

Spatial Analysis of Incidence Rates: A Bayesian Approach

Silvio A. da Silva, Luiz L.M. Melo and Ricardo Ehlers

July 2004

Abstract

Spatial models have been used in many fields of science where the data are collected in different locations, i.e. each observation is associated to a point in space. In particular, the analysis of spatial dispersion of the risk of occurrence of a certain event is in general performed via maps of incidence, where a set of areas is shaded according to the values of a variable of interest. The goal of this mapping is to infer the geographic distribution of the rates thus identifying areas of higher or lower incidence.

In this work, maps of incidence rates will be constructed using a Bayesian approach. In particular, situations where the geographic units have small populations will be studied. The variability of the rates due to the differences among areas of underlying risk is separated from that due to random fluctuations via appropriate prior distributions. Some combinations of co-variables will also be evaluated in order to check for possible influence of social factors on the incidence rates. The risks will be estimated via Markov chain Monte Carlo (MCMC) methods and we select the most appropriate model among those analysed via the Deviance information criterion (DIC).

Key-words: Spatial analysis, Bayesian inference, MCMC, WinBUGS, Gibbs sampler, conditionally autoregressive models.

1 Introduction

Spatial models have been used in many fields of science where the data are collected in different locations, i.e. each observation is associated with a point or region in space. The data can be points when the exact location of the occurrence is known, or counts by area where the cases are aggregated in geographically defined areas. For a review of modelling techniques for spatially referenced data see for example Schmidt et. al. (2004).

In many applications the data are counts corresponding to areas which are administratively specified, thus leading to an artificial level of aggregation and the size of the population at risk may vary considerably among the areas under study. Consequently, the spatial variability might be artificial and should not be confused with real underlying heterogeneity (Richardson, 2003). Most works in

spatial modelling for count data define the model in the aggregate level already specified and this will be the approach here too. However, it is worth noting that inferences will be relevant for this level of aggregation only.

In particular, the analysis of spatial dispersion of the risk of occurrence of an event (e.g. disease, homicide, etc.) is usually done via maps of incidence rates, where a set of areas is shaded according to the values of a variable of interest. The objective of this mapping is to infer the geographic distribution of the rates thus identifying areas of higher or lower incidence. When the tones on the map tend to change smoothly we can call this feature a spatial trend (in analogy to time trend).

For a map divided into n contiguous areas let $\mathbf{y} = (y_1, \dots, y_n)$ and $\mathbf{e} = (e_1, \dots, e_n)$ where y_i is the number of occurrences of the event of interest in area i during the period of investigation, and e_i is the expected number of occurrences. For rare events a Poisson model is commonly adopted for each y_i , and if the event is noncontagious the numbers of cases are mutually independent. Each y_i has mean $e_i\psi_i$ where $\Psi = (\psi_1, \dots, \psi_n)$ are the relative risks, specific of each area and these are the parameters of interest. So, the adopted model is given by

$$Y_i|e_i, \psi_i \sim \text{Poisson}(e_i\psi_i), \quad i = 1, \dots, n.$$

The expected count e_i is assumed to be a known quantity which value is based on known risk factors. In the analysis in this work we assume that there are no confounding factors and the expected values are computed as $e_i = n_i \sum_j y_j / \sum_j n_j$ where n_i is the population in area i .

In the classical approach the maximum likelihood estimate of the relative risks is given by $\hat{\psi}_i = y_i/e_i$, with estimated standard error $s_i = \sqrt{y_i}/e_i$, and is usually called the standardised mortality rate (SMR) in the epidemiological literature. Several problems with this approach have already been identified in the literature. For example, more extreme values of these estimates (which visually dominate the map) may be based on a few cases only in areas with small population (see for example Molli, 1996).

Another problem is that rare events in small areas can lead to extra-Poisson variation, i.e. there is more heterogeneity in the population than is assumed by the Poisson model. Another characteristic commonly found in this context and that is not taken into account by the classical approach is the possibility of spatial correlation in the relative risks. Such correlation may be due for example to spatially correlated covariates and not included in the model.

This can be taken into account by allowing the relative risk to vary within each area and in this case the Bayesian approach is appropriate specifying prior distributions to the parameters ψ_i .

Example: Homicide Rates in Curitiba

Information was collected concerning the number of homicides in 2000 at each district of Curitiba city, Brazil. We note from the dispersion in Figure 1 that

there is a large variation in the homicide rates for small populations. In the map in Figure 2 the city of Curitiba was subdivided in its 75 districts existing in 2000. Maximum likelihood estimates (SMR) were calculated and the districts received different tones on the map according to the SMR value. As we can see, districts with larger populations received a positive evaluation relative to those with small populations, which can lead to wrong decisions concerning the homicide rates by district. It seems as if the homicide rates are being underestimated in some districts and it is also not possible to identify any sort of spatial pattern on the map.

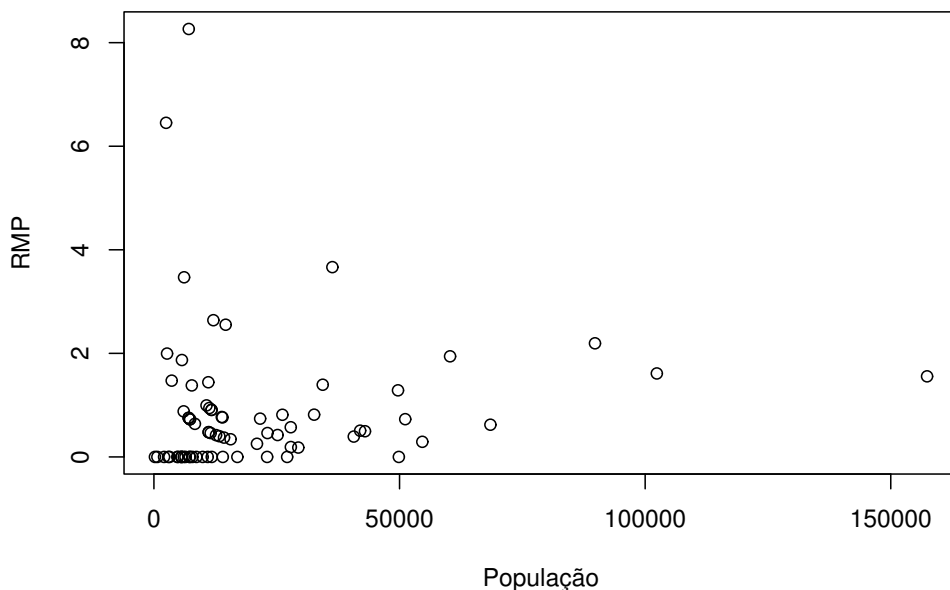


Figure 1: Relative risks obtained via maximum likelihood estimates (SMR) for the homicide numbers in districts of Curitiba city versus the districts population.

2 Hierarchical Models for Relative Risks

In the Bayesian approach, further to information in the data, i.e. the number of occurrences in each area, we need to specify a prior distribution $p(\Psi)$ for the relative risks which provides information concerning their variability along the map. Bayesian inference is then based on the combination of these two sources of information via the posterior distribution of the relative risks $p(\Psi|\mathbf{y})$ obtained via Bayes theorem. The prior distribution usually depends on hyperparameters γ so that the marginal posterior of Ψ is given by

$$p(\Psi|\mathbf{y}) = \int p(\Psi, \gamma | \mathbf{y})d\gamma. \quad (1)$$

Point estimates of the relative risks can be obtained via location measures of the distribution (1) while scale measures provide information on the uncertainty of these estimates. The integrals involved in the computation of these measures cannot in general be obtained analytically or even by numerical integration and approximation methods are necessary. See for example Clayton and Kaldor (1987) and Molli and Richardson (1991) for applications of the empirical Bayes method in the context of rare diseases in small areas.

2.1 Prior Specification

The prior distribution of the relative risks should be structured so as to accommodate the effect of factors measured at the level of each area as well as the possibility of spatial variation. One form of combining unstructured priors with information concerning the spatial structure proposed in the literature consists in modelling the logarithm of the relative risk as the sum of two independent components: the nonspatial random effect θ_i and the spatial random effect ϕ_i . Besag et al. (1991) propose to model the nonspatial components independently as $\theta_i \sim N(0, 1/\tau_\theta)$, describing the unstructured heterogeneity, and the spatial components so as to indicate that geographically close areas tend to present similar risks. One way of expressing this spatial structure is via Markov random fields models where the distribution of each ϕ_i given all the other elements $\{\phi_1, \dots, \phi_{i-1}, \phi_{i+1}, \dots, \phi_n\}$ depends only on its neighbourhood. A commonly used model is the Gaussian intrinsic conditional autoregression where the conditional distribution of each ϕ_i is given by

$$\phi_i|\phi_j, j \neq i \sim N\left(\frac{\sum_{j \in \delta_i} w_{ij}\phi_j}{\sum_{j \in \delta_i} w_{ij}}, \frac{1}{\tau_\phi \sum_{j \in \delta_i} w_{ij}}\right) \quad (2)$$

where δ_i represents the set of areas which are neighbour of area i . One important restriction in this specification is that the matrix of weights \mathbf{W} must be symmetric. It should be noted that the specification of this CAR structure leads to a prior joint distribution for the relative risks given by

$$p(\phi|\tau_\phi) \propto \frac{1}{\tau_\phi^n} \exp\left\{-\frac{1}{2\tau_\phi^2} \sum_{i=1}^n \sum_{j<i} w_{ij}(\phi_i - \phi_j)^2\right\}.$$

This prior is improper as it is based on paired differences between the ϕ_i 's. In practice, a constraint is imposed to these random effects in order to guarantee identifiability. Besag and Kooperberg (1995) showed that imposing a sum to zero constraint and specifying an intercept with location invariant prior $\text{Uniform}(-\infty, \infty)$ is equivalent to an unrestricted parameterisation without an intercept.

Although other possibilities exist, the simplest and most commonly used neighbourhood structure is defined by the existence of a common border of any

length between the areas. In this case, the weights w_{ij} are specified as $w_{ij} = 1$ if $j \in \delta_i$ and $w_{ij} = 0$ otherwise so that $\sum_{j \in \delta_i} w_{ij}$ is simply the number of neighbours of area i . So, the conditional prior mean of ϕ_i is given by the arithmetic average of the spatial effects from its neighbours and the conditional prior variance is proportional to the number of neighbours. This will also be the structure adopted in this work.

We also relate a set of covariates to the relative risks. The vector β of covariate coefficients will be specified from a multivariate normal prior distribution with mean zero and variance-covariance matrix $\Sigma = I\sigma_\beta^2$, i.e. the β_j 's are assumed independent *a priori*.

This class of models has been used in most recent works on disease mapping (see for example Best *et al.*, 1999). The use of information from other areas in the region under study should reduce the effect of random fluctuations not associated to the relative risk. Taking spatial correlation among neighbouring areas into account is expected to produce smoother and more informative maps.

2.2 Fully Bayesian Estimation

In the fully Bayesian approach the idea is to add another level in the model hierarchy by specifying a prior distribution for the hyperparameters γ . In This case the inference on the relative risks will be based on the marginal posterior distribution (1) which often cannot be obtained analytically. Analytical or numerical approximations are needed. In particular, Markov chain Monte Carlo methods (MCMC) will be employed to obtain a sample from the joint posterior distribution of (Ψ, γ) , automatically generating samples from the marginal posteriors of Ψ and γ (see for example, Gamerman 1997). Defining the parameter vectors $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ and $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)$, the joint posterior distribution of all parameters is expressed as

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}, \tau_\theta, \tau_\phi | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta} | \tau_\theta) p(\boldsymbol{\phi} | \tau_\phi) p(\tau_\theta) p(\tau_\phi)$$

thus taking into account a conditional independence structure. From a sample from this posterior we can obtain estimates of the relative risks via $\psi_i = \exp(\theta_i + \phi_i)$.

In the highest level of the hierarchy prior distributions are specified to the prior precisions τ_θ and τ_ϕ . The Gamma family of prior distributions is conditionally conjugate, i.e. the full posterior conditional distribution is also Gamma. This conditional conjugacy allows that τ_θ and τ_ϕ be easily updated in the algorithm used here. A common choice in the literature is the non-informative (proper) prior $\text{Gamma}(\epsilon, \epsilon)$ with small values for ϵ . However, this specification attributes low prior probability to small values of the standard deviation and consequently a spatial structure for example might be imposed *a priori*. Kelsall and Wakefield (1998) verified that the estimation of relative risks can be highly dependent of the choice of prior parameters and within a class of Gamma priors they suggest a $\text{Gamma}(0,5;0,0005)$ distribution as a sensible choice.

The posterior distribution is clearly analytically intractable. One method to obtain values from the joint posterior is via simulation of a sufficiently large number of dependent observations of the parameter vector as an ergodic Markov chain. In particular, the algorithm known as Gibbs sampler (Gilks, 1995) is useful in the context of Markov random fields where the joint posterior distribution is complicated but the full conditional posterior distributions have simple forms. This method is implemented here with the software WinBUGS (Spiegelhalter et. al. 1999), freely available on the Internet in the address <http://www.mrc-bsu.cam.ac.uk/bugs>. The commands used in this work are provided in the appendix.

The relative contribution of the spatial and nonspatial effects to the total overdispersion can be estimated from the posterior distribution of the quantity

$$\xi = \frac{sd_{\phi}}{sd_{\theta} + sd_{\phi}}$$

where sd_{θ} and sd_{ϕ} are empirical marginal standard deviations of θ and ϕ respectively, computed at each chain iteration (Best *et al*, 1999). If the estimate of ξ is close to 1 then the total variation is dominated by the spatial effects while a value close to zero indicates that the spatial variation is negligible.

2.3 Example: The Homicide Rates Revisited

In this section, the dataset concerning homicides in 2000 by district in Curitiba city was again analysed. The Bayesian hierarchical model as described above was estimated with a conditionally conjugate Gamma prior for the hiperparameters τ_{θ} and τ_{ϕ} . The variation in the relative risks along the districts was modelled according to the spatial prior given in (2), plus the effects of selected covariates. The covariates used in this study regard to social aspects of the city and are described below.

- Income: Median Income by district of Curitiba in 2000.
- Illiteracy: Percentage of people aged 15 years or over considered to be illiterate.
- Household: Percentage of households in areas considered to be at risk, or not in regular conditions.

The maps in in Figures 5, 6 and 7 were constructed with the values of these covariates.

The WinBugs software was then used to perform 50 000 simulations from the full conditional posterior distributions, from which the first 20 000 were discarded as burn-in. So, all the results here are based on a sample of 30 000 values. From this sample several characteristics of the posterior distribution may be estimated, and the main interest here is the spatial variation of the relative risks.

In order to compare and select the more appropriate model among those considered here we use the Deviance information criterion (DIC), where lower values indicate a good model fit relative to the number of parameters in the model (Spiegelhalter *et al.*, 2002). One advantage of the DIC is that it is easily computed during the simulation of the Markov chains. In Table 1 the computed values for the DIC are presented, from which we can see that the best model is the one that incorporates spatial and nonspatial random effects plus the covariate Illiteracy. We note however the similarity with the DIC for the model which also includes Income. Also, the DIC is subject to Monte Carlo sampling error since it is a function of stochastically simulated quantities. This might cast some doubt whether the inclusion of Income in the model is substantially improving model fit. One way round this problem, as suggested in Burnham and Anderson (1998) Section 4.2, is to use DIC weights obtained by subtracting from each DIC the value associated with the “best” model and then setting

$$w_k \propto \exp(-\Delta DIC(k)/2)$$

where $\Delta DIC(k)$ denotes the transformed DIC value for model k . The normalised weights (summing to 1) are shown in the third column of Table 1 and model comparison is much easier in this transformed scale.

Table 1: DIC values for each model obtained from WinBugs based on 30 000 simulations (lowest value in bold) and computed DIC normalised weights.

Model	DIC	weights
No spatial effect	274.551	0.0000
No covariates	261.888	0.0062
Illiteracy	252.448	0.6936
Household	292.265	0.0000
Income	259.717	0.0183
Income+Illiteracy	254.249	0.2819
Illiteracy+Household	285.940	0.0000
Income+Household	291.764	0.0000
Income+Illiteracy+Household	286.244	0.0000

The map in Figure 4 was constructed with the posterior means of the ψ_i as point estimates of the relative risks. We note that there were no big changes from the SMRs in areas with large populations, as we expected as these were good estimates of the relative risks. On the other hand, areas with small populations benefited from the information coming from neighbouring areas. We can also note that there is a risk pattern that concentrates in the South/South-West areas, then reduces in all directions and raises again in two central and two eastern areas. All in all, it is easier to identify areas in the city with similar relative risks.

3 Discussion

In this work we adopted a Bayesian approach to estimate relative risks of a rare event occurrence in small areas. The problem of overdispersion found in the usual classical estimation was tackled via specification of suitable priors. The method was illustrated with a real data example. Estimates of the posterior distribution were obtained via MCMC methods where inference is based on an approximate sample from the posterior distribution.

The Bayesian hierarchical model adopted is intrinsically spatial thus incorporating a component that captures the large scale smooth variation of the risk in the whole region under study. For the real data example, a set of covariates which might be of potential influence on the relative risks were included and tested via DIC.

There are various interesting extensions of the model adopted in this work that are currently being investigated by the authors. In the CAR prior, other weight structures describing the neighbourhood may be adopted, e.g. $w_{ij} = \exp(-u_{ij})$ where u_{ij} is the distance between areas i and j (Paez and Gamerman, 2003). Alternatively, the matrix of weights of neighbourhood structure might be defined in terms of the length of the border between two areas. In this case, rather than a binary structure the weight w_{ij} is equal to the length of the border between areas i and j so that the spatial influence from neighbouring areas increases or decreases (*a priori*) according to the length of the border. Another neighbourhood structure, suggested in Molli (1996) takes into account that natural borders (e.g. mountains) may influence the spatial analysis of the relative risks. Thus, the weights can be specified by subtracting the length of the natural border (Ferreira and Schmidt, 2004 compare these different specifications in the spatial modelling of Dengue cases in Rio de Janeiro). Trying to include other covariates, time structures and studying space-time interactions will also be the subject of future work.

Finally, an area still to be explored in more depth in the literature of spatial and space-time models is model comparison and selection via trans-dimensional MCMC algorithms. Recently, Green and Richardson (2002) and Fernandez and Green (2002) used mixture models for the relative risks where the number of components in the mixtures is variable. Surely the only sensible approach is to calculate posterior model probabilities and there is still considerable room for more research in this area.

References

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

- [2] Best, N. G., Arnold, R.A., Thomas, A., Waller, L.A. and Collon, E.M. (1999) Bayesian models for spatially correlated disease and exposure data. In *Bayesian Statistics 6*. Bernardo, J. M., Smith, A. F. M., Dawid, A. P. and Berger, J. O (Eds.) 131-156. Oxford University Press.
- [3] Clayton, D. and Kaldor, J. (1987) Empirical Bayes Estimates of age-standardised relative risks for use in disease mapping. *Biometrics*, **43**, 671–81.
- [4] Fernandez, C. and Green, P.J. (2002) Modelling spatially correlated data via mixtures: a Bayesian approach. *Journal of the Royal Statistical Society, Series B*, **64**, 805-826.
- [5] Gamerman, D. (1997). *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall Texts in Statistical Science Series. London: Chapman and Hall.
- [6] Gilks, W.R., Richardson, S. and Spiegelhalter, D. J. (1995) Introducing Markov chain Monte Carlo. In *Markov Chain Monte Carlo in Practice*. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.) 1-19. Chapman & Hall, London.
- [7] Green, P.J. and Richardson, S. (2002) *Hidden Markov Models and disease mapping*. *Journal of the American Statistical Association*, **97**, 1055-70.
- [8] Kelsall and Wakefield (1999) Modelling spatial variation in disease risk. *Technical Report*, Imperial College, London.
- [9] Molli, A. (1996) Bayesian Mapping of disease. In *Markov Chain Monte Carlo in Practice*. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.) 359-379. Chapman & Hall, London.
- [10] Molli, A. and Richardson, S. (1991) Empirical Bayes estimates of cancer mortality rates using spatial models. *Statistics and Medicine*, **10**, 95-112.
- [11] Paez and Gamerman (2003) *Study of the space-time in the concentration of airborne pollutants in the Metropolitan area of Rio de Janeiro*. *Environmetrics*, **14**, 387-408.
- [12] Richardson, S. (2003) Spatial models in epidemiological applications. In *Highly Structured Stochastic Systems*. P. Green, N.L. Hjort and S. Richardson (Eds.) 237–259. Oxford University Press.
- [13] Schmidt, A.M., Nobre, A.A. and Ferreira, G.S. (2004) *Alguns aspectos da modelagem de dados espacialmente referenciados*. *Revista Brasileira de Estatística*, (In print).
- [14] D.J. Spiegelhalter, A. Thomas and N. G. Best (1999). WinBUGS Version 1.3 User Manual, MRC Biostatistics Unit.

- [15] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van der Linde, A. (2002). *Bayesian Measures of Model Complexity and Fit (with discussion)*. Journal of the Royal Statistical Society, Series B, **64**, 1-34.

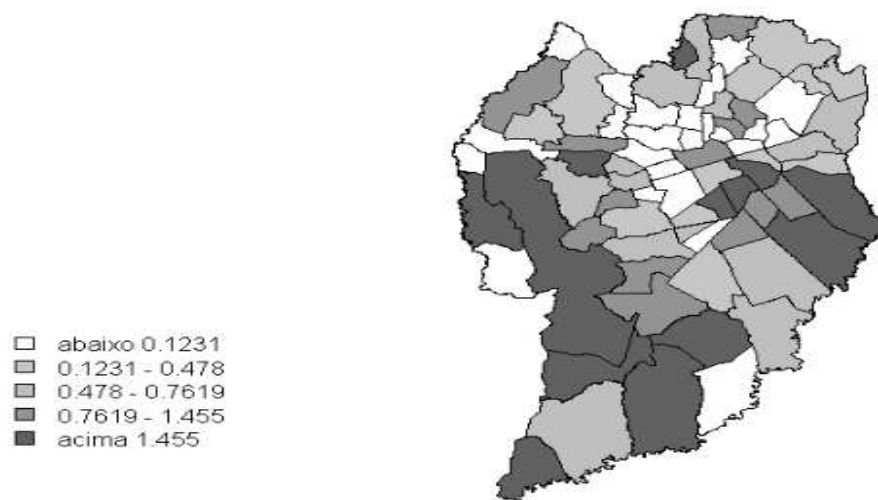


Figure 2: Relative risks obtained via maximum likelihood estimates (SMR) for the homicide numbers in districts of Curitiba city.

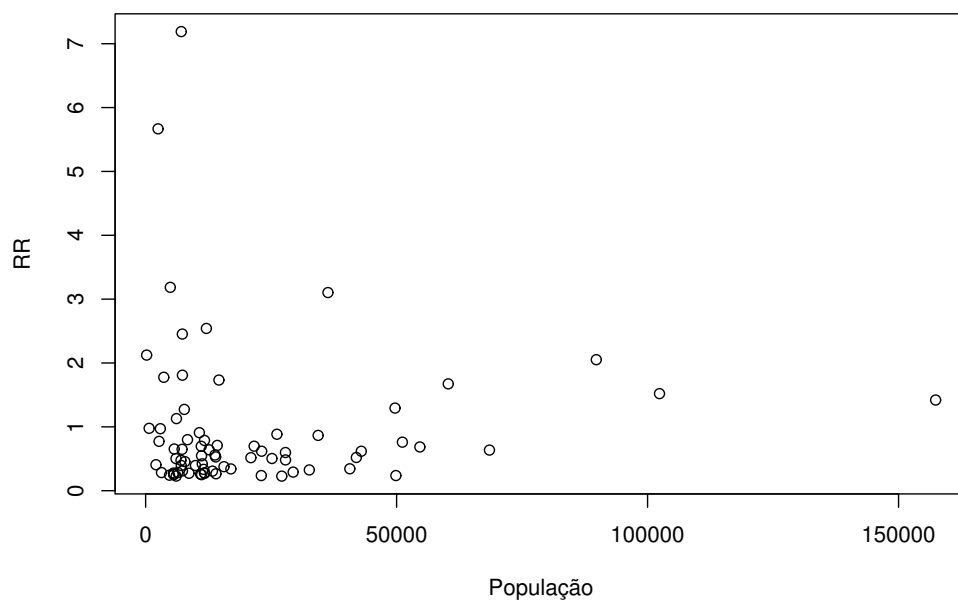


Figure 3: Relative risks estimates in the Bayesian hierarchical model versus district populations.

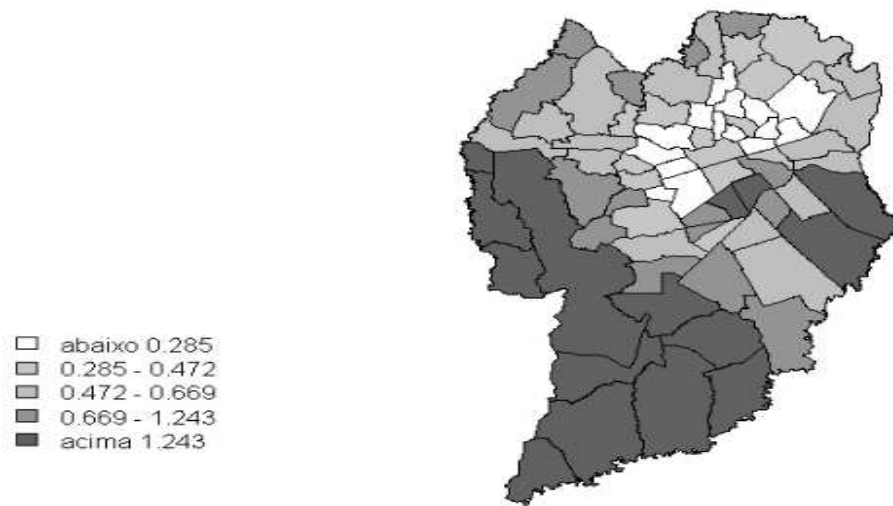


Figure 4: Relative risks estimates in the Bayesian hierarchical model with covariate Illiteracy included.

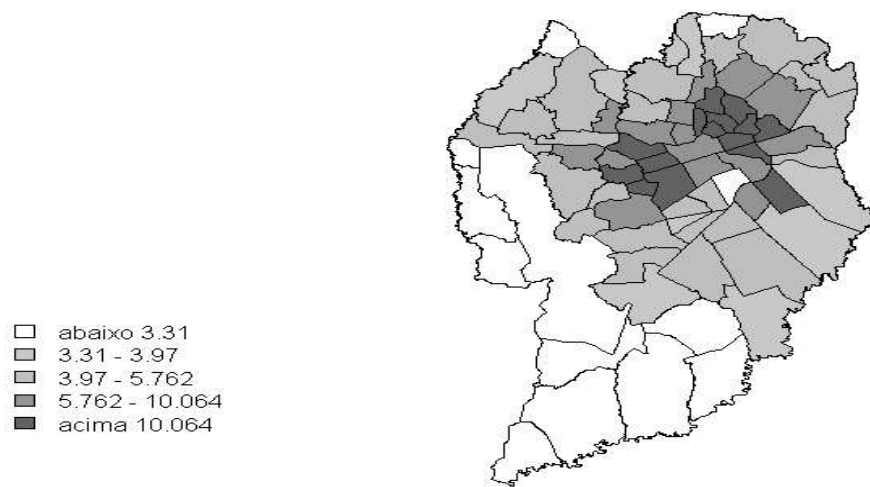


Figure 5: Median Income by district in Curitiba in 2000.

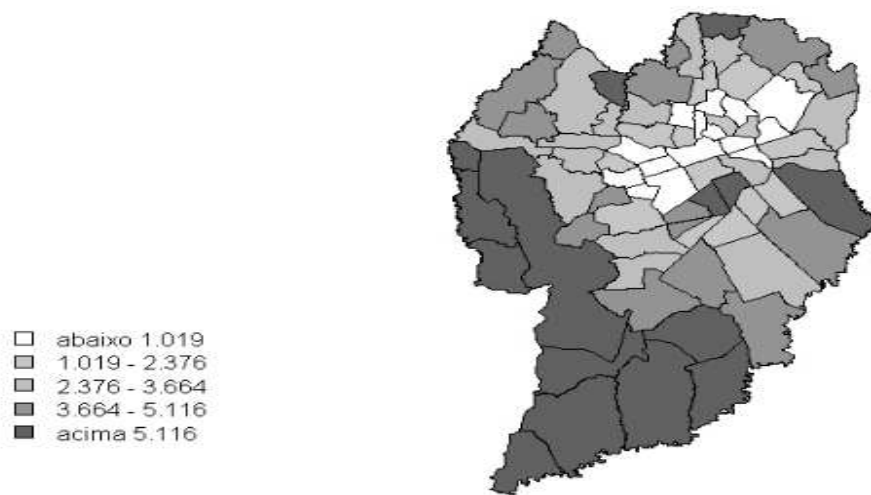


Figure 6: Percentage of people aged 15 year old or more considered to be illiterate by district in Curitiba in 2000.

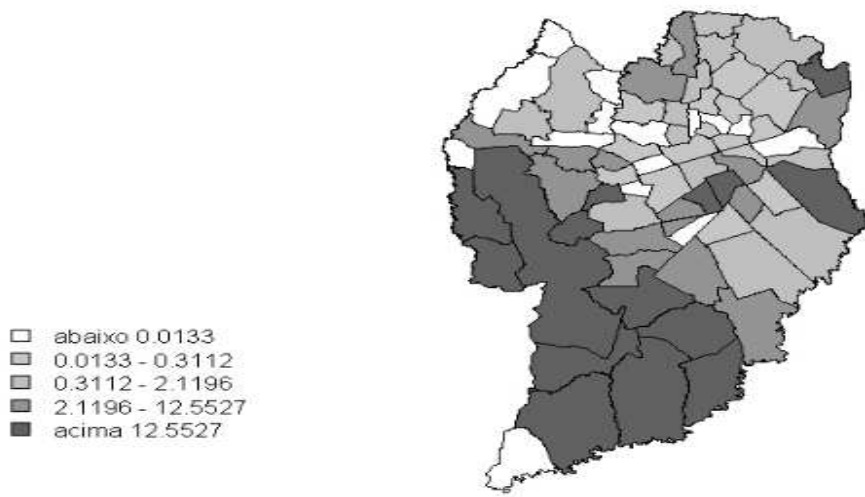


Figure 7: Percentage of households in areas considered to be at risk, or not in regular conditions by district in Curitiba in 2000.

A WinBUGS commands

```
model {
for(i in 1:N){
theta[i]~dnorm(0,tau.theta)
psi[i]<-beta+beta1*reg[i]+theta[i]+phi[i]
lambda[i]<-e[i]*exp(psi[i])
y[i]~dpois(lambda[i])
RR[i]<-exp(psi[i])
}

# CAR prior distribution for spatial random effects
phi[1:N] ~ car.normal (adj[], weights[], num[], tau.phi)
for (k in 1:sumNumNeigh){
weights[k] <- 1
}

# Other priors
beta ~ dflat()
beta1 ~ dnorm(0.0, 1.0E-5)
tau.phi ~ dgamma(0.5,0.0005)
tau.theta ~ dgamma(0.5,0.0005)

# relative contribution of random effects
psi<-sd(theta[])/(sd(phi[])+sd(theta[]))
}
```