

A More Robust Random Effects Model for Disease Mapping

Tonui Benard Cheruiyot^{1,2,*}, Mwalili Samuel^{2,3}, Wanjoya Anthony²

¹Department of Mathematics & Computer Science, University of Kabianga, Kericho, Kenya

²Department of Statistics & Actuarial Science, Jomo Kenyatta University of Science & Technology, Juja, Kenya

³Kenya Medical Research Institute/Centre for Disease Control, Nairobi, Kenya

Email address:

tonuib@kabianga.ac.ke (T. B. Cheruiyot), samuel.mwalili@gmail.com (M. Samuel), awanjoya@gmail.com (W. Anthony)

*Corresponding author

To cite this article:

Tonui Benard Cheruiyot, Mwalili Samuel, Wanjoya Anthony. A More Robust Random Effects Model for Disease Mapping. *American Journal of Theoretical and Applied Statistics*. Vol. 7, No. 1, 2018, pp. 29-34. doi: 10.11648/j.ajtas.20180701.14

Received: December 20, 2017; **Accepted:** January 8, 2018; **Published:** January 19, 2018

Abstract: Disease mapping studies have found wide applications within geographical epidemiology and public health and are typically analysed within a Bayesian hierarchical model formulation. The most popular disease mapping model is the Besag-York-Mollié model. A distinguishing feature of this model is the use of two sets of random effects: one spatially structured to model spatial autocorrelation and the other spatially unstructured to describe residual unstructured heterogeneity. Very often the spatially unstructured random effect is assumed to be normally distributed. Under practical situations, this normality assumption is found to be over restrictive. In this study, we investigate a more robust spatially unstructured random effect distribution by considering the Inverse Gaussian (IG) distribution in the disease mapping problem. The distribution has the normal distribution as special case. The inferences under this model are carried out within a Bayesian hierarchical model formulation implemented in WinBUGS. The IG distribution is introduced in WinBUGS using zero tricks. The usefulness of the proposed model is investigated with a simulation study and applied in real data; mapping HIV in Kenya. In this work we showed that the IG distribution can produce better results when the normality assumption is violated due to the skewness of the data. For the case of data in which the random effects are truly normal, the IG distribution adjusts to a normal distribution as dictated by the data itself. On the other hand, the spatially structured random effect is normally modelled using the intrinsic conditional autoregressive (iCAR) prior. This prior is improper and has the undesirable large scale property of leading to a negative pairwise correlation for regions located further apart. In addition, the BYM model presents some identifiability problems of the spatial and non-spatial effects. In this work, we consider Leroux CAR (named ICAR hereafter) prior, a less widely used prior in disease mapping, as the prior distribution for the spatially structured random effects.

Keywords: Disease Mapping, Spatial Analysis, Random Effects, Bayesian Analysis, Markov Chain Monte Carlo, Inverse Gaussian Distribution

1. Introduction

Disease mapping is the study of the geographical or spatial distribution of health outcomes. Over the past few decades and with the advent of computational methods and statistical methodology, and availability of spatially-referenced data and software tools, disease mapping has increased in popularity in epidemiological research [1, 2]. Particularly for modeling rare diseases, it is critical to consider spatial variabilities in order to reliably conduct inference or prediction about the process under study. In disease mapping, usually the object of analysis is to provide (estimate) the true

relative risk of a disease of interest across a geographical study area (map). Disease mapping is useful for several purposes such as health services resource allocation, disease atlas construction and in formulation of hypotheses about disease aetiology. Several statistical reviews on disease mapping have been done [3, 4, 5, 6, 7].

Most of the work in spatial disease mapping considers the Besag-York-Mollié model (BYM) proposed by [8] which includes two spatial effects: one assuming a Gaussian exchangeable prior to model unstructured heterogeneity and another one assuming an iCAR prior for the spatially structured variability. Assumption of normality on the

uncorrelated random effect in models is common. However, under practical situations, this normality assumption can be incorrect because some random effects can be skewed violating this general normality assumption, see [9, 10]. In this study, we investigate a more robust spatially unstructured random effect distribution by considering the Inverse Gaussian (IG) distribution in the disease mapping problem. The IG distribution is a member of the exponential family and it offers a convenient modeling for positive right skewed data. It has two parameters, the location parameter, μ and the scale parameter λ . The normal distribution is obtained as a limiting case for the IG distribution. On the other hand, the iCAR prior that is usually used for modeling spatially structured random effect is improper and has the undesirable large scale property of leading to a negative pairwise correlation for regions located further apart [11, 12]. In addition, the spatial and non-spatial effects in the BYM convolution model are not identifiable from the data [13, 14]. In this work, we also consider the Leroux CAR prior proposed by [15] as the prior distribution for the spatially structured random effects. This prior has been shown to outperform the iCAR prior [16].

This paper is organized as follows. In section 2, a review of the BYM and Leroux CAR models is given. Section 3 focuses on the IG distribution and discusses its limiting distribution, while a simulation to study the effect of misspecifying the random effects is conducted in Sections 4. In section 5, the models discussed are used to analyse the HIV data from 2012 Kenya Aids Indicator Survey (KAIS). In

section 6, we present discussions and finally conclusion is given in section 7.

2. Review of the BYM and the Leroux CAR Models

2.1. BYM Model

Consider a study region subdivided into n contiguous areas. Let Y_i and E_i denote, respectively, the observed and expected counts for a disease in the i th area ($i = 1, \dots, n$). Expected counts for each area can be obtained by applying a standard table of group-specific sex and age rates to each area-specific background population, subdivided by age and sex. The Besag-York-Mollié (BYM) model has the following set of assumptions:

$$\begin{aligned} Y_i &\sim \text{Poisson}(E_i \theta_i) \\ \log(\theta_i) &= \mu_i + v_i + u_i \end{aligned} \quad (1)$$

in which $\mu_i = x_i' \beta$ describes the effect of area-level covariates x_i , whereas vectors $v = (v_1, \dots, v_n)$ and $u = (u_1, \dots, u_n)$ represents spatially unstructured and structured random effects respectively. The implied prior on $b = v + u$ is termed a ‘convolution prior’ since it is the sum of two independent components. Specifically, the following priors are assumed:

$$v | \tau_v \propto \tau_v^{\frac{n-1}{2}} \exp\left(-\frac{\tau_v}{2} \sum_{i=1}^n v_i^2\right) \quad (2)$$

$$u | \tau_u \propto \tau_u^{\frac{n-1}{2}} \exp\left(-\frac{\tau_u}{2} \sum_{j \sim i} (u_i - u_j)^2\right) = \tau_u^{\frac{n-1}{2}} \exp\left(-\frac{\tau_u}{2} u' Q u\right) \quad (3)$$

where $j \sim i$ denotes the set of all unordered pairs of neighbours, i.e., regions sharing a common border, and hence the sum over all such sets can be written using a precision matrix Q . Here v is a zero mean white noise Gaussian process with precision τ_v ; i.e $v \sim N(0, \tau_v^{-1} I)$ and u is a first order intrinsic Gaussian Markov random field [17] with precision τ_u ; i.e $u \sim N(0, \tau_u^{-1} Q^-)$ where Q^- denotes the generalized inverse of Q . The resulting covariance matrix of b is

$$\text{Var}(b | \tau_v, \tau_u) = \tau_v^{-1} I + \tau_u^{-1} Q^- \quad (4)$$

The full conditional distributions of u_i given all the remaining components $u_{-i} = (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n)$ can be expressed as follows:

$$u_i | u_{-i}, \tau_u \sim \text{Normal}\left(\frac{1}{n_i} \sum_{j \sim i} u_j, \frac{1}{n_i \tau_u}\right) \quad (5)$$

This conditional distribution for u is called the intrinsic conditional autoregressive (iCAR) prior distribution.

2.2. Leroux CAR Model

In the BYM model, the structured and unstructured components cannot be seen independently from each other, and are thus not identifiable [13, 14]. That is, each data point is represented by two random effects but only their

sum $v_i + u_i$ is identifiable. An additional challenge, which makes the choice of hyperpriors more difficult, is that v_i and u_i do not represent variability on the same level. While τ_v^{-1} can be interpreted as the marginal variance of the unstructured random effects, τ_u^{-1} controls the variability of the structured component u_i , conditional on the effects in its neighbouring areas [7, 18]. [15] proposed an alternative model formulation to make the compromise between unstructured and structured variation more explicit. Here, b is assumed to follow a normal distribution with mean zero and covariance matrix

$$\text{Var}(b | \tau_b, \rho) = \tau_b^{-1} [(1 - \rho)I + \rho Q]^{-1} \quad (6)$$

where $\rho \in [0, 1]$ denotes a mixing parameter. The model reduces to a model with independent (exchangeable) random effects if $\rho = 0$ and to the iCAR model when $\rho = 1$, see [16, 14, 11].

The univariate full conditional distribution of u_i is given by:

$$u_i | u_{-i}, \tau_u \sim \text{Normal}\left(\frac{\rho}{(1-\rho) + n_i \rho} \sum_{j \sim i} u_j, \frac{1}{[(1-\rho) + n_i \rho] \tau_u}\right) \quad (7)$$

The basic idea of these models is that, conditional upon its neighbours, u_i is independent of all u_j s at non-neighbouring areas. A neighbouring structure needs to be

specified. Many structures have been proposed in the literature. In this study, we follow the most popular approach: areas are considered to be neighbours if they share a boundary, and information about neighbouring is summarised in the $n \times n$ matrix W , whose entries $w_{ij} = 1$ if areas i and j share a boundary and 0 otherwise ($w_{ii} = 1$ completes the specification).

The posterior distribution can be sampled using MCMC algorithms such as the Gibbs or Metropolis-Hastings samplers. In this study, we use a Gibbs sampler as conditional distributions for the parameters were available in that formulation.

3. The Inverse Gaussian Distribution

Several authors, [9, 7, 8, 19], have suggested that it is possible to replace the normality assumption of the spatially unstructured random effect with other choices such as the Laplace distribution or the Student's t distribution. For instance, [9] used generalized Gaussian distribution. In this work we explore the use of the IG distribution as a candidate for the unstructured random effects.

Definition: A random variable Y is said to have an IG distribution if its probability density function is given by:

$$f(y|\mu, \lambda) = \left(\frac{\lambda}{2\pi y^3}\right)^2 \exp\left(\frac{-\lambda(y-\mu)^2}{2\mu^2 y}\right), y > 0, \mu \in \mathbb{R}, \lambda > 0 \quad (8)$$

The mean and the variance of this parametrized IG distribution is given by $E(Y) = \mu$ and $Var(Y) = \mu^3/\lambda$. In GLMs, the parametrization $\mu, \sigma^2 = 1/\lambda$ is normally used.

The normal (Gaussian) distribution is obtained as a limiting case of IG distribution when $\lambda \rightarrow \infty$ (or $\sigma^2 \rightarrow 0$). See [20] for further discussions on statistical properties of this distribution.

4. Simulation

In this section, we carry out a simulation study to determine the effect of wrongly specifying the distribution of the random effect in a BYM model. Two scenarios were considered in the simulation. In the first simulation, the dataset is generated through a random effect v which assumes IG distribution as follows: Assuming that there are 50 geographical regions and O_i is the number of disease counts observed in region i and E_i is the corresponding

expected counts in that region. Without loss of generality, we further assume that no covariates are available for use.

Step 1: Generate 50 values of $v \sim IGD(0.5, 1)$.

Step 2: Set $E = 30$ for all the regions.

Step 3: Calculate the relative risk $\theta = \log(v)$

Step 4: Calculate $O = E \times \theta$

Step 5: Generate the observed counts as

$$O \sim \text{Poisson}(50, \text{mean} = \lambda)$$

In the second scenario, the procedure above is repeated but changing the random effect generation mechanism in step 1, as $v \sim N(0.5, 0.125)$. We fitted two Bayesian hierarchical models for the data set. The models were specified based on different assumptions on the random effects as follows:

$$\begin{aligned} Y_i &\sim \text{Poisson}(E_i \theta_i) \\ \log(\theta_i) &= \log(E) + v_i \end{aligned} \quad (9)$$

with

a. *Model a:* $v_i \sim IGD(0.5, \lambda)$

b. *Model b:* $v_i \sim N(0.5, \sigma^2)$

The estimated relative risk $\hat{\theta}_i = \log(v_i)$. The simulation steps above were repeated $m = 1000$ times. To compare the two models, we calculated the mean squared error (MSE), for each model, using the formula:

$$MSE = \frac{1}{1000} \sum_{j=1}^{1000} \frac{1}{50} \sum_{i=1}^{50} \frac{1}{50} (\hat{\theta}_{ij} - \theta_{ij})^2$$

The model with a small MSE provides the best fit.

In the second scenario, the procedure above is repeated but changing the random effect generation mechanism in step 1, as $v \sim N(0.5, 0.125)$. From the simulation results, the inverse gaussian random effect is seen to perform well even in cases where the random effect strictly follows a normal distribution. In Table 1, the IG distribution produces lower mean squared error values as compared to the normal distribution in both cases. When the random effects are generated using IG distribution, the loss in efficiency incurred for using normal random effects is 4.7%. When the random effects are generated using normal distribution, the IGD still has a lower MSE compared to the normal counterpart. The loss in efficiency for this case is low at 2.1%.

Table 1. Model comparison under simulation.

Generating distribution	Parameters used	Model used	MSE	% loss in Efficiency
Inverse Gaussian	$\mu = 0.5, \lambda = 1$	Normal	0.09836	4.7
		IGD	0.09395	0
Normal	$\mu = 0.5, \sigma^2 = 0.125$	Normal	0.05135	2.1
		IGD	0.05028	0

5. Application to HIV Data

In this section we apply the model to HIV data collected by the Ministry of Health, Kenya. The data was extracted

from the 2012 Kenya Aids Indicator Survey (KAIS), conducted by the Government of Kenya. The main objective of survey was to collect high quality data on the prevalence of HIV and sexually transmitted infections (STI) among adults, and to assess knowledge of HIV and STI in the

populations. The survey collected a representative sample of households selected from the eight provinces in the country. In this study, we focus only on HIV cases among adults, that is, men and women in the age of 15-64 years. Two questionnaires were used in the survey. The first one is a household questionnaire which collected information about the household head and the characteristics of the dwelling place. The second one, the individual questionnaire, collected information from men and women aged 15-64 years, about their demographic characteristics, and their knowledge on

HIV and STI. All women and men aged 15-64 years in selected households who were either usual residents or visitors present the night before the survey were eligible to participate in the individual interview and blood draw, provided they gave informed consent.

The following eight models were fitted:

$$Y_i \sim \text{Poisson}(\mu_i)$$

with

1. Model 1: $\log(\mu_i) = \alpha + \log(E_i) + v_i; v_i \sim N(0, \sigma_v^2)$
2. Model 2: $\log(\mu_i) = \alpha + \log(E_i) + v_i; v_i \sim \text{IGD}(0.5, \lambda)$
3. Model 3: $\log(\mu_i) = \alpha + \log(E_i) + u_i; u_i \sim \text{iCAR}(\sigma_u^2)$
4. Model 4: $\log(\mu_i) = \alpha + \log(E_i) + u_i; u_i \sim \text{lCAR}(\sigma_u^2)$
5. Model 5: $\log(\mu_i) = \alpha + \log(E_i) + v_i + u_i; v_i \sim N(0, \sigma_v^2), u_i \sim \text{iCAR}(\sigma_u^2)$
6. Model 6: $\log(\mu_i) = \alpha + \log(E_i) + v_i + u_i; v_i \sim N(0, \sigma_v^2), u_i \sim \text{lCAR}(\sigma_u^2)$
7. Model 7: $\log(\mu_i) = \alpha + \log(E_i) + v_i + u_i; v_i \sim \text{IGD}(0.5, \lambda), u_i \sim \text{iCAR}(\sigma_u^2)$
8. Model 8: $\log(\mu_i) = \alpha + \log(E_i) + v_i + u_i; v_i \sim \text{IGD}(0.5, \lambda), u_i \sim \text{lCAR}(\sigma_u^2)$

where Y_i and E_i denote, respectively, the observed and expected cases of HIV in the i th county ($i = 1, \dots, 47$). Model estimation was carried out using a Bayesian approach. All parameters in the models were assigned prior distributions. In this analysis, a non-informative normal prior was assigned to the fixed effect coefficient α . The shape parameter λ was given a gamma prior distribution, and the variance parameters were assigned inverse gamma distributions. The models were implemented using WinBUGS version 1.4 [21, 22]. For each model, 6,000 Markov chain Monte Carlo (MCMC) iterations were ran, with the initial 4,000 discarded to cater for the burn-in. The 2,000 iterations left were used for assessing convergence of the MCMC and parameter estimation. We monitor MCMC convergence using trace plots, see [23]. For

model comparison, we adopt the deviance information criterion (DIC) proposed by [24]. DIC is defined as $\text{DIC} = D + pD$, where D is a measure of current model fit and pD is a penalty for model complexity. The best fitting model is one with the smallest DIC value. The overall loss across the data was assessed by the use of the Mean Squared Predictive Error (MSPE), which is an average of the item-wise squared error loss, $\text{MSPE} = \sum_i \sum_j (y_i - y_{ij}^{pr})^2 / (G \times m)$ with y_{ij}^{pr} being the predictive data item at iteration j , m being the number of observations and G is the sampler sample size. The best model for prediction is the one with the lowest MSPE value. The results for this analysis are given in Table 2 below.

Table 2. Model comparison in mapping HIV in Kenya.

Model	α	σ_v^2	σ_u^2	λ	ρ	pD	DIC	MSPE
Model 1	-0.09 (-0.09, -0.09)	0.86 (0.86, 0.86)	-	-	-	75.30	693.13	50590
Model 2	-2.94 (-2.96, -2.92)	0.28 (0.00, 3.69)	-	0.12 (0.10, 0.15)	-	46.61	637.16	50320
Model 3	-0.21 (-0.21, -0.21)	-	1.24 (1.11, 1.37)	-	-	133.23	928.30	76130
Model 4	0.02 (-0.01, 0.05)	-	1.03 (0.91, 1.15)	-	0.73 (0.60, 0.87)	46.97	636.48	50610
Model 5	-0.23 (-0.25, -0.20)	0.24 (0.20, 0.28)	1.22 (1.07, 1.36)	-	-	67.05	676.60	50490
Model 6	0.09 (0.06, 0.11)	0.27 (0.21, 0.29)	0.95 (0.84, 1.06)	-	0.75 (0.62, 0.88)	46.95	636.44	50880
Model 7	-0.90 (-0.97, -0.84)	0.48 (0.00, 11.28)	0.79 (0.68, 0.91)	0.67 (0.50, 0.84)	-	54.23	651.03	58030
Model 8	-1.00 (-1.02, 0.99)	0.24 (0.00, 2.59)	0.77 (0.66, 0.89)	0.57 (0.40, 0.74)	0.66 (0.50, 0.82)	46.73	636.03	50190

From table 2 above, it can be seen that models whose unstructured random effects follow IG distribution have quite small DIC values in comparison to the models with normally distributed unstructured random effects. Similarly models whose structured random effects are modelled with ICAR prior distribution have smaller DIC values as compared to models with the iCAR distributed structured random effects. This confirms that the IG and lCAR prior models produce better results than the popular normal and iCAR prior

models. In particular, the IG convolution model (model 8) provides the smallest DIC value. Thus, Model 8 is the best model in terms of goodness-of-fit measures. It also has the lowest MSPE value as compared to the other models indicating that it has a good predictive behaviour. Figure 1 shows the spatial distribution of HIV in Kenya in 2016 based on the best fitting model (model 8). This is a map of relative risk and its corresponding credible interval.

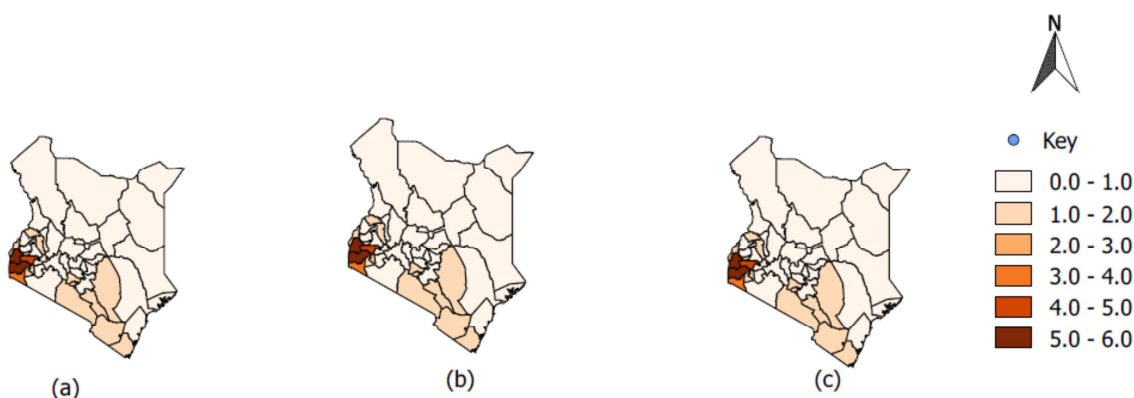


Figure 1. HIV relative risk map (a) and the corresponding 95% lower (b) and upper (c) credible limits maps, respectively, produced by model 8.

6. Discussion

Correlated data are usually modelled using generalized linear mixed effects model through the inclusion of subject-specific, random effects. Very often such random effects are assumed to be normally distributed. The normality assumption for the random effects has been both criticised and supported by several authors [25, 26, 27].

A similar situation arises in disease mapping context. The most popular model is the BYM model. This model has two sets of random effects: one which is spatially unstructured and the other which is spatially structured. The spatially unstructured random effect is usually assumed to have a Gaussian exchangeable prior distribution. However, this normality assumption can be incorrect and too restrictive because some random effects can be skewed violating this general normality assumption, see [9, 10]. The spatially structured component is usually modelled by an iCAR prior distribution. However, this prior has been shown to produce negative correlations for regions located further apart [11, 12]. In addition, the variance components in the BYM convolution model are not identifiable from the data; see [13].

This study has revealed that more robust and flexible distributions for the random effects can be used to improve the model fit in disease mapping models. In particular, we have considered IG distribution for the spatially unstructured random effects. This distribution contains the normal as a special case. For the spatially structured random effects, we propose Leroux CAR prior [15] which is a less widely used prior in disease mapping.

7. Conclusion

In this paper, we have shown that models whose unstructured random effects follow IG distribution perform better than those with normally distributed unstructured random effects. Similarly models whose structured random effects are modelled with the Leroux CAR prior distribution perform better than those models with the iCAR distributed structured random effects.

We adopt a bayesian modeling approach and assign prior distributions to all model parameters. This approach offers

several potential advantages over classical (e.g. maximum likelihood) estimation procedures. First, bayesian inference allows us to express uncertainty about model parameters through prior distributions. Second, by incorporating recent developments in McMC methods [28], including Gibbs sampling, bayesian models provide a flexible way to handle complex non-linear regressions such as ours. This was implemented easily within WinBUGS. Since the IG distribution is not a standard distribution in the WinBUGS software, it was introduced in the software using zero tricks. We assessed McMC convergence of all models parameters by checking trace plots and autocorrelation plots of the McMC output. The models were compared using the DIC as suggested by [24]. In addition, the overall loss across the data was assessed by the use of the Mean Squared Predictive Error (MSPE).

The models were compared using a simulation study and a real data set. In the simulation study, it was noted that the effect of misspecification of the random effects when the normal distribution is used in place of the IG distribution was high as compared to using the IG distribution in place of the normal distribution. IG distribution is often used as alternative to the normal distribution because of the similarities between the inference methods for these distributions. In fact the normal distribution is a special case of the IG distribution. When the random effect distribution fails to adhere to the normality assumption due to skewness, the IG distribution plays a big role in capturing this, something that the normal distribution cannot.

In the real data sets comparison, it can be seen that models whose unstructured random effects follow IG distribution generally perform better than models with normally distributed unstructured random effects. Similarly, models whose spatially structured random effects are modelled with Leroux CAR prior distribution perform better than those modelled with the iCAR prior. The IG convolution model (model 8) is the best fitting model. This model was used to produce the county relative risk maps of HIV in Kenya for the year 2016. Disease maps play a key role in descriptive spatial epidemiology. Maps are useful for several purposes such as identification of areas with suspected elevations in risk, formulation of hypotheses about disease aetiology, and assessing needs for health care resource allocation.

References

- [1] Lawson, AB. (2013). Bayesian disease mapping: Hierarchical modelling in spatial Epidemiology. *Chapman and Hall/CRC press*.
- [2] Elliott, P. and Wartenberg, D. (2004). Spatial epidemiology: Current Approaches and Future Challenges. *Environmental Health Perspectives*; 112, 998-1006.
- [3] Clayton D. and Bernardinelli L. (1992). Bayesian methods for mapping disease risk: Geographical and environmental epidemiology, methods for small-area studies. *Oxford University Press, Oxford*.
- [4] Bithell J. (2000). A classification of disease mapping methods. *Statistics in Medicine*; 19: 2203-2215.
- [5] Diggle PJ. (2000). Overview of statistical methods for disease mapping and its relationship to cluster detection. In: *Spatial Epidemiology: Methods and Applications*; Edited by: Elliot P, Wakefield JC, Best NG, Briggs DJ. *Oxford, Oxford University Press*; 87-103.
- [6] Lawson AB. (2001). Tutorial in Biostatistics: Disease map reconstruction. *Statistics in Medicine*; 20: 2183-2203.
- [7] Wakefield J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*; 8 (2): 158-183.
- [8] Besag J, York J and Mollié A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Ann Inst Stat Math.*, 43, 1-59.
- [9] Ngesa O, Achia T and Mwambi H. (2014). A Flexible Random Effects Distribution in Disease Mapping Models. *South African Statistical Journal*; 48 (1), 83-93.
- [10] Box G. and Tiao G. (1973). Bayesian inference in statistical analysis. *Addison-Wesley Pub. Co*.
- [11] MacNab YC. (2011). On Gaussian Markov random fields and Bayesian disease mapping. *Statistical Methods in Medical Research*; 20: 49-68.
- [12] Botella-Rocamora P, Lo'pez-Quílez A and Martínez-Beneito MA. (2013). Spatial Moving Average Risk Smoothing. *Statistics in Medicine*; 32: 2595-2612.
- [13] MacNab YC. (2014). On identification in Bayesian disease mapping and ecological-spatial regression models. *Statistical Methods in Medical Research*; 23 (2) 134-155; Originally published online 8 May 2012 DOI: 10.1177/0962280212447152. Available at: <http://smm.sagepub.com/content/23/2/134>.
- [14] Rampaso RC, de Souza ADP & Flores EF (2016). Bayesian analysis of spatial data using different variance and neighbourhood structures, *Journal of Statistical Computation and Simulation*, 86:3, 535-552, DOI: 10.1080/00949655.2015.1022549.
- [15] Leroux BG, Lei X and Breslow N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment and clinical trials*. New York: Springer-Verlag; pp.135-178.
- [16] Lee, D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology*; 2 (2): 79-89.
- [17] Rue H and Held L. (2005). Gaussian Markov random ...fields: theory and applications. *CRC Press*.
- [18] Bernardinelli L, Clayton D and Montomoli C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine*; 14 (21-22): 2411-2431.
- [19] Best N., Thomas A., Waller L., Conlon E., and Arnold R. (1999). Bayesian models for spatially correlated disease and exposure data. In *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting, volume 6*. *Oxford University Press, USA*, pp. 131-156.
- [20] Sato S. and Inoue J. (1994). Inverse Gaussian Distribution and Its Application. *Electronics and Communications in Japan*; 77 (1), 32-42.
- [21] Spiegelhalter DJ, Thomas A, Best NG, et al. (2007). WinBUGS version 1.4 user's manual. *MRC Biostatistics unit*.
- [22] Ntzoufras I. (2011). Bayesian modeling using WinBUGS, volume 698. *Wiley*.
- [23] Gelman A, Carlin JB., Stern HS. and Rubin DB. (2004). Bayesian Data Analysis; 2nd edn. *Boca Raton: Chapman and Hall*.
- [24] Spiegelhalter DJ, Best NG, Carlin BP, et al. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*; 64 (4): 583-639.
- [25] Litière S., Alonso A., and Molenberghs G. (2007). Type i and type ii error under random effects misspecification in generalized linear mixed models. *Biometrics*, 63 (4), 1038-1044.
- [26] Litière S., Alonso A., and Molenberghs G. (2008). The impact of a misspecified random effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in medicine*, 27 (16), 3125-3144.
- [27] McCulloch CE and Neuhaus JM. (2011). Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical Science*; 26 (3), 388-402.
- [28] Gelfand, AE. and Smith, AFM. (1990). Sampling-based approaches to calculating marginal densities. *Journal of America Statitistical Association*; 85, 398-409.