}))$. In the context of exponential family, there is no limitation with this fact.

Two special cases are of special interest. The following examples show some reduced forms for (5.1.4).

Example 5.1 Consider a random vector $\boldsymbol{y}$ following a Poisson distribution with $\mathbb{E}(y_i) = \exp(\eta_i)$. Then, it is known from exponential family theory that $g(\eta_i) = \exp(\eta_i)$ and $\phi_i = 1$, $i = 1, \ldots, n$, then (5.1.4) can be re written as,

$$\pi(\boldsymbol{v}|\boldsymbol{y}) \propto \mathbb{E}_{\boldsymbol{\eta}^*}(\exp\{(\boldsymbol{y} - \mathbf{1})'\boldsymbol{\eta} - \exp(\boldsymbol{\eta})\}|\boldsymbol{v}) \times N(\mathbf{0}, \sigma^2 \mathbf{I})$$

where,

$$\boldsymbol{\eta}^* \sim SMN\left(\boldsymbol{v}, \frac{\sigma_u^2}{\psi} D_w^{-1}\right)$$

Example 5.2 Consider a random vector y following a Bernoulli distribution with $\mathbb{E}(y_i) = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}$. In this case, $g(\eta_i) = \log(1 + \exp(\eta_i))$ and $\phi_i = 1$, $\forall i = 1, \ldots, n$; under this conditions, (5.1.4) can be expressed by,

$$\pi(\boldsymbol{v}|\boldsymbol{y}) \propto \mathbb{E}_{\boldsymbol{\eta}^*}(\exp\{(\boldsymbol{y}-\mathbf{1})'\boldsymbol{\eta} + \log(1+\exp(-\boldsymbol{\eta}))\}|\boldsymbol{v}) \times N_{\boldsymbol{v}}(\mathbf{0}, \sigma^2\mathbf{I})$$

with $\boldsymbol{\eta}^*$ distributed as in example 4.1.

Figure 5.1 shows the expectation behavior for probability functions (5.1.4), in the Poisson (example 5.1) and Bernoulli (example 5.2) case, when $\pi(\boldsymbol{\eta}|\boldsymbol{v})$ is equivalent to a standard normal distribution. The importance of these plots is related to their positive behavior, as part of a distribution function. Even more, these expressions could include the normalization constant.

On the other hand, notice that this class of models keep the initial SMN distributions properties. A particular case is exposed in the next corollary.

Corollary 5.1.2. *If $\mathbb{P}(\psi = 1) = 1$, then probabilistic expressions for p_1 in (5.1.4) and p_2 in (5.1.5) are reduced to,*

$$\begin{aligned} p_1 &= \int_{Rec\boldsymbol{\eta}} \exp\{(\boldsymbol{\phi}\otimes \boldsymbol{y})'\boldsymbol{\eta} - \boldsymbol{\phi}' g(\boldsymbol{\eta})\} h'(\boldsymbol{\eta}) N_{\boldsymbol{\eta}}\left(\boldsymbol{v}, \sigma_u^2 D_w^{-1}\right) d\boldsymbol{\eta} \times N_{\boldsymbol{v}}(\mathbf{0}, \sigma^2\mathbf{I}), \\ p_2 &= N(\boldsymbol{\mu}_v, \Sigma_v), \end{aligned}$$

with

$$\boldsymbol{\mu}_v = \boldsymbol{\eta}'\left[\frac{1}{\sigma_u^2} D_w\right]\Sigma_v \qquad \textit{and} \qquad \Sigma_v = \left[\frac{1}{\sigma_u^2} D_w + \frac{1}{\sigma^2}\mathbf{I}\right]^{-1},$$

where, $\boldsymbol{\mu}_v$ and Σ_v remains as the previous location and dispersion parameters in Proposition 5.1.1.

5.1.2 L_1 distance measure

Vidal et al. [87] analyze what they called local sensitivity, when a specific parameter for skewness is considered or not. They compare this discrepancy through L_1 distance ([73], [90]) between two proposed models, and measures the maximum discrepancy between them. This q-divergence measure is considered in Arellano-Valle et al. [4] in order to quantify data influence in posterior distributions, when elliptical regression models are considered. L_1 distance between two models f and g can be defined by,

$$L_1(f,g) = \frac{1}{2}\int |f(x) - g(x)|dx = \sup_{A \in \mathbf{B}} |\mathbb{P}(A|f) - \mathbb{P}(A|g)|,$$

where $\mathbf{B}$ represents Borel's sets and $\mathbb{P}(A|f)$ denotes the probability measure defined by the density f. Therefore, distance L_1 let measure the maximum discrepancy between specific models f and g. More detailed interpretations can be found in Vidal et al. [87] and references therein.

Taking this idea, learning differences can be obtained using differences between distributions in proposition 5.1.1, in order to measure sensitive elements to decide how much can be learnt about the set of unidentifiable parameters, once $\mathbf{y}$ has been observed.

Proposition 5.1.3. *For any fixed σ^2 and σ_u^2, L_1 distances between distributions (5.1.3), (5.1.4) and (5.1.5) specified in Proposition 5.1.1 are,*

$$L_1(p_2, p_0) = \mathcal{C}\left(sign(\eta_i)\left[\frac{1}{2} - \int_{[-\infty,0]^n} \pi(v_i|\eta_i)dv_i\right]; i = 1,\dots,m\right) \tag{5.1.6}$$

$$L_1(p_2, p_1) = \frac{1}{2}\int_{[-\infty,\infty]^n} N_{\boldsymbol{v}}(\mathbf{0},\sigma^2)|\pi(\boldsymbol{\eta}|\boldsymbol{v}) - \mathrm{K}\mathrm{E}_{\boldsymbol{\eta}}[\exp\{(\boldsymbol{\phi}\otimes \boldsymbol{y})'\boldsymbol{\eta} - \boldsymbol{\phi}' g(\boldsymbol{\eta})\}h'(\boldsymbol{\eta})|\boldsymbol{v}]|d\boldsymbol{v}, \tag{5.1.7}$$

where, $\boldsymbol{\eta}|\boldsymbol{v} \sim \pi(\boldsymbol{\eta}|\boldsymbol{v}) \equiv SMN(\boldsymbol{v}, \sigma_u^2 D_w^{-1}, \nu)$, $K = f(\boldsymbol{y})^{-1}$ and $\mathcal{C}$ is an appropriate function that captures the joint behavior from each marginal element of $L_1(p_2, p_0)$.

Proof. Proof of proposition 5.1.1 is oriented to show the reduced expression obtained for (5.1.6). The symmetric behavior of normal and scale of normal distributions is relevant

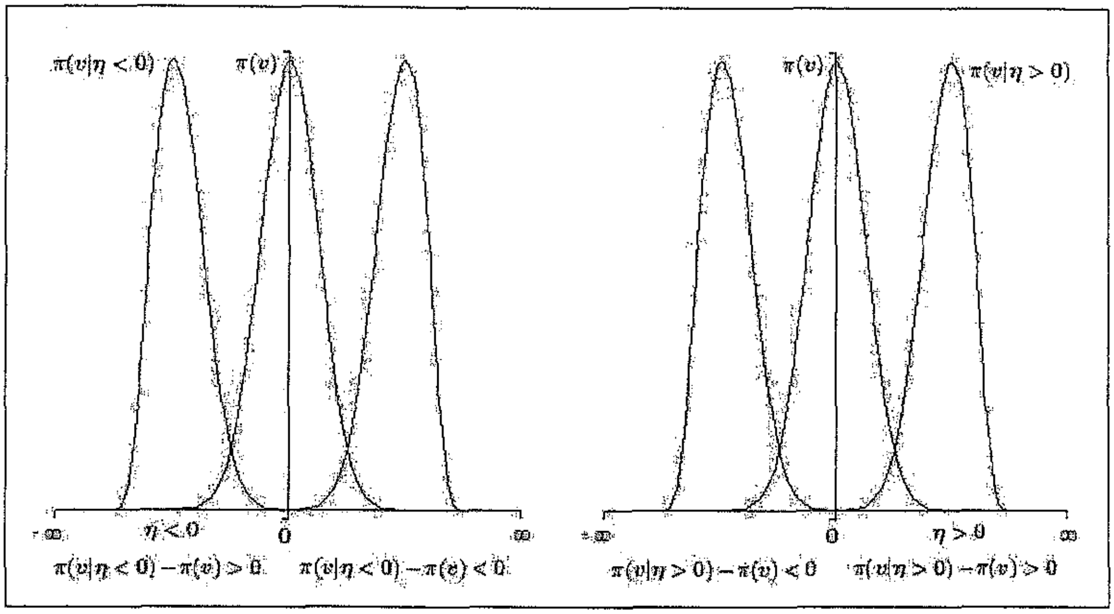

Figure 5.2: Regions for L_1 distances under symmetry.

in the demonstration process. Prior distribution for $\boldsymbol{v}$ is assumed to be zero centered. Multivariate case considers a construction from the univariate cases. Demonstration is reflected for the univariate case.

$$\int |\pi(v_i|\eta_i) - \pi(v_i)| = \int_{-\infty}^{0} [\pi(v_i|\eta_i) - \pi(v_i)]dv_i + \int_{0}^{\infty} [\pi(v_i|\eta_i) - \pi(v_i)]dv_i$$

Assume that $\eta_i < 0$.

If $v_i < 0$ then, $\pi(v_i|\eta_i < 0) - \pi(v_i) > 0$. Conversely, if $v_i > 0$ then, $\pi(v_i|\eta_i < 0) - \pi(v_i) <$

0,

$$\begin{aligned}
\Rightarrow \quad & \int_{-\infty}^{0} |\pi(v_i|\eta_i < 0) - \pi(v_i)|dv_i + \int_{0}^{\infty} |\pi(v_i|\eta_i < 0) - \pi(v_i)|dv_i \\
= \quad & \int_{-\infty}^{0} \pi(v_i|\eta_i < 0)dv_i - \frac{1}{2} - \int_{0}^{\infty} \pi(v_i|\eta_i < 0)dv_i + \frac{1}{2} \\
= \quad & \int_{-\infty}^{0} \pi(v_i|\eta_i < 0)dv - 1 + \int_{-\infty}^{0} \pi(v_i|\eta_i < 0)dv \\
= \quad & 2\int_{-\infty}^{0} \pi(v_i|\eta_i < 0)dv - 1.
\end{aligned}$$

Similarly, if $\eta_i > 0$,

$$\begin{aligned}
\Rightarrow \quad & -\int_{-\infty}^{0} \pi(v_i|\eta_i > 0)dv_i + 1 - \int_{-\infty}^{0} \pi(v_i|\eta_i > 0)dv_i \\
= \quad & 1 - 2\int_{-\infty}^{0} \pi(v_i|\eta_i > 0)dv_i,
\end{aligned}$$

then $L_1(p_0, p_2)$ depend on the collection generated by the signs of $\boldsymbol{\eta}$. Figure 5.2 shows individual regions for L_1 distances when $\eta_i < 0$ and $\eta_i > 0$. Approximations using Monte Carlo methods can be used. Other alternatives are related to obtain this quantity using functions C, that approximate the joint behavior and keep the dependence between its components.

With respect to (5.1.7), its computation is relegated to use estimations obtained from MCMC sampling methods. □

A reduced mathematical expression for L_1 distance between full posterior distribution (5.1.5) and prior distribution (5.1.3) is obtained, due to the symmetry properties that SMN and Normal distributions present. The latter present a simple structure, which just depends on the full posterior distribution $\pi(\boldsymbol{v}|\boldsymbol{\eta})$.

Computation for expression (5.1.7) relies on simulation methods because involved probability distribution (5.1.4) depends on complex functions of unknown parameters $(\boldsymbol{\eta}, \psi)$

and cannot be computed exactly.

Both expressions (5.1.6) and (5.1.7), represent learning discrepancies. Expression (5.1.6) can be interpreted as the prior (potential) learning difference and expression (5.1.7) is related to the posterior (remaining) learning difference, once y was observed. These interpretations are appropriate, if Xie and Carlin's work is followed.

5.1.3 Kullback Leibler divergence

A second alternative to measure discrepancy between two models is proposed in this section, following divergence measures considered by Arellano-Valle et al. [4], who applied them for sensitivity detection purposes in elliptical Bayesian regression models; they proposed this class of measures as part of q-divergences measures family to quantify the effect produced by a subset of data, in order to examine its influence in posterior distributions.

Xie and Carlin [94] compare Bayesian learning among prior and posterior distributions, when non-normality is addressed. They suggested that KL divergence is adequate to quantify the learning process. Computations and applications are oriented to illustrate this methodology when methods for density estimation and spatial models including multiple random effects are considered.

In this work, non symmetrized KL divergence between two probability models f and g will be defined by

$$KL(f,g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx.$$

Following Xie and Carlin [94], let define the following expressions.

- Potential Learning Divergence

$$KL(p_0, p_2) = \int_{[-\infty,\infty]^n} \pi(v|\eta) \log \frac{\pi(v|\eta)}{\pi(v)} dx. \tag{5.1.8}$$

- Remaining Learning Divergence

$$KL(p_1, p_2) = \int_{[-\infty,\infty]^n} \pi(\boldsymbol{v}|\boldsymbol{\eta}) \log \frac{\pi(\boldsymbol{v}|\boldsymbol{\eta})}{\pi(\boldsymbol{v}|\boldsymbol{y})} dx. \tag{5.1.9}$$

These expressions do not present analytical closed forms, therefore, MCMC approaches can be adopted to obtain estimations for 5.1.8 and 5.1.9. Explicit expressions are omitted, due to closed forms are unavailable and difficult to evaluate.

5.2 Markov chain Monte Carlo approach

Monte Carlo algorithms proposed by Xie and Carlin [94], can be adapted to estimate expressions difficult to evaluate, such as 5.1.7, 5.1.8 and 5.1.9. Following this work, it is possible to obtain empirical Bayes estimations for a given model, even within a simulation study. Under this idea, MCMC scheme can be reviewed into the following steps.

i. Obtain marginal samples from $\boldsymbol{v}$ and $\boldsymbol{\eta}$ given $\boldsymbol{y}$ posterior distributions, running an appropriate MCMC sampler. In particular, for this work, a hybrid Gibbs-Metropolis algorithm can be applied.

ii. Once $\boldsymbol{\eta}$ samples are available, run the MCMC algorithm replacing $\boldsymbol{y}$ with $\boldsymbol{\eta}$ to obtain samples from the posterior distribution $\pi(\boldsymbol{v}|\boldsymbol{\eta})$.

iii. Estimate 5.1.7, 5.1.8 and 5.1.9, using available samples and Monte Carlo integration techniques.

In this thesis work, the computational stage is proposed to be further developed.

Chapter 6

Concluding remarks and discussion

Specific results were obtained through this work, which are relevant for Bayesian small area estimation proposing a new methodology as an alternative to usual parametric models. This approach is particularly useful to obtain estimation of rates or relative risks when subjective geographical dependence is assumed and related results are too smooth for the region under study.

In this work, an extension to obtain robust inference from GMRF theory is proposed. As a first step, the extension is applied to classical normality assumptions, considering SMN family of distributions to capture regional spatial behavior. Conditions are required to ensure the propriety of these intrinsic spatial random effect posterior distributions, which must be associated to sum zero constraint and existence of mixing random variable expectations. When spatial correlation structure is available, one Proposition led to provide sufficient conditions to guarantee posterior distribution integrability for Bayesian GLMM. Another Proposition is presented, when a non spatial correlation scenario is assumed.

The general methodology is applicable to situations where small area parameters must be estimated. Variability parameters are of interest, their incorporation in the proposed hierarchical models allowed the computation of the marginal spatial proportion of variability, through the empirical marginal standard deviation function, to quantify excess of variability explained by the spatial effect. This fact keeps direct relation with the spatial random effect contribution considered for the analysis.

Applications study gave better results, considering the complex structure of Chilean geography. Both applications were best modeled by Poisson regression with spatial random effects following a joint Slash distribution. It is possible to notice that β_0 do not produce changes when the three models are fitted to both applications. That is an important consideration that shows the robust properties of the Student-t MRF and Slash MRF. Specifically, for IDDM incidence estimation, relevant results are,

- Smooth effect on several distant rural communes was better adjusted considering that the observed number of cases are zero with associated small population at risk.
- The Slash MRF model increase the degree of spatial aggregation. This is an important result from epidemiologist researchers, since diabetes spatial hypothesis is not clearly understood.

In the case of female cancer mortality relative risks estimation, some important results are:

- The smoothing effect of Arica and Parinacota region matches with a better estimation of risks in the area.
- Highly geographical influence that cancer mortality present has been studied in most references. For this application, high spatial excess of variability was not the exception, allowing for a 78.8%.

In the last chapter, Bayesian learning analysis is considered in order to show that unknown parameters are updated by the observed data. Simulation study is proposed in order to investigate model behavior under heavier tailed distributions for spatial random effects. General expressions to measure Bayesian learning were obtained. These quantities are based on L_1 distance and Kullback Leibler divergence, and allow interpretations facts such as potential and remaining learning, defined in a previous work by Xie and Carlin [94]. These measures were specifically proposed to measure the discrepancy between prior and posterior distributions that take part on the Bayesian identifiability definitions, when a spatial random effect that follows a SMN RF is assumed. A formal simulation study is proposed to be further developed.

This thesis proposed an extension and the Bayesian methodology to fit robust Bayesian inference models, extending GMRF to a heavier tailed family distributions such as SMN RF. Bayesian model identifiability was performed, to obtain results related to measure Bayesian learning on a set of specific parameters. The results were computationally implemented using simulated data.

As future works, several topics can be explored in the spatial context. Diagnostic approaches, model sensibility and extensions of model assumptions which include asymmetry in the distribution of the random effects, are related topics to be developed. Simulation studies to validate proposed models under different scenarios can be performed too.

Bayesian space time models can be proposed, with the subsequent problem of sparseness of data that could affect estimation on communes with low population. Therefore, robust models will become more necessary. Temporal trends and geographical patterns are estimated simultaneously, allowing for additional random effects to represent temporal and spatio temporal interaction variations.

Multivariate extensions related to simultaneous modeling of epidemiological events, using multivariate generalized linear regression models. Develop the related Bayesian theory of this class of models, which include estimation and decision problems under informative and non informative prior elicitation. Model comparison with other different developed methodologies actually proposed in the literature.

Bibliography

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle. *Proceedings of the 2nd. International Symposium of Information Theory*, pages 267–281, 1973.

[2] H. Akaike. A new look at the statistical model identification. *IEEE Transantions on Automatic Control*, AC19:716–723, 1974.

[3] D.F. Andrews and C.L. Mallows. Scale mixture of normal distributions. *Journal of the Royal Statistical Society, Series B*, 36(1):99–102, 1974.

[4] R. Arellano-Valle, M. Galea, and P. Iglesias. Bayesian sensitivity analysis in elliptical linear regression models. *Journal of Statistical Planning and Inference*, 86:175–199, 2000.

[5] R.M. Assunção, J.E. Potter, and S.M. Cavenaghi. A Bayesian space varying parameter model applied to estimating fertility schedules. *Statistics in Medicine*, 21:2057–2075, 2002.

[6] S. Banerjee, B. Carlin, and A. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Monographs on Statistics and Applied Probability 101. Chapman and Hall, Boca Ratón, Florida, 2004.

[7] L. Bernardinelli, D. Clayton, C. Pascutto, C.L. Montomoli, M. Ghislandi, and M. Songini. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, 14(21-22):2433–2443, 1995.

[8] J.M. Bernardo and A. Smith. *Bayesian Theory*. Wiley Sons, New York, 2000.

[9] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36(2):192–236, 1974.

[10] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48(3):259–302, 1986.

[11] J. Besag. Towards Bayesian image analysis. *Journal of Applied Statistics*, 16:395–407, 1989.

[12] J. Besag and D. Hidgon. Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society, Series B*, 61(4):691–746, 1999.

[13] J. Besag and C. Kooperberg. On conditional and intrinsic autoregression. *Biometrika*, 82(4):733–746, 1995.

[14] J. Besag and J. Newell. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society, Series A*, 154:143–155, 1991.

[15] J. Besag, J. York, and A. Molliè. Bayesian image restoration with applications in spatial statistics (with discussion). *Annals of the Institute of Mathematical Statistics*, 43:1–59, 1991.

[16] N. Best, R. Arnold, A. Thomas, L. Waller, and E. Collon. Bayesian models for spatially correlated disease and exposure data. In Bernardo J., Smith A., Dawid A., and Berger J., editors, *Bayesian Statistics 6*, pages 131–156, Oxford, 1999. Oxford University Press.

[17] N. Best, S. Richardson, and A. Thomson. A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14:35–59, 2005.

[18] N. Breslow and D. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25, 1993.

[19] N. Breslow and N. Day. *Statistical Methods in Cancer Research, Volume II - The Design and Analysis of Cohort Studies*. International Agency for Research on Cancer, Lyon, 1987.

[20] H. Bunke and O. Bunke. Identifiability and estimability. *Mathematische Operationsforschung und Statistik, Berlin*, 5:222–233, 1974.

[21] R. Cangnon and M. Clayton. Bayesian detection and modeling of spatial disease clustering. *Biometrics*, 56(3):922–935, 2000.

[22] R. Cangnon and M. Clayton. A hierarchical model for spatially clustered disease rates. *Statistics in Medicine*, 22:3213–3228, 2003.

[23] G. Casella and E. George. Explaining the Gibbs Sampler. *The American Statistician*, 46(3):167–174, 1992.

[24] A. Casu, C. Pascutto, V.M. Cambuli, L. Bernardinelli, M. Songini, and The Sardinian IDDM Epidemiology Study Group. Type 1 diabetes incidence under 15 years is still increasing in Sardegna: data from the Eurodiab register 1989-1999. *Diabetologia*, 45(S2):A68, 2002.

[25] M.H. Chen, Q.M. Shao, and D. Xu. Necessary and sufficient conditions on the properiety of posterior distributions for Generalized Linear Mixed Models. *Sankhya, Series A*, 64(1):57–85, 2002.

[26] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[27] D. Clayton. Generalized linear mixed models. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 359–379, New York, 1996. Chapman and Hall.

[28] D.G. Clayton and J. Kaldor. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43:671–681, 1987.

[29] P. Congdon. *Applied Bayesian Modelling.* Wiley Sons, Chichester, 2003.

[30] N.A.C. Cressie. *Statistics for Spatial Data.* Wiley - Interscience, New Jersey, 1993.

[31] P. Damien and S. Walker. Sampling truncated normal, beta and gamma densities. *Journal of Computational and Graphical Statistics*, 10(2):206–215, 2001.

[32] G.S. Datta and P. Lahiri. Robust hierarchical Bayes estimation of small area characteristics in the presence of covariates and outliers. *Journal of Multivariate Analysis*, 54:310–328, 1995.

[33] A.P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B*, 41:1–31, 1979.

[34] N. Díaz. Atlas de mortalidad por enfermedades isquémicas del corazón para la Región Metropolitana. Chile 1997-2002. Master's thesis, Pontificia Universidad Católica de Chile, 2005.

[35] P.J. Diggle. *Statistical Analysis of Spatial Point Patterns, 2nd. Ed.* Arnold, London, 2003.

[36] K.T. Fang and T.W. Anderson. *Statistical Inference in Elliptically Contoured and Related Distributions.* Allerton Press Inc., New York, 1990.

[37] K.T. Fang, S. Kotz, and K.W. NG. *Symmetric Multivariate and Related Distributions.* Chapman and Hall, New York, 1990.

[38] C. Fernandez and M. Steel. Multivariate student-t regression models: Pitfalls and inference. *Biometrika*, 86:153–167, 1999.

[39] C. Fernández and M. Steel. Bayesian regression analysis with scale mixtures of normals. *Econometric Theory*, 16(1):80–101, 2000.

[40] A.E. Gelfand and S.K. Ghosh. Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85:1–11, 1998.

[41] A.E. Gelfand and S.K. Sahu. Identifiability, improper priors and gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, 94:247–253, 1999.

[42] A.E. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[43] J. Geweke. Bayesian treatment of the independent Student- t linear model. *Journal of Applied Econometrics*, 8:519–540, 1993.

[44] M. Ghosh, K. Natarajan, T.W.F. Stroud, and B.P. Carlin. Generalized linear models for small-area estimation. *Journal of the American Statistical Association*, 93(441):273–282, 1998.

[45] W.R. Gilks and P. Wild. Adaptive rejection sampling. *Applied Statistics*, 41:337–348, 1990.

[46] P. Giudicci, L. Knorr-Held, and G. Rasser. Modelling categorical covariates in bayesian disease mapping by partition structures. *Statistics in Medicine*, 19:2579–2593, 2000.

[47] E. Gómez-Sánchez-Manzano, M.A. Gómez-Villegas, and Marín J.M. Sequences of elliptical distributions and mixtures of normal distributions. *Journal of Multivariate Analysis*, 97:295–310, 2006.

[48] D.A. Griffith and R. Haining. Beyond mule kicks: The Poisson distribution in geographical analysis. *Geographical Analysis*, 38:123–139, 2006.

[49] A.K. Gupta and T. Varga. Normal mixture representations of matrix variate elliptically contoured distributions. *Sankhya, Series A*, 57(1):68–78, 1995.

[50] G. Icaza and L. Núñez. *Atlas de mortalidad por enfermedades cardiovasculares en Chile, 1997- 2003.* Universidad de Talca y Ministerio de Salud, Talca, Chile, 2006.

[51] G. Icaza, L. Núñez, F. Torres, N. Díaz, and D. Varela. Distribución geográfica de mortalidad por tumores malignos de estómago, tráquea, bronquios y pulmón, Chile 1997-2004. *Revista Médica de Chile*, 135(11):1397–1405, 2007.

[52] G. Joner, O. Sovik, and T. Riise. Clustering of type one diabetes mellitus in Norway. *Diabetologia*, 41(S1):A21 – 75, 1998.

[53] J.B. Kadane. The role of identification in Bayesian theory. In *Studies in Bayesian Econometrics and Statistics*, pages 175–191, 1974.

[54] M. Kaiser and N. Cressie. The construction of multivariate distributions from markov random fields. *Journal of Multivariate Analysis*, 73:199–220, 2000.

[55] Y. Kano. Consistency property of elliptical probability density functions. *Journal of Multivariate Analysis*, 51:139–147, 1994.

[56] L. Knorr-Held and G. Raßer. Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, 56(1):13–21, 2000.

[57] A. Kottas, J. Duan, and A. Gelfand. Modeling disease incidence data with spatial and spatio-temporal dirichlet process mixtures. *Biometrical Journal*, 50(1):29–42, 2008.

[58] S. Kullback. *Information Theory and Statistics.* Wiley Sons, New York, 1959.

[59] K. Lange, R.J.A. Little, and J.M.G. Taylor. Robust statistical modeling using the t-distribution. *Journal of the American Statistical Association*, 84:881–896, 1989.

[60] K. Lange and J.S. Sinsheimer. Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, 2(2):175–198, 1993.

[61] P.W. Laud and J.G. Ibrahim. Predictive model selection. *Journal of the Royal Statistical Society, Series B*, (57):247–262, 1995.

[62] A. Lawson. *Statistical Methods in Spatial Epidemiology.* Wiley Sons, London, 2006.

[63] J. Lee, M. Kaiser, and N. Cressie. Multiway dependence in exponential family conditional distributions. *Journal of Multivariate Analysis*, 79:171–190, 2001.

[64] S. Lyu and E.P. Simoncelli. Statistical modeling of images with fields of Gaussian scale mixtures. In Schölkopf B., Platt J., and Hoffman T., editors, *Advances in Neural Information Processing Systems, 19*, pages 945–952, Cambridge, 2007. MIT Press.

[65] Y. MacNab, P. Farrell, P. Gustafson, and S. Wen. Estimation in Bayesian disease mapping. *Biometrics*, 60(4):865–873, 2004.

[66] P. McCullagh and J. Nelder. *Generalized Linear Models, 2nd. Edition.* Chapman & Hall/CRC, 1989.

[67] A. Molliè. Bayesian mapping of Hodgkin's disease in France. In P. Elliott, J. Wakefield, N.G. Best, and D.J. Briggs, editors, *Spatial Epidemiology: Methods and Applications*, pages 267–285, New York, 1957. Oxford University Press.

[68] A. Molliè. Bayesian mapping of disease. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 359–379, New York, 1996. Chapman and Hall.

[69] B. Nandram. Bayesian generalized linear models for inference about small areas. In D.K. Dey, S.K. Ghosh, and B.K. Mallick, editors, *Generalized Linear Models: A Bayesian Perspective*, pages 91–114, New York, 2000. Marcel Dekker.

[70] B. Nandram, J. Sedransk, and L.W. Pickle. Bayesian analysis of mortality rates for U.S. health service areas. *Sankhya, series B*, 61(1):145–165, 1999.

[71] M. Orellana. Análisis espacial de la mortalidad infantil en Chile. Master's thesis, Pontificia Universidad Católica de Chile, 2001.

[72] C. Pascutto, J.C. Wakefield, N.G. Best, S. Richardson, L. Bernardinelli, A. Staines, and P. Elliott. Statistical issues in the analysis of disease mapping data. *Statistics in Medicine*, 19(17–18):2493–519, 2000.

[73] F. Peng and D.K. Dey. Bayesian analysis of outlier problems using divergences measures. *Canadian Journal of Statistics*, 23:199–213, 1995.

[74] J. Picka. Gaussian markov random fields: Theory and applications. *Technometrics*, 48(1):S146–147, 2006.

[75] D.J. Poirier. Revising beliefs in non identified models. *Econometric Theory*, 14:483–509, 1998.

[76] S. Richardson, A. Thomson, N.G. Best, and P. Elliot. Interpreting posterior relative risks estimates in disease mapping studies. *Environmental Health Perspectives*, 112:1016–1025, 2004.

[77] J. Roislien and O. Omre. T-distributed random fields: A parametric model for heavy-tailed well-log data. *Mathematical Geology*, 38(7):821–849, 2006.

[78] T.J. Rothenberg. Identification in parametric models. *Econometrica*, 39:577–591, 1971.

[79] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications.* Monographs on Statistics and Applied Probability 104. Chapman and Hall/CRC, Boca Ratón, Florida, 2005.

[80] M. Rytkönen. *Geographical study on childhood type 1 diabetes mellitus (T1DM) in Finland.* PhD thesis, Department of Geography, University of Oulu, 2004.

[81] U. Samuelsson, C. Johansson, J. Carstensen, and J. Ludvigsson. Space-time clustering in insulin-dependent diabetes mellitus (iddm) in South-East Sweden. *International Journal of Epidemiology*, 23:138–42, 1994.

[82] G. Schwarz. Estimating the dimmension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

[83] M. Songini, L. Bernardinelli, D. Clayton, C. Montomoli, C Pascutto, M. Ghislandi, D. Fadda, G.F. Bottazzo, and The Sardinia IDDM study 1. Epidemiology and geographical distribution of iddm in Sardinia during 1989 to 1994. *Diabetologia*, 41:221–227, 1998.

[84] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. Van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, (64):583–639, 2002.

[85] D.J. Spiegelhalter, A. Thomas, and N.G. Best. Winbugs user manual, v. 1.4.1, 2006.

[86] F. Torres, F. Pérez-Bravo, P. Iglesias, G. Icaza, and E. Carrasco. Spatio-temporal analysis of type 1 diabetes mellitus in Metropolitan Region, Chile, from 2000 to 2005. *Diabetic Medicine Journal*, 2008. submitted.

[87] I. Vidal, P. Iglesias, M. Branco, and R. Arellano-Valle. Bayesian sensitivity analysis and model comparison for skew-elliptical models. *Journal of Statistical Planning and Inference*, 136:3435–3457, 2006.

[88] J. Wakefield and P. Elliot. Issues in the statistical analysis of small area health data. *Statistics in Medicine*, 18:2377–2399, 1999.

[89] L.A. Waller. Bayesian thinking in spatial statistics. In D.K. Dey and C.R. Rao, editors, *Handbook of Statistics 25. Bayesian Thinking: Modeling and Computation*, pages 589–622, North Holland, 2005. Elsevier.

[90] R.E. Weiss. An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society, Series B*, 58:739–750, 1996.

[91] M. West. Outlier models and prior distribution in bayesian linear regression. *Journal of the Royal Statistical Society, Series B*, 46:431–439, 1984.

[92] M. West. On scale mixture of normals. *Biometrika*, 74(3):646–648, 1987.

[93] R. Wolfinger and M. O'Connell. Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 4:233–243, 1993.

[94] Y. Xie and B.P. Carlin. Measures of bayesian learning and identifiabiity in hierarchical models. *Journal of Statistical Planning and Inference*, 136:3458–3477, 2006.