

An Extension to the Scale Mixture of Normals for Bayesian Small-Area Estimation

Una extensión a la mezcla de escala de normales para la estimación
Bayesiana en pequeñas áreas

FRANCISCO J. TORRES-AVILÉS^{1,a}, GLORIA ICAZA^{2,b},
REINALDO B. ARELLANO-VALLE^{3,c}

¹DEPARTAMENTO DE MATEMÁTICA Y CIENCIA DE LA COMPUTACIÓN, FACULTAD DE CIENCIA,
UNIVERSIDAD DE SANTIAGO DE CHILE, SANTIAGO, CHILE

²INSTITUTO DE MATEMÁTICA Y FÍSICA, UNIVERSIDAD DE TALCA, TALCA, CHILE

³DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE MATEMÁTICAS, PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE, SANTIAGO, CHILE

Abstract

This work considers distributions obtained as scale mixture of normal densities for correlated random variables, in the context of the Markov random field theory, which is applied in Bayesian spatial intrinsically autoregressive random effect models. Conditions are established in order to guarantee the posterior distribution existence when the random field is assumed as scale mixture of normal densities. Lung, trachea and bronchi cancer relative risks and childhood diabetes incidence in Chilean municipal districts are estimated to illustrate the proposed methods. Results are presented using appropriate thematic maps. Inference over unknown parameters is discussed and some extensions are proposed.

Key words: Disease mapping, Markov random field, Hierarchical model, Incidence rate, Relative risk.

Resumen

Este trabajo aborda las distribuciones obtenidas como mezcla de escala de normales para variables aleatorias correlacionadas, en el contexto de la teoría de los campos markovianos, la cual es aplicada a modelos bayesianos espaciales con efectos aleatorios autoregresivos intrínsecos. Se establecen condiciones para garantizar la existencia de la distribución a posteriori cuando se

^aAssistant professor. E-mail: francisco.torres@usach.cl

^bAssociate professor. E-mail: gicaza@utalca.cl

^cProfessor. E-mail: reivalle@mat.puc.cl

asume una distribución mezcla de escala de normales para el campo markoviano propuesto. Para ilustrar los métodos propuestos, se estiman los riesgos relativos de cáncer de tráquea, bronquios y pulmón, y tasas de incidencia de diabetes tipo 1 en distritos municipales de Chile. Los resultados son presentados usando mapas temáticos apropiados. Se discute la inferencia sobre los parámetros desconocidos y se proponen algunas extensiones.

Palabras clave: campo aleatorio markoviano, mapeo de enfermedades, modelo jerárquico, riesgo relativo, tasa de incidencia.

1. Introduction

Over the last two decades, Bayesian spatial models have become increasingly popular for epidemiologists and statisticians. In particular, small-area modeling is oriented to illustrate the behavior of rates or relative risks associated to each district that form a region or a country, that is, recognition of spatial patterns through maps is the main aim of these methodologies. The conventional assumption to estimate the standardized mortality ratios (SMRs) or incidence rates is based on the Poisson distribution. This assumption may cause several problems in this class of studies, mainly because of the extra-Poisson variation. This extra-Poisson variation generally arises when the observed number of cases on each small-area are more variable than the variation contributed by the standard Poisson model (Mollié 2000). Bayesian models have been developed to solve this problem, introducing random effects to account for unobserved spatial heterogeneity; even more, Markov chain Monte Carlo (MCMC) methods led to an explosive increment of the use of Bayesian analysis in these areas of application.

Important works that develop and use Bayesian theory are mentioned in the following lines. The pioneering work in this direction was done by Clayton & Kaldor (1987) who proposed an empirical Bayes approach with application to lip cancer data in Scotland. In Ghosh, Natarajan, Stroud & Carlin (1998), conditions to demonstrate Bayesian generalized linear model (GLM) integrability are formalized under improper prior assumptions in order to represent lack of knowledge over unknown parameters. Best, Arnold, Thomas, Waller & Collon (1999) investigated several spatial prior distributions based on Markov random field (MRF) theory, and discussed methods for model comparison and diagnostics. Pascutto, Wakefield, Best, Richardson, Bernardinelli, Staines & Elliott (2000) examined some structural and functional assumptions of these models and illustrated their sensitivity through the presentation of results related to informal sensitivity analysis for prior distributions choices. They also explored the effect caused by outlying areas, assuming a Student-t distribution for the nonstructured effect.

Recently, Parent & Lesage (2008) proposed a linear Bayesian hierarchical model to study the knowledge spillovers in European countries, under different specifications of the proximity structures. They also compared this effect through different strategies, for example allowing different prior distributions or Student-t errors, to include heterogeneity in the disturbances.

As it was previously mentioned, the class of spatial models has been related to GLM theory, considering random effects to represent the influence of geography. Besag (1974) presented a pioneering work in the context of the MRF theory, with applications to regular lattice systems when spatial heterogeneity is considered.

Furthermore, the most used structure follows the work developed by Besag (1986), who presents a definition in the context of a MRF: Let $\mathbf{u} = (u_1, u_2, \dots, u_m)'$ be a set of m random variables and \mathbf{u}_{-i} the random vector without the i -th component, where m represents a number of different and contiguous areas. If the joint distribution of \mathbf{u} can be expressed by each conditional distribution, $u_i | \mathbf{u}_{-i}, i = 1, \dots, m$, then it is called a MRF. Intrinsically conditional autoregressive (CAR) random effects are defined as a particular MRF, initially proposed by Besag, York & Mollié (1991), which name is related to the impropriety of the joint distribution generated by the univariate conditional distributions of $u_i | \mathbf{u}_{-i}, i = 1, \dots, m$ (see details in Banerjee, Carlin & Gelfand 2004).

In this work, an intrinsically Gaussian MRF is considered and its properties are extended to a more general family of continuous distributions. Scale mixture of normal (SMN) distributions have been proposed as robust extensions of the normal model. The genesis of this class of models is presented by Andrews & Mallows (1974). The SMN class of distributions is generated if the vector of interest, \mathbf{u} , can be represented as

$$\mathbf{u} = \boldsymbol{\mu} + \psi^{-1/2} \mathbf{z} \quad (1)$$

where $\boldsymbol{\mu}$ is a location vector parameter, and \mathbf{z} and ψ are independent, with \mathbf{z} following a multivariate zero centered normal distribution with covariance matrix Σ and ψ being a non-negative random scale factor with c.d.f. $F_\psi(\cdot | \nu)$, so that $F_\psi(0 | \nu) = 0$. Here ν is an additional or set of parameters controlling the kurtosis of the distribution of ψ . The SMN distributions have been shown to be a subclass of the elliptical distributions family by Fang, Kotz & Ng (1990). This subfamily presents properties similar to the normal distributions, except that their behavior allows capturing unusual patterns present in the data. In a Bayesian context, robust linear models have been studied since Zellner (1976). The multivariate Student-t and the multivariate Slash distributions are examples of this class of distributions.

Following the above ideas, heavier-tailed models will be assumed instead of working with the usual assumption of normality for \mathbf{u} , through the Student-t and Slash distributions developed by Geweke (1993) and Lange & Sinsheimer (1993), respectively. Specifically, the Slash distribution considered in this work corresponds to the distribution of the random vector $\psi^{-1/2} \mathbf{z}$, where \mathbf{z} and ψ are independent, with \mathbf{z} having a multivariate normal distribution as in (1) and $\psi | \nu$ following a distribution $Beta(\nu/2, 1), \nu > 0$. The Student-t distribution is obtained through the same representation as the Slash distribution, with the difference that $\psi | \nu$ follows a Gamma distribution, where both parameters are equal to $\nu/2, \nu > 0$. There are some potential problems with the Slash distribution that probably has resulted in more use of the Student-t. However, the Slash distribution may allow for fatter tails (more extreme values) than the Student-t.

From a MRF context, the class of SMN 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 & Omre (2006) using a frequentist approach. Lyu & Simoncelli (2007) made the extension of Gaussian MRF theory to what they called Gaussian scale mixture fields, for image reconstruction modeling.

In this work, Bayesian non-Gaussian spatial models are developed to detect unusual rates or relative risks in a particular area under the following scheme. Standard small-area models are presented in Section 2. SMN theory is applied to extend the Gaussian MRF model (Besag 1974) in Section 3. In Section 4, non-Gaussian models are developed through extensions of the spatial random effect following a Gaussian MRF to the scale mixture of Normal random field (SMN RF) proposed previously. Three different models are used to estimate the incidence rates of Insulin Dependent Diabetes Mellitus (IDDM) in the Chilean Metropolitan Region, and Respiratory Cancer mortality in the northern regions of Chile. These results are presented in Section 5. Finally, some comments and discussion are made in Section 6.

2. Spatial Models with Random Effects

Let $\mathbf{y} = (y_1, \dots, y_m)'$ be a set of m random variables indexed to a specific region. A general formulation is assumed from the generalized linear mixed model theory (Breslow & Clayton 1993), which includes the following elements:

1. A specification of the likelihood function as member of the exponential family, namely

$$f(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\phi}) = \prod_{i=1}^m \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i)) + \rho(\phi_i; y_i)\}, \quad (2)$$

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

2. A random specification for the link function, $h(\theta_i) = E(y_i \mid \theta_i)$, is typically represented by the normal linear mixed model

$$h(\theta_i) \mid \mathbf{x}_i, \boldsymbol{\beta}, u_i, \sigma^2 \stackrel{ind.}{\sim} Normal(\mathbf{x}_i' \boldsymbol{\beta} + u_i, \sigma^2), \quad (3)$$

where the \mathbf{x}_i s are a $p \times 1$ vectors of covariates associated to a $p \times 1$ vector of coefficients $\boldsymbol{\beta}$, the u_i s represent spatial random effects, and σ^2 measures the nonstructured variability.

3. A model specification for the spatially structured random effects u_i s. Typically, Gaussian assumptions for the u_i s are made. In the literature it is recurrent to find that these spatial random effects are influenced by a predefined neighborhood represented by an adjacency matrix \mathbf{D}_w , controlling the

local variability. Hence, the mean is smoothed by the information given by its neighbors.

Let $\pi(\mathbf{u} \mid \sigma_u^2, \mathbf{D}_w) = \pi(u_1, \dots, u_m \mid \sigma_u^2, \mathbf{D}_w)$ be the joint probability distribution derived from a MRF given a dispersion parameter σ_u^2 and a $m \times m$ adjacency matrix \mathbf{D}_w . A multivariate Gaussian distribution is then obtained when,

$$\pi(\mathbf{u} \mid \sigma_u^2, \mathbf{D}_w) \propto \frac{1}{(\sigma_u^2)^{m/2}} \exp \left\{ -\frac{1}{2\sigma_u^2} \mathbf{u}' \mathbf{D}_w \mathbf{u} \right\}. \quad (4)$$

A specific case is considered in this work, where \mathbf{D}_w has 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, denoted by $i \sim j$, and 0 in other case, i.e.,

$$w_{ij} = \begin{cases} w_{i+} & i = j \\ -1 & i \neq j; i \sim j \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Under 5, equation 4 is reduced to

$$\pi(\mathbf{u} \mid \sigma_u^2, \mathbf{D}_w) \propto \frac{1}{(\sigma_u^2)^{m/2}} \exp \left\{ -\frac{1}{2\sigma_u^2} \sum_{i \sim j} (u_i - u_j)^2 \right\} \quad (6)$$

A basic discussion and treatment of several proximity matrices can be found in Banerjee et al. (2004). A constraint will be imposed to this expression to guarantee integrability.

4. 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

- i. $\pi(\boldsymbol{\beta}) \propto \text{constant}$
- ii. $\sigma^{-2} \sim \text{Gamma}(a/2, b/2)$
- iii. $\sigma_u^{-2} \sim \text{Gamma}(c/2, d/2)$

(7)

where the improper prior $\pi(\boldsymbol{\beta})$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)' \in \mathbb{R}^p$, is assumed according with Ghosh et al. (1998). The hyperparameters $a, b, c, d > 0$ are known constant. Here, both σ^2 and σ_u^2 represent the dispersion parameters included in the model; σ_u^2 is the local dispersion parameter related to a specific spatial structure. Another useful measure in spatial models is the percentage of spatial aggregation explained by the model, which usually is measured by the ratio

$$\frac{\sigma_u^2}{\sigma_u^2 + \sigma^2} \times 100\% \quad (8)$$

Its interpretation is related to obtain the relative contribution given by the spatial aggregation effect. Here, a common estimation of σ_u^2 is the empirical variance s_u^2 , which can be obtained from the estimation of \mathbf{u} for each MCMC iteration.

3. Scale mixture of Intrinsically CAR Models

In this section an extension of the usual multivariate Gaussian MRF is proposed, assuming a multivariate SMN distribution. The next definition will provide an extension of (4) to the SMN random field (SMN RF).

Definition 1. A spatial random vector $\mathbf{u} = (u_1, \dots, u_m)'$ follows an SMN RF, if the kernel distribution can be obtained as

$$\pi(\mathbf{u} \mid \sigma_u^2, \mathbf{D}_w, \nu) \propto \int_0^\infty \left(\frac{\psi}{\sigma_u^2} \right)^{m/2} \exp \left\{ -\frac{\psi}{2\sigma_u^2} \mathbf{u}' \mathbf{D}_w \mathbf{u} \right\} dF(\psi \mid \nu) \quad (9)$$

where $F(\cdot \mid \nu)$ is the c.d.f. of $\psi \mid \nu$, σ_u^2 is a dispersion parameter, and \mathbf{D}_w denotes a adjacency matrix. A SMN RF with scale parameter σ_u^2 will be denoted as $SMNRF(\mathbf{0}, \sigma_u^{-2} \mathbf{D}_w, \nu)$.

For the Gaussian case, it is known that specification of \mathbf{D}_w in (5) makes (4) improper (Banerjee et al. 2004), since the matrix \mathbf{D}_w is singular, so that its inverse does not exist, hence

$$\int_{\mathbb{R}^m} \pi(\mathbf{u} \mid \sigma_u^2, \mathbf{D}_w, \nu) d\mathbf{u} \propto \int_{\mathbb{R}^m} \frac{1}{(\sigma_u^2)^{m/2}} \exp \left\{ -\frac{1}{2\sigma_u^2} \mathbf{u}' \mathbf{D}_w \mathbf{u} \right\} d\mathbf{u} = \infty$$

The last equation implies that a density function is available, but not integrable. This result is the intrinsic autoregressive 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 (9), then consistent property (Kano 1994) fails. Therefore, integration theory can not be applied.

In the same way as the joint distribution of the Gaussian MRF treated in the spatial literature, for every SMN RF, the joint distribution will also be 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 distribution, as discussed in Banerjee et al. (2004) and Assunção, Potter & Cavenaghi (2002). The next proposition establishes conditions to make proper the associated SMN RF. The proof of this proposition can be found in the appendix.

Proposition 1. *Suppose that a set of spatial indexed random variables, represented by the vector $\mathbf{u} = (u_1, \dots, u_m)'$, is available. Consider the SMN RF in (9) as the distribution of \mathbf{u} . Additionally, let us suppose that $F(\cdot \mid \nu)$ is a known positive c.d.f. If $\sum_{i=1}^m u_i = 0$ and $\mathbb{E}(\psi^{1/2} \mid \nu) < \infty$, then (9) is proper.*

Specific choices for $F(\cdot \mid \nu)$ in (9) lead to different scale mixture probability distributions. Student-t and Slash MRFs will be used in this work, which can be obtained using stochastic representations which depend on the selected mixing distribution $F(\cdot \mid \nu)$.

The SMN RF can be represented hierarchically in terms of two stages:

At the first stage of the hierarchy, a Gaussian MRF is specified with an additional random scale factor ψ . At the second stage, a mixing distribution for the scale perturbation ψ is then specified. Specifically:

1. For the Student- t MRF:

$$\text{i) } \mathbf{u} \mid \sigma_u^2, \psi, \mathbf{D}_w \sim \text{Normal}(\mathbf{0}, \sigma_u^{-2}\psi\mathbf{D}_w) \quad (10)$$

$$\text{ii) } \psi \mid \nu \sim \text{Gamma}(\nu/2, \nu/2) \quad (11)$$

In this case, the Student- t MRF with ν degrees of freedom follows, which is denoted by $\mathbf{u} \mid \mathbf{D}_w, \sigma_u^2, \nu \sim t(\mathbf{0}, \sigma_u^{-2}\mathbf{D}_w, \nu)$.

2. For the Slash MRF:

$$\text{i) } \mathbf{u} \mid \sigma_u^2, \psi, \mathbf{D}_w \sim \text{Normal}(\mathbf{0}, \sigma_u^{-2}\psi\mathbf{D}_w) \quad (12)$$

$$\text{ii) } \psi \mid \nu \sim \text{Beta}(\nu/2, 1) \quad (13)$$

In this case, the Slash MRF, denoted by $\mathbf{u} \mid \mathbf{D}_w, \sigma_u^2, \nu \sim \text{Slash}(\mathbf{0}, \sigma_u^{-2}\mathbf{D}_w, \nu)$, is obtained.

The model described above is useful to implement the MCMC method. It is important to mention that the distribution of both of the above random fields has the finite condition exposed in Proposition 1.

A prior distribution for ν is required in order to assume a valid Bayesian model. Usually, an exponential distribution prior is considered for this parameter, which is assumed independent of (6), that is,

$$\nu \mid \delta_0 \sim \exp(\delta_0), \quad \delta_0 > 0 \quad (14)$$

Assuming (2), (3), (7), (9) and (14), the full joint posterior distribution is specified as,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma^2, \psi, \nu \mid \mathbf{y}, \mathbf{D}_w, \delta_0) &\propto \prod_{i=1}^m \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\} \\ &\times \prod_{i=1}^m \exp\{-(1/2\sigma^2)(h(\theta_i) - \mathbf{x}'_i\boldsymbol{\beta} - u_i)^2\} h'(\theta_i) \\ &\times \exp\{-(\psi/2\sigma_u^2)\mathbf{u}'\mathbf{D}_w\mathbf{u}\} \psi^{m/2} (\sigma^2\sigma_u^2)^{-m/2}, \\ &\times \exp\{-a/2\sigma_u^2\} (\sigma^2)^{-(b/2+1)} \\ &\quad \exp\{-c/2\sigma^2\} (\sigma_u^2)^{-(d/2+1)} \\ &\times f(\psi \mid \nu) \exp\{-\delta_0\nu\} \end{aligned} \quad (15)$$

where $f_\psi(\cdot \mid \nu)$ represents the conditional density or probability function of $\psi \mid \nu$. See item 3 of the appendix for the computational aspects.

4. Proposed Bayesian Small-Area Models

An important point is to demonstrate the integrability of the proposed model. Under the generalized linear model (2), link function (3) and prior assumption given by (7) and (14), Theorem 2 from the work developed by Ghosh et al. (1998) gives the conditions to obtain a proper posterior distribution for $\boldsymbol{\theta} \mid \mathbf{y}$ when $P(\psi = 1 \mid \nu) = 1$ (the Gaussian MRF). Following that theorem, it is possible to find a generalization towards the SMN case.

The next proposition gives conditions when the spatial random effect follows an SMN RF. Its proof is given in the appendix.

Proposition 2. *Consider the model (2), link function (3) and prior assumption given by (7) and (14). Consider also the assumptions of Proposition 1 and the following additional conditions:*

- i. $\theta_i \in (\underline{\theta}_i, \bar{\theta}_i)$, for some $-\infty < \underline{\theta}_i < \bar{\theta}_i < \infty$, $i = 1, \dots, m$;
- ii. $m - p + a - 1 > 0$;
- iii. $b > 0$, $d > 0$, $m + c > 0$

If the condition of integrability

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

is verified for all $i = 1, \dots, m$, then posterior distribution $\pi(\boldsymbol{\theta} \mid \mathbf{y})$ is proper.

The main interest is focused in establishing a non-Gaussian parametric spatial random effect. A MCMC structure seems to be adequate to make inferences from this class of model. Most full conditional distributions computed for this scheme are known distributions, therefore, a hybrid Gibbs sampling - metropolis Hastings algorithm is used to generate samples from the joint posterior distribution. The algorithm given in item 3 of the appendix presents the full conditional distributions for this particular model.

5. Applications

The proposed spatial Bayesian models will be applied assuming SMN random effects for two real data in the epidemiological framework to control for excessive smoothness in small areas with sparse data. One dataset is related to IDDM incidence rates in the Chilean municipal districts from Metropolitan region and the other dataset contains female lung, trachea and bronchi cancer standardized mortality ratios in the municipal districts of the country's northern zone. The municipal district is the smallest administrative area in Chile. In this country there are only few published studies related to spatial epidemiology (Andia, Hsing, Andreotti & Ferreccio 2008, Ferreccio, Rollán, Harris, Serrano, Gederlini, Margozzini, Gonzalez, Aguilera, Venegas & Jara 2007, Icaza, Núñez, Torres, Díaz & Varela 2007, Icaza, Núñez, Díaz & Varela 2006, Torres-Avilés, Icaza, Carrasco &

Pérez-Bravo 2010). Results from non-Gaussian spatial Bayesian modeling related to both diseases are presented in the next subsections.

The specific model that is considered for these two applications is the Poisson hierarchical model given by

$$\begin{aligned} y_i | e_i, \lambda_i &\stackrel{ind.}{\sim} Poisson(e_i \lambda_i) \\ \log(\lambda_i) | \beta_0, u_i, \sigma^2 &\stackrel{ind.}{\sim} Normal(\beta_0 + u_i, \sigma^2) \end{aligned}$$

$i = 1, \dots, m$, where $\mathbf{y} = (y_1, \dots, y_m)'$ represents the observed sample vector associated to m different regions under study, $\mathbf{e} = (e_1, \dots, e_m)'$ represents the population at risk or the expected population associated to the m different regions, and $\mathbf{u} = (u_1, \dots, u_m)'$ is the vector of random effects which is assumed to have a SMN distribution constrained to sum zero. Diffuse prior distributions are considered for the location and scale parameters, as those presented in (7). For the variance parameters, σ^2 and σ_u^2 , the hyperparameters $a = b = c = d = 0.001$ were assumed.

Posterior estimations are obtained from a single run of the Gibbs sampler, with a burn-in of 1000 iterations followed by 10000 further cycles. Convergence have been checked through trace and autocorrelation plots. Three common ways to measure model assessment are taken into account. The first two are oriented to penalize the observed deviance: The deviance information criterion (DIC) (Spiegelhalter, Best, Carlin & Van der Linde 2002) and a modified BIC (Congdon 2003) will be used. A third model choice criterion is applied, proposed by Gelfand & Ghosh (1998), which 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. As was described in the introduction, the competing models are related to Gaussian, Student-t and Slash MRF. The percentage of spatial variability is computed using expression (8).

5.1. Insulin Dependent Diabetes Mellitus Incidence, Metropolitan Region, Chile

The objective of this study is to describe spatial patterns of type 1 diabetes in children under 15 of age, diagnosed between 2000 and 2005 with residence in the Metropolitan Region of Chile. The Metropolitan Region is located in the centre of Chile. According to the Chilean National Institute of Statistics (INE), this region represents an area of approximately 15,403 km². Total population living at Metropolitan Region was 6,061,185 inhabitants, according to the 2002 census. Metropolitan population represents 40% of the whole country. The region is divided into 52 districts, 18 are considered as rural and 34 as highly urbanized, known as Greater Santiago, in the centre of the region, with the 96.9% of the metropolitan population. With respect to the population at risk, children under 15 years of age represent the 24.9% of the metropolitan region population, which is composed by 1,509,218 children. A population-based registry of type 1 diabetes in children under than 15 years of age has been available in the Metropolitan Region since 2000. See Carrasco, Pérez-Bravo, Dorman, Mondragón & Santos (2006)

for details about the registry. Torres-Avilés et al. (2010) show an aggregation on incidence rates in urban areas of the Chilean Metropolitan Region, using the Bayesian methodology proposed by Mollié (2000).

TABLE 1: IDDM model selection criteria, DIC, BIC and predictive check.

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

Model selection criteria results are presented in Table 1. According to previously mentioned goodness of fit criteria, small values imply better adjustment. Therefore, a spatial model that includes Slash random effects with 7 d.f. is a strong candidate to model geographic dependence. This result seems to be adequate due to those extreme values, which match with the higher socioeconomic areas of the region, as is explained in next paragraphs. The predictive measure $G&G$ disagrees with the other methods; this can be interpreted as a “failure of the model for prediction”, pointing out a better performance of the usual Gaussian MRF.

TABLE 2: Posterior mean, standard deviation and 95% HPD credibility intervals for unknown parameters when a Gaussian MRF, Student-t MRF and Slash MRF are assumed.

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

Robust Bayesian models proposed in the previous section were applied to this problem. Inferences over unknown parameters are displayed in Table 2, when Gaussian MRF, Student-t MRF and Slash MRF are assumed to control spatial variability. Similar values are estimated for β_0 and σ^2 , under the three MRF models, showing the models’ robustness. In contrast, σ_u^2 presents different values, depending on the distribution assumed for the MRF. The non-Gaussian model (Slash MRF) increases the degree of spatial aggregation from 44.1 % to 54.6 %, that is, the excess of spatial variability presented in these data seems mostly due

to a clustering effect. Notice that the estimated degrees of freedom are small, which implies that the excess of variability is better captured by one of the SMN RF model.

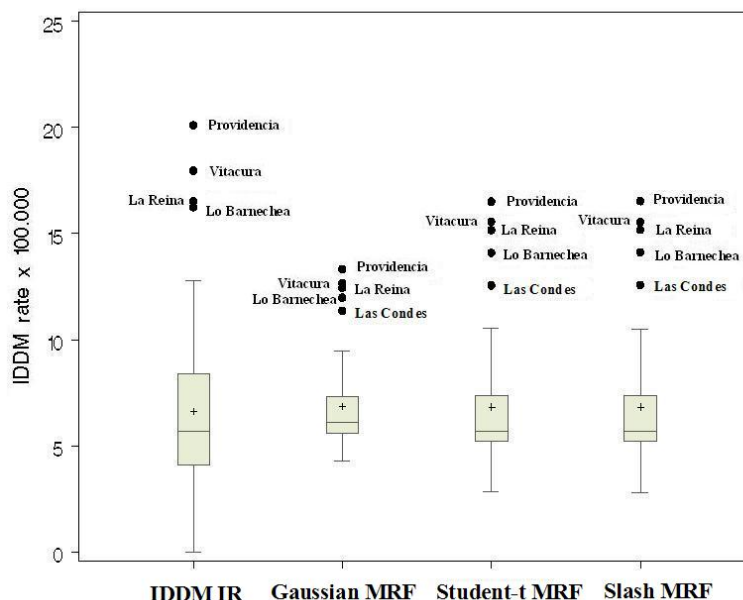


FIGURE 1: IDDMM incidence rate (IR) variability: Raw estimates, Mollié’s convolution model (Gaussian MRF), Student-t convolution model (Student-t MRF) and Slash convolution model (Slash MRF).

Figure 1 shows that fully Bayesian estimates of IDDMM incidence rates present less variation than raw incidence rate. The three Bayesian variation plots seem to have a similar behavior, due to the presence of several municipal districts with high incidence rates, which are considered as outliers. Comparing the four box-plots, the three fitted models (Gaussian, Student-t and Slash) present an additional municipal district, named Las Condes, as part of the higher incidence group. The normal MRF assumption leads to estimate smoother rates; however, Student-t, and Slash MRF’s present slight variability differences. Those differences allow controlling the excess of smoothness, i.e., non-Gaussian shrinkage gives a more adequate estimate of the pattern of underlying risk of disease than that provided by the Mollié’s convolution estimates.

From Figure 2, high incidence estimates remain in municipal districts with high socioeconomic level, such as Vitacura and Providencia, located at the northeast side of the map. These results were already found by Torres-Avilés et al. (2010). Slight differences are seen when Slash MRF (d) and Student-t MRF (c) models are assumed, but these differences are clinically important since are in rural municipal districts with zero cases of diabetes located at southwest side of the map.

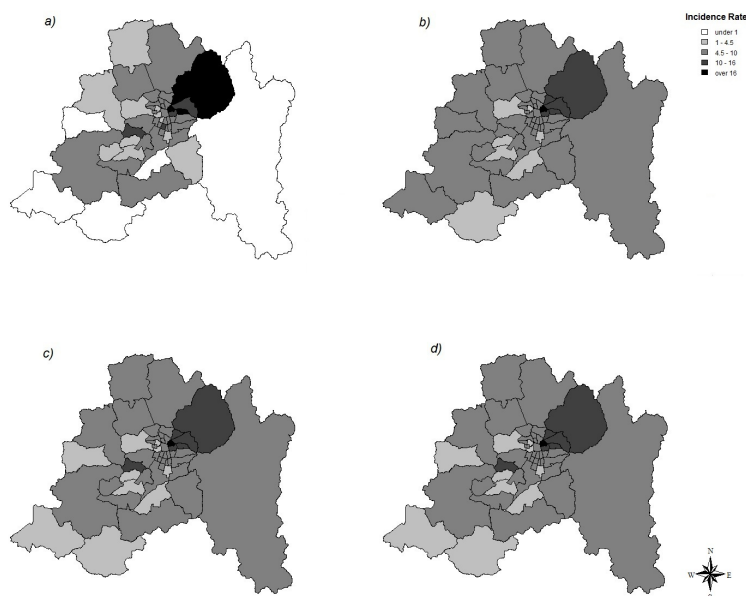


FIGURE 2: IDDm incidence rate by district: a) Raw incidence rates. b) Mollie's convolution model (Gaussian MRF). c) Student-t convolution model (Student-t MRF). d) Slash convolution model (Slash MRF).

5.2. Female Trachea, Bronchi and Lung Cancer Mortality, Chilean Northern Regions

Bayesian methods that have been applied to several real problems to estimate relative risks of cancer mortality in small-areas can be found in the literature, e.g., Ghosh et al. (1998) and Pascutto et al. (2000), and Mollié (2000). In particular, this application is related to estimate female lung, bronchi and trachea cancer mortality relative risks in the northern regions of Chile. The northern region of Chile represents an area of approximately 300,904 km². According to the 2002 census there were 819,177 women inhabitants in this part of the country. The region is divided into 43 districts, many of them (20 or 47%) with less than 10,000 inhabitants. The aim of this study is to describe the geographical distribution of this class of mortality, which has presented smoothness problems in comparison with the usual model.

Mortality statistics for the years 1997-2004 published by the Chilean Ministry of Health were used. The SMR was calculated for 341 districts in the country. Results show an excess of mortality caused by trachea, bronchi and lung cancer in the region. A previous work can be found, where the analysis for both sexes was done for the whole country and published by Icaza et al. (2007). The problem arised when Mollié's model estimates for women cancer mortality risks were too smooth and high in municipal districts where zero cases occurred.

TABLE 3: Cancer mortality model selection criteria, DIC, BIC and predictive check.

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

TABLE 4: Posterior mean, standard deviation and 95% HPD credibility intervals for unknown parameters when a Gaussian MRF, Student-t MRF and Slash MRF 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)

For this application, Table 3 shows a better fit for the model that includes the Slash spatial random effect with approximately 32 degrees of freedom, as can be seen in Table 4. Once again, the Slash can not be considered as a predictive alternative, in contrast to a parsimonious model such as the Student-t or the Gaussian MRF. One important result is referred to the 79% estimated proportion of spatial variability associated to this model. Notice that this proportion is almost the same for the three proposed models. This could be related to the estimated degrees of freedom. One important issue is related to the estimation for the other parameters, such as β_0 or baseline risk, which is not affected by the model.

Standardized mortality ratios and Risk estimations are compared in Figure 3. It is important to add that variability estimation is reduced when any of the Bayesian models is considered. All of them show an improvement in contrast to the SMR, and a district called Mejillones is separated from the rest of the distribution, showing the highest risk in the north for this mortality.

Figure 4 displays the cancer mortality relative risk estimation using three different models, with Mollíe'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.

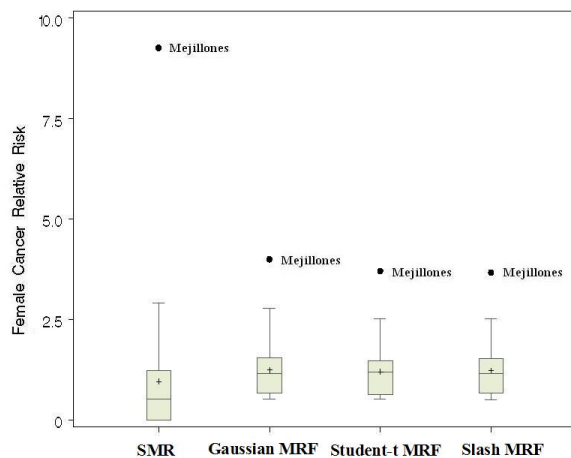


FIGURE 3: Female trachea, bronchi and lung cancer SMR variability: Standardized mortality ratio, Mollié’s convolution model (Gaussian MRF), Student-t convolution model (Student-t MRF) and Slash convolution model (Slash MRF).

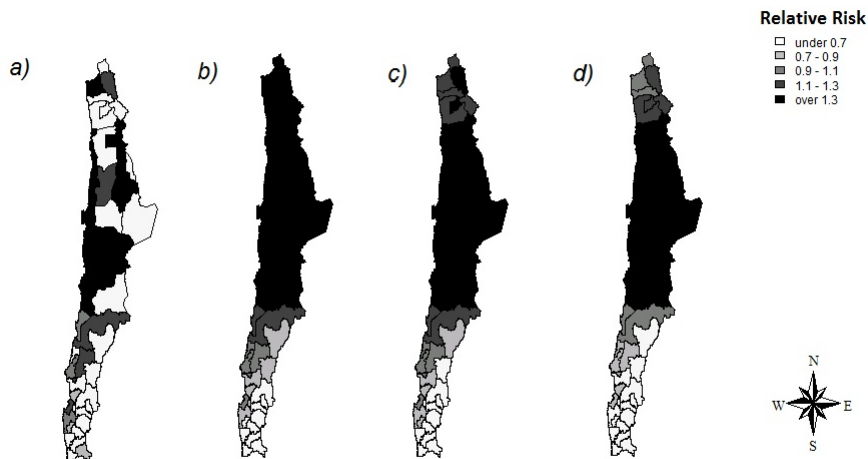


FIGURE 4: Female trachea, bronchi and lung cancer SMR by district: a) Standardized mortality ratio (SMR). b) Mollié’s convolution model. c) Student-t convolution model (Student-t MRF). d) Slash convolution model (Slash MRF).

According to the DIC and BIC criteria, the selected Slash MRF model presented better fitted rates, even when Figure 4(d) shows that the first and darkest area in the extreme north, the most populated municipal district (Arica) in that

region, presents the highest rates compared to its closer neighbors. It was not possible to reduce the effect produced by the larger areas in the next darkest zones, which correspond to Tarapacá and Antofagasta regions, which are located in the Atacama Desert. The over-smoothing effect lead to flat true variations in risk, even by the selected model.

6. Concluding Remarks

In this work, a non-Gaussian Bayesian-small area estimation is proposed as an alternative to usual parametric models. This approach is particularly useful to obtain estimations of rates or relative risks when subjective geographical dependence is assumed and related results are too smooth for the region under study.

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 was available, Proposition 2 provided sufficient conditions to guarantee posterior distribution integrability for Bayesian GLM.

The general methodology is applicable to situations where small area parameters must be estimated. Variability parameters are of interest, since 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 has direct relation with the spatial random effect contribution considered for the analysis. As mentioned in Banerjee et al. (2004, p. 166), differences may exist in percentage of variability estimation, when other prior distributions are considered. A prior sensitivity analysis is not studied in this work.

Considering the complex structure of Chilean geography, better results were obtained using our proposed strategy. Both applications were best modeled by Poisson regression with spatial random effects following a joint Slash distribution. It can be seen that β_0 does not produce changes when the three models are fitted to both applications. That is an important consideration that shows the non-Gaussian properties of the Student-t MRF and Slash MRF.

In the future, several topics can be explored in the spatial context. Diagnostic approaches 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 also be made.

Bayesian space time models can be proposed, with the subsequent problem of sparseness of data that could affect estimation in municipal districts with low population. Therefore, non-Gaussian 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.

Acknowledgements

The authors acknowledge the helpful and constructive comments of the two anonymous referees, which significantly improved the quality of this paper. Torres-Avilés's research was partially supported by DICYT 040975TA and FONDECYT de Iniciación 11110119 grants. Arellano-Valle's research was partially supported by FONDECYT 1085241-Chile grant. We are very grateful to Ms. Elena Carrasco, Dr. Francisco Pérez-Bravo, the *Fundación de Diabetes Juvenil* and the Department of Statistics of the *Ministerio de Salud de Chile* for providing the data used in this article. We acknowledge the helpful initial discussions with Dr. Pilar Iglesias Z.

[Recibido: septiembre de 2011 — Aceptado: febrero de 2012]

References

- Andia, M., Hsing, A. W., Andreotti, G. & Ferreccio, C. (2008), 'Geographic variation of gallbladder cancer mortality and risk factors in Chile: A population-based ecologic study', *International Journal of Cancer* **123**(6), 1411–1416.
- Andrews, D. F. & Mallows, C. L. (1974), 'Scale mixture of normal distributions', *Journal of the Royal Statistical Society Series B* **36**(1), 99–102.
- Assunção, R. M., Potter, J. E. & Cavenaghi, S. M. (2002), 'A Bayesian space varying parameter model applied to estimating fertility schedules', *Statistics in Medicine* **21**, 2057–2075.
- Banerjee, S., Carlin, B. & Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Monographs on Statistics and Applied Probability 101. Chapman and Hall, Boca Raton, Florida.
- Besag, J. (1974), 'Spatial interaction and the statistical analysis of lattice systems', *Journal of the Royal Statistical Society Series B* **36**(2), 192–236.
- Besag, J. (1986), 'On the statistical analysis of dirty pictures', *Journal of the Royal Statistical Society Series B* **48**(3), 259–302.
- Besag, J., York, J. & Mollié, A. (1991), 'Bayesian image restoration, with two applications in spatial statistics', *Annals of the Institute of Statistical Mathematics* **43**, 1–59.
- Best, N., Arnold, R., Thomas, A., Waller, L. & Collon, E. (1999), Bayesian models for spatially correlated disease and exposure data, in J. Bernardo, A. Smith, A. Dawid & J. Berger, eds, 'Bayesian Statistics 6', Oxford University Press, Oxford, pp. 131–156.
- Breslow, N. & Clayton, D. (1993), 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association* **88**, 9–25.

- Carrasco, E., Pérez-Bravo, F., Dorman, J., Mondragón, A. & Santos, J. L. (2006), 'Increasing incidence of type 1 diabetes in population from Santiago of Chile: Trends in a period of 18 years (1986-2003)', *Diabetes/Metabolism Research and Reviews* **22**, 34-37.
- Clayton, D. & Kaldor, J. (1987), 'Empirical Bayes estimates of age-standardized relative risks for use in disease mapping', *Biometrics* **43**, 671-681.
- Congdon, P. (2003), *Applied Bayesian Modelling*, Wiley & Sons, Chichester.
- Damien, P. & Walker, S. (2001), 'Sampling truncated normal, beta and gamma densities', *Journal of Computational and Graphical Statistics* **10**(2), 206-215.
- Fang, K. T., Kotz, S. & Ng, K. W. (1990), *Symmetric Multivariate and Related Distributions*, Chapman and Hall, New York.
- Ferreccio, C., Rollán, A., Harris, P., Serrano, C., Gederlini, A., Margozzini, P., Gonzalez, C., Aguilera, X., Venegas, A. & Jara, A. (2007), 'Gastric cancer is related to early Helicobacter pylori infection in a high prevalence country', *Cancer Epidemiology, Biomarkers & Prevention* **16**, 662-667.
- Gelfand, A. E. & Ghosh, S. K. (1998), 'Model choice: A minimum posterior predictive loss approach', *Biometrika* **85**, 1-11.
- Geweke, J. (1993), 'Bayesian treatment of the independent Student-t linear model', *Journal of Applied Econometrics* **8**, 519-540.
- Ghosh, M., Natarajan, K., Stroud, T. W. F. & Carlin, B. P. (1998), 'Generalized linear models for small-area estimation', *Journal of the American Statistical Association* **93**(441), 273-282.
- Icaza, G., Núñez, L., Díaz, N. & Varela, D. (2006), *Atlas de mortalidad por enfermedades cardiovasculares en Chile, 1997- 2003*, Universidad de Talca y Ministerio de Salud, New York.
- Icaza, G., Núñez, L., Torres, F., Díaz, N. & Varela, D. (2007), 'Distribución geográfica de mortalidad por tumores malignos de tráquea, bronquios y pulmón, Chile 1997-2004', *Revista Médica de Chile* **135**(11), 1397-1405.
- Kano, Y. (1994), 'Consistency property of elliptical probability density functions', *Journal of Multivariate Analysis* **51**, 139-147.
- Lange, K. & Sinsheimer, J. S. (1993), 'Normal/independent distributions and their applications in robust regression', *Journal of Computational and Graphical Statistics* **2**(2), 175-198.
- Lyu, S. & Simoncelli, E. P. (2007), Statistical modeling of images with fields of Gaussian scale mixtures, in B. Schölkopf, J. Platt & T. Hoffman, eds, 'Advances in Neural Information Processing Systems, 19', MIT Press, Cambridge, pp. 945-952.

- Mollié, A. (2000), Bayesian mapping of Hodgkin's disease in France, *in* P. Elliott, J. Wakefield, N. G. Best & D. J. Briggs, eds, 'Spatial Epidemiology: Methods and Applications', Oxford University Press, New York, pp. 267–285.
- Parent, O. & Lesage, J. P. (2008), 'Using the variance structure of the conditional autoregressive specification to model knowledge spillovers', *Journal of Applied Economics* **23**, 235–256.
- Pascutto, C., Wakefield, J. C., Best, N. G., Richardson, S., Bernardinelli, L., Staines, A. & Elliott, P. (2000), 'Statistical issues in the analysis of disease mapping data', *Statistics in Medicine* **19**(17-18), 2493–519.
- Roislien, J. & Omre, O. (2006), 'T-distributed random fields: A parametric model for heavy-tailed well-log data', *Mathematical Geology* **38**(7), 821–849.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society, Series B* **64**, 583–639.
- Torres-Avilés, F., Icaza, G., Carrasco, E. & Pérez-Bravo, F. (2010), 'Clustering of cases of type 1 diabetes in high socioeconomic communes in Santiago de Chile: Spatio-temporal and geographical analysis.', *Acta Diabetologica* **47**(3), 251–257.
- Zellner, A. (1976), 'Bayesian and non-Bayesian analysis of the regression model with multivariate student-t error terms', *Journal of the American Statistical Association* **71**(354), 400–405.

Appendix

1. **Proof of Proposition 1.** As was showed by Assunção et al. (2002), the $\sum_{i=1}^m u_i = 0$ constraint makes the Gaussian kernel (4) proper; i.e., on the set $C = \{\mathbf{u} \in \mathbb{R}^m : \sum_{i=1}^m u_i = 0\}$, we have

$$\int_C \frac{1}{(\sigma_u^2)^{m/2}} \exp \left\{ -\frac{1}{2\sigma_u^2} \mathbf{u}' \mathbf{D}_w \mathbf{u} \right\} d\mathbf{u} < \infty$$

Hence, under the $\sum_{i=1}^m u_i = 0$ constraint, by applying the Fubini's theorem and the change variable $\mathbf{y} = \psi^{1/2} \mathbf{x}$, we have in (9) that

$$\begin{aligned} \int_C \pi(\mathbf{u} | \sigma_u^2, \mathbf{D}_w, \nu) d\mathbf{u} &= \int_0^\infty \psi^{m/2} \int_C \frac{1}{(\sigma_u^2)^{m/2}} \exp \left\{ -\frac{\psi}{2\sigma_u^2} \mathbf{u}' \mathbf{D}_w \mathbf{u} \right\} d\mathbf{u} dF(\psi | \nu) \\ &\propto \int_0^\infty \psi^{1/2} dF(\psi | \nu) < \infty \quad * \end{aligned}$$

2. **Proof of Proposition 2.** From (2), (3), (7), (9) and (14) we have for the full joint posterior distribution that

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma^2, \psi, \nu \mid \mathbf{y}, \mathbf{D}_w, \delta_0) &\propto \prod_{i=1}^m \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\} \\ &\times \prod_{i=1}^m \exp\{-(1/2\sigma^2)(h(\theta_i) - \mathbf{x}'_i\boldsymbol{\beta} - u_i)^2\}h'(\theta_i) \\ &\times \exp\{-(\psi/2\sigma_u^2)\mathbf{u}'\mathbf{D}_w\mathbf{u}\}\psi^{m/2}(\sigma^2\sigma_u^2)^{-m/2}, \\ &\times \exp\{-a/2\sigma_u^2\}(\sigma^2)^{-(b/2+1)} \\ &\quad \exp\{-c/2\sigma^2\}(\sigma_u^2)^{-(d/2+1)} \\ &\times f(\psi \mid \nu) \exp\{-\delta_0\nu\} \end{aligned}$$

where $f(\cdot \mid \nu)$ is the conditional density (or probability) function of $\psi \mid \nu$. Integrating with respect to $\boldsymbol{\beta}$, σ^2 and σ_u^2 , we obtain

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{u}, \psi, \nu \mid \mathbf{y}, \mathbf{D}_w, \delta_0) &\propto \prod_{i=1}^m \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\}h'(\theta_i) \\ &\times \psi^{m/2}(a + \psi\mathbf{u}'\mathbf{D}_w\mathbf{u})^{-(m+b-1)/2} \\ &\times f(\psi \mid \nu) \exp\{-\delta_0\nu\} \end{aligned}$$

Notice that this last result has a multivariate Student-t kernel. Now, integrating over $\mathbf{u} \in \mathbb{R}^m$ under the constraint $\sum_{i=1}^m u_i = 0$, the following result is obtained,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \psi, \nu \mid \mathbf{y}, \mathbf{D}_w, \delta_0) &\leq K \prod_{i=1}^m \exp\{\phi_i^{-1}(y_i\theta_i - g(\theta_i))\}h'(\theta_i) \\ &\times f(\psi \mid \nu) \exp\{-\delta_0\nu\} \end{aligned}$$

where K is a constant that does not depend on $\boldsymbol{\theta}$ or any of the parameters previously integrated. Finally, integration over ψ and then over ν leads to the desired result.*

3. **Proposed MCMC Algorithm.** To implement the Gibbs sampling, the full conditional distributions associated with the full joint posterior distribution (15) are given in the following, in which $\mathbf{h}(\boldsymbol{\theta}) = (h(\theta_1), \dots, h(\theta_m))'$ denotes the link vector and \mathbf{X} is the $m \times p$ design matrix which has rows $\mathbf{x}_1, \dots, \mathbf{x}_m$.

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

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

b) $\mathbf{u} \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2, \sigma_u^2, \psi, \mathbf{X}, \mathbf{D}_w \sim \text{Normal}(\boldsymbol{\mu}_u, \mathbf{V}_u)$, where

$$\boldsymbol{\mu}_u = \frac{1}{\sigma^2} \mathbf{V}_u (\mathbf{h}(\boldsymbol{\theta}) - \mathbf{X}\boldsymbol{\beta}), \mathbf{V}_u = \left(\frac{1}{\sigma^2} \mathbf{I}_m + \frac{\psi}{\sigma_u^2} \mathbf{D}_w \right)^{-1}$$

and \mathbf{I}_m is the identity matrix of size m

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

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

d) $\sigma_u^{-2} \mid \mathbf{u}, \psi, \mathbf{D}_w, c, d \sim \text{Gamma}(c^*, d^*)$ where,

$$c^* = \frac{m + c}{2} \text{ and } d^* = \frac{1}{2} (\psi(\mathbf{u}'\mathbf{D}_w\mathbf{u}) + d)$$

e) Choice of a distribution for the scale random factor ψ :

i. If $\psi \mid \nu \sim \text{Gamma}(\nu/2, \nu/2)$, then

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

ii. If $\psi \mid \nu \sim \text{Beta}(\nu/2, 1)$, then

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

where $\mathbf{1}_A$ represents the indicator function. Notice the presence of a truncated Gamma distribution in the $[0, 1]$ interval. To draw from this distribution, the Damien & Walker (2001) algorithm can be performed.

f) Degrees of freedom are estimated from

fa. If $\psi \mid \nu \sim \text{Gamma}(\nu/2, \nu/2)$, then

$$\pi(\nu \mid \psi, \delta_0) \propto \Gamma(\nu/2)^{-1} \nu^{\nu/2} \exp\{-\nu(\delta_0 + 0.5(\psi - \ln(\psi)))\}$$

fb. If $\psi \mid \nu \sim \text{Beta}(\nu/2, 1)$, then

$$\nu \mid \psi, \delta_0 \sim \text{Gamma}(2, \delta_0 - \ln(\psi/2)) \mathbf{1}_{(0,1)}(\psi)$$

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

The algorithm must be iterated until convergence is detected in order to start to take a sample.