

Bayesian multi-scale modeling for aggregated disease mapping data

Peer-reviewed author version

AREGAY, Mehreteab; LAWSON, Andrew; FAES, Christel & Kirby, R.S. (2017)

Bayesian multi-scale modeling for aggregated disease mapping data. In: Statistical methods in medical research 26(6), p. 2726-2742.

DOI: 10.1177/0962280215607546

Handle: <http://hdl.handle.net/1942/21026>



Published in final edited form as:

Stat Methods Med Res. 2017 December ; 26(6): 2726–2742. doi:10.1177/0962280215607546.

Bayesian multi-scale modeling for aggregated disease mapping data

Mehreteab Aregay¹, Andrew B Lawson¹, Christel Faes², and Russell S Kirby³

¹Department of Public Health Sciences, Division of Biostatistics and Bioinformatics, MUSC, Charleston, USA

²Interuniversity Institute for Biostatistics, statistical Bioinformatics, Hasselt University, Hasselt, Belgium

³Department of Community and Family Health, University of South Florida, Tampa, FL, USA

Abstract

In disease mapping, a scale effect due to an aggregation of data from a finer resolution level to a coarser level is a common phenomenon. This article addresses this issue using a hierarchical Bayesian modeling framework. We propose four different multiscale models. The first two models use a shared random effect that the finer level inherits from the coarser level. The third model assumes two independent convolution models at the finer and coarser levels. The fourth model applies a convolution model at the finer level, but the relative risk at the coarser level is obtained by aggregating the estimates at the finer level. We compare the models using the deviance information criterion (DIC) and Watanabe-Akaike information criterion (WAIC) that are applied to real and simulated data. The results indicate that the models with shared random effects outperform the other models on a range of criteria.

Keywords

Deviance information criterion; Watanabe-Akaike information criterion; predictive accuracy; shared random effect model; scaling effect

1 Introduction

In spatial epidemiology, an important goal is to study the spatial distribution of diseases. Often, public health workers are interested in identifying areas which have a higher risk for a certain disease so that resources can be allocated accordingly. Spatial epidemiology can help us to examine such risk by taking into account population variation. Several authors have studied the relative risk using a standardized mortality/morbidity ratio (SMR), which is defined as the ratio of observed disease incidence relative to what is expected under standard

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

Corresponding author: Mehreteab Aregay, Department of Public Health Sciences, Division of Biostatistics and Bioinformatics, MUSC, 135 Cannon Street Suite 303, MSC 835 Charleston, SC 29425-8350, USA. aregay@musc.edu.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

conditions.¹ However, this is a crude approach and SMRs can provide unstable estimates due to their ratio form. Moreover, they do not accommodate the correlation between neighbors. To overcome this limitation, most notably, Besag et al.² proposed a convolution model (BYM) that models the relative risks as a function of spatially structured and unstructured random effects.

Convolution models have been widely used in disease mapping.³ Besag et al.² first considered two additive components in the log of the relative risk: a correlated and an uncorrelated component. The correlated component (CH) was assumed to have an intrinsic conditional autoregressive (ICAR) prior distribution. On the other hand, the uncorrelated component (UH) was assumed to have a zero mean Gaussian distribution. Alternative correlation models are available. For example, a proper CAR model has been proposed for the CH component, corresponding to a proper multivariate Gaussian distribution with a full rank covariance matrix.^{4,5} On the other hand, Moraga and Lawson⁶ considered Gaussian component mixture models (GCM) and compared it to the ICAR assumption in a simulation study. Even though the convolution model has been widely used for spatial epidemiology data, it does not accommodate a spatial scaling effect associated with aggregation in the data space.

A spatial scaling effect is induced from a lower geographical level (finer) to a higher geographical level (coarser) by smoothing out variation (i.e. averaging).⁷ Modeling scale effects is of special interest in disease mapping. Scaling reduces the variability of the data and makes them more alike, and hence the results from the finer data may not be consistent with the results from the coarser data. For example, we may find a relationship between an outcome and a covariate at the higher level, which is not retained at the lower level. On the other hand, an atomistic fallacy will occur if we try to generalize directly from the lower level (e.g. census tract) to the higher aggregation level (e.g. county) without adjusting for the variability of individuals' response to the diseases. In general, these scale change issues are called the modifiable areal unit problem^{3,7} or change of support problem in geostatistics.⁸⁻⁹ In particular, Wong⁷ discussed in detail the two sub-problems (zoning and scale problem) of MAUP. Here, we deal with the scale problem.

To encompass scaling effects, Kolaczyk and Haug¹⁰ proposed a multiscale modeling approach by factorizing the likelihood into the individual components of local information. The model that they developed assumes that the hierarchical partitions correspond to the successive aggregation of an initial data space. Here, we describe their model briefly. Assume that D is an arbitrary spatial area and $B_{i,k}$ are subareas in D at the k th spatial scale and relative positions i within the scale, such that $\bigcup_{i=1}^{N_k} B_{i,k} = D$ and $\bigcup_{i' \in ch(i,k)} B_{i',k+1} = B_{i,k}$, where for a given choice of $k = 0, 1, \dots, K$, and $i = 1, \dots, N_k$, $ch(i, k)$ represents the collection of spatial indices i' at scale level k for which $B_{i',k+1} \subseteq B_{i,k}$. Note that $k = 0$ denotes the coarsest level with $N_0=1$ while $k = K$ represents the finest level. Further, let

$\mathbf{Y}_k = (Y_{1,k}, \dots, Y_{N_k,k})^T$ be the vector of aggregated counts at the k th scale and assume $\mathbf{Y}_{ch(i,k)k+1}$ is the vector of aggregated counts $Y_{i',k+1}$ for whom $B_{i',k+1} \subseteq B_{i,k}$. Similar to wavelet decomposition, Louie and Kolaczyk¹¹ assume that the information in the original measurement at the finest scale K , i.e. \mathbf{Y}_K , decomposes into conditionally independent

components as those measurements are aggregated across each successively coarser level. Mathematically, it can be expressed as

$$\Pr(\mathbf{Y}_K | \boldsymbol{\mu}_K) = \Pr(Y_{1,0} | \mu_{1,0}) \prod_{k=0}^{K-1} \prod_{i=1}^{N_k} \Pr(\mathbf{Y}_{ch(i,k),k+1} | Y_{i,k}, \boldsymbol{\omega}_{i,k})$$

where the count at the coarsest level follows a Poisson distribution, $Y_{1,0} \sim \text{Poisson}(\mu_{1,0})$, and conditioning on the count at the coarser scale, the aggregated counts at the finer level follow a multinomial distribution, $\mathbf{Y}_{ch(i,k),k+1} | Y_{i,k}, \boldsymbol{\omega}_{i,k} \sim \text{Multinomial}(Y_{i,k}; \boldsymbol{\omega}_{i,k})$, with

$$\boldsymbol{\omega}_{i,k} = \mu_{i,k}^{-1} \times \boldsymbol{\mu}_{ch(i,k),k+1}.$$

Nevertheless, their approach is limited by the conditional assumption, denoted by $\mathbf{Y}_{ch(i,k),k+1} | Y_{i,k}$. This means that conditioning on the counts at the k th coarser level ($Y_{i,k}$), the counts at $k+1$ finer scale level ($\mathbf{Y}_{ch(i,k),k+1}$) follow a multinomial distribution. This is not true in reality because health data are collected at the finer level and not at the coarser level. Moreover, the data at the coarser level are an aggregation of the data at the finer level. In addition, their method does not help to assess whether a general clustering effect exists in the overall region because it does not contain the spatial component which captures the extra-Poisson variation. To address this issue, Louie and Kolaczyk¹¹ considered a Poisson log-normal model at each individual and independent scale k . However, this method does not accommodate the individual information at the finest level and it is not flexible enough to estimate the relative risks at the lowest and highest levels at the same time. To overcome such issues, in this paper, we develop a multiscale log-normal modeling framework that can be used to make inference both at the higher (areas) and lower levels (subareas) simultaneously using Bayesian models. We also evaluate the performance of the different multiscale models via a simulation study.

This paper proceeds as follows. Section 2 is devoted to the description of the data set, followed by the statistical methodologies and the design of the simulation study that will be elaborated in Section 3. The simulation results and the application of the multiscale modeling to a real data set will be presented in Section 4. Finally, in Section 5, we discuss the main findings and draw conclusions from the results.

2 Georgia oral cancer data

As part of our analysis of multiscale effects, we chose to examine a real data set: Georgia county and public health (PH) district oral cancer incidence. This was chosen as it provides a reasonably large set of spatial units at each scale. The outcome of interest is the number of persons discharged from non-federal acute-care inpatient facilities for oral cancer in 2008 at both the county and public health levels. These PH districts are the administrative units that provide health services. The counts of disease at the public health level were created by aggregating the counts at the county level nested within the PH district. Hence, the data exhibit a scaling effect when aggregated from the county to the PH level. This effect should be considered during modeling, a point on which we will elaborate in Section 3. There are

159 counties that are nested within 18 public health districts, as shown in Figure 1. Our analysis of these data is deferred to Section 4.2.

3 Multi-scale modeling

In disease mapping, the information conveyed by maps varies with scale. Louie and Kolaczyk¹¹ proposed a factorization of the likelihood that contains the information on the scaling effect in a multiscale fashion under the assumed Poisson model. Conditioning on the coarser level, they assumed a multinomial distribution for the data at the finer level. In this paper, we incorporate the scale effect using a convolution (BYM) multiscale modeling approach. We have proposed four different models to account for the scaling effect. We discuss each of these models in the following subsections.

3.1 Model 1

The framework, which was developed by Louie and Kolaczyk,¹² can be further extended to estimate relative risk in Poisson-based models for count data. We are motivated by their use of convolution models in estimating the relative risk of data available at multiple scales. However, these authors considered the convolution model at the finer and coarser levels separately. Here, we study the model both at the finer and coarser levels simultaneously by including a shared random effect to handle the scaling effect.

Several studies have looked at multiscale modeling of spatial data within the Bayesian framework.^{13–15} The models proposed by many of the previous studies involve complex statistical and computational techniques that may not always be easily implemented in standard software. Taking into account the need for simpler and more user-friendly methods, we propose multiscale convolution models to obtain smoother risk estimates for multiscale data.

Suppose \mathbf{y}_k is a vector of observed aggregated outcomes $y_{i,k}$ for spatial unit i at the k th scale level, $k = 0, 1, \dots, K$; $i = 1, \dots, N_k$, where N_k is defined to be the number of units at the k th level and K denotes the number of levels. We use $ch(i, k)$ to denote the set of spatial subunits at the $k+1$ level uniquely allocated within the i th unit of the k th level. For example, for count data, we can express the aggregation as $y_{i,k} = \sum_{l \in ch(i,k)} y_{l,k+1}$ for $k = 0, 1, \dots, K$. The aggregated counts at each level are assumed to be conditionally independently distributed according to a Poisson distribution with mean $\mu_{i,k} = e_{i,k} \theta_{i,k}$, i.e. $y_{i,k} | \theta_{i,k} \sim \text{Poisson}(\mu_{i,k} = e_{i,k} \theta_{i,k})$, where $e_{i,k}$ is the expected number of cases in the i th area at scale k and $\theta_{i,k}$ is the relative risk given by

$$\log(\theta_{i,k}) = a_{0,k} + v_{i,k} + u_{i,k}. \quad (1)$$

Here $a_{0,k}$ is the intercept at scale level k , whereas $u_{i,k}$ and $v_{i,k}$ are the spatially structured and unstructured random effects for unit i at scale level k , respectively. Let $p_k(\boldsymbol{\theta}_k)$ be the joint prior distribution of the N_k components of $\boldsymbol{\theta}_k = (\theta_{1,k}, \dots, \theta_{N_k,k})$ for scale k , i.e.

$p_k(\boldsymbol{\theta}_k) = \prod_{i=1}^{N_k} p_{i,k}(\theta_{i,k})$, and $p_k(\mathbf{y}_k | \boldsymbol{\theta}_k) = \prod_{i=1}^{N_k} p_{i,k}(y_{i,k} | \theta_{i,k})$ is the joint distribution of the

sample which is represented as the likelihood $L(\boldsymbol{\theta}_k|\mathbf{y}_k)$ when viewed as a function of $\boldsymbol{\theta}_k$. Then, the posterior distribution at scale k , which is a combination of the prior distribution and the likelihood function, is defined as

$$P_k(\boldsymbol{\theta}_k|\mathbf{y}_k) \propto L_k(\boldsymbol{\theta}_k|\mathbf{y}_k)P_k(\boldsymbol{\theta}_k). \quad (2)$$

To obtain the posterior parameter estimates, specifically the relative risk at each scale, we can sample from the updated posterior distribution $P_k(\boldsymbol{\theta}_k|\mathbf{y}_k)$ using an MCMC sampling method.

Linkage between the different levels can be achieved via the spatially structured and unstructured random effects or directly through the relative risk. Following that, we will discuss how we can implement these models for two level count data.

Let $y_{l,2}$, $l = 1, \dots, N_2$, be the subunit level counts of disease and $y_{i,1} = \sum_{l \in ch(i,1)} y_{l,2}$, $i = 1, \dots, N_1$ be the i th unit level counts of disease obtained by summing the counts at the subunit level; N_1 and N_2 are the number of units and subunits, respectively. In Model 1, we considered a joint convolution model at the subunit and unit levels. The linkage between these two levels was incorporated in the model by including a shared spatially structured unit-level random effect $u_{i,1(l \in i)}$, where $(l \in i)$ denotes the subunit l within the i th unit. For example, if unit i has m_i subunits, each of the subunits (children) inherits a common characteristic from their unit (parent) via the shared random effect $u_{i,1(l \in i)}$. In other words, $u_{i,1(l \in i)}$ will be common for all the children who belong to the same parent. For unit i and subunit l within i , the model is given by

$$\begin{aligned} y_{i,1} &\sim \text{Poisson}(e_{i,1}\theta_{i,1}), \\ \log(\theta_{i,1}) &= a_{0,1} + v_{i,1} + u_{i,1}, \\ y_{l,2} &\sim \text{Poisson}(e_{l,2}\theta_{l,2}), \\ \log(\theta_{l,2}) &= a_{0,2} + v_{l,2} + u_{i,1(l \in i)} \end{aligned} \quad (3)$$

Note that in this model the spatially correlated random effect is shared from the unit-level model to the subunit-level model. In other words, the spatial association is defined at the coarser (unit) level, and assumed to be the same at the finer (subunit) level. Here,

$e_{i,1} = p_{i,1} \frac{\sum y_{i,1}}{\sum p_{i,1}}$ and $e_{l,2} = p_{l,2} \frac{\sum y_{l,2}}{\sum p_{l,2}}$ are the expected rates with $p_{i,1}$ and $p_{l,2}$ denoting the population size at the coarser and finer levels, respectively. For this model and for the other models below, we have assumed a flat prior for the intercept parameters, $a_{0,1}$ and $a_{0,2}$. Further, the uncorrelated heterogeneity (UH) random effects, $v_{i,1}$ and $v_{l,2}$, were assumed to be normally distributed, i.e. $v_{i,1} \sim N(0, sd_{v,1}^2)$ and $v_{l,2} \sim N(0, sd_{v,2}^2)$, whereas the correlated heterogeneity random effect, $u_{i,1}$, was assumed to have a conditional autoregressive (ICAR) distribution, which is the most widely used method because of its theoretical properties, computational and interpretation advantages.^{3,16} The ICAR structure can be expressed as

$$u_{i,1}|u_{-i,1} \sim N\left(\bar{u}_{\delta_{i,1}}, \frac{sd_{u,1}^2}{n_{\delta_{i,1}}}\right), \quad (4)$$

where

$$\bar{u}_{\delta_{i,1}} = \frac{1}{n_{\delta_{i,1}}} \sum_{i' \in \delta_{i,1}} u_{i',1}.$$

Here, $n_{\delta_{i,1}}$ is the cardinality of $\delta_{i,1}$, which denotes the set of labels of the neighbors of unit i and $u_{-i,1}$ is the set of all random effects not including the i th. For the hyperparameters, $sd_{v,2}$, $sd_{v,1}$, and $sd_{u,1}$, we considered a uniform prior distribution, $U(0, 100)$.¹⁷

3.2 Model 2

Model 2 is similar to Model 1 except that in this model an extra spatially structured random effect is included at the subunit level, i.e. $u_{l,2}$ is added to the model. This extra random effect captures the additional spatial structure at the finer level, which is not captured by the coarser level random effect. The model can be written as

$$\begin{aligned} y_{i,1} &\sim \text{Poisson}(e_{i,1}\theta_{i,1}), \\ \log(\theta_{i,1}) &= a_{0,1} + v_{i,1} + u_{i,1}, \\ y_{l,2} &\sim \text{Poisson}(e_{l,2}\theta_{l,2}), \\ \log(\theta_{l,2}) &= a_{0,2} + v_{l,2} + u_{l,2} + u_{i,2}(l \in i). \end{aligned} \quad (5)$$

$v_{i,1}$ and $v_{l,2}$ were assumed to follow an independent normal distribution as in Model 1 and we assumed a conditional autoregressive (ICAR) distribution for the spatially structured random effect given by

$$u_{l,2}|u_{-l,2} \sim N\left(\bar{u}_{\delta_{l,2}}, \frac{sd_{u,2}^2}{n_{\delta_{l,2}}}\right), \quad (6)$$

where

$$\bar{u}_{\delta_{l,2}} = \frac{1}{n_{\delta_{l,2}}} \sum_{l' \in \delta_{l,2}} u_{l',2}.$$

Here, $n_{\delta_{l,2}}$ is the cardinality of $\delta_{l,2}$, which represents the set of labels of the neighbors of subunit and $sd_{u,2} \sim U(0, 100)$. Note that two counties are neighbors if they share a common boundary and it is possible a pair of counties can be neighbors even if they lie within different (but adjoining) public health districts.

3.3 Model 3

The third model assumes two separate convolution models at both the subunit and unit levels. The model for unit i and subunit l within i is of the form

$$\begin{aligned} y_{i,1} &\sim \text{Poisson}(e_{i,1}\theta_{i,1}), \\ \log(\theta_{i,1}) &= a_{0,1} + v_{i,1} + u_{i,1}, \\ y_{l,2} &\sim \text{Poisson}(e_{l,2}\theta_{l,2}), \\ \log(\theta_{l,2}) &= a_{0,2} + v_{l,2} + u_{l,2}. \end{aligned} \quad (7)$$

Note that this model does not introduce linkage between the subunit and unit levels; rather it assumes they are independent. The distributions for $v_{i,1}$, $v_{l,2}$, $u_{i,1}$, and $u_{l,2}$ were assumed to be the same as in Model 2.

3.4 Model 4

Finally, in model 4, we assumed a convolution model to calculate the relative risk at the subunit level, $\theta_{l,2}$, while the relative risk at the unit level, $\theta_{i,1}$, is obtained by aggregating over the subunit estimated effects. The model is defined as

$$\begin{aligned} y_{l,2} &\sim \text{Poisson}(e_{l,2}\theta_{l,2}), \\ \log(\theta_{l,2}) &= a_{0,2} + u_{l,2} + v_{l,2}, \\ \theta_{i,1} &= \mu_{i,1} / e_{i,1}, \\ \mu_{i,1} &= \sum_{l \in ch(i,1)} e_{l,2}\theta_{l,2}. \end{aligned} \quad (8)$$

Note that Model 4 is different from Models 1–3 in the sense that we did not model the unit level data. Indeed, we did not include spatially correlated and uncorrelated random effects at the unit level, as in Models 1–3. This procedure simplifies the model because we have fewer free parameters to estimate as compared with Models 1–3. The motivation to consider this model is the fact that the data at the coarser level are an aggregation of the data at the finer level, and hence, we argue that the relative risk at the coarser level can be obtained by averaging the relative risk at the finer level within the coarser level.

3.5 Model assessment and goodness of fit

To investigate the performance of the models, the deviance information criterion (DIC^{18,19}), which is a combination of the likelihood function (deviance) and a model complexity term (PD_{dic}; number of effective parameters), was calculated. We have also considered other criteria for model comparison such as WAIC^{20,21} (Watanabe-Akaike information criterion). For a predictive accuracy assessment, the mean absolute prediction error (MAPE) and mean square prediction error (MSPE) were used. In addition, we examined local measures of fit including the conditional predictive ordinate (CPO³).

WAIC is a fully Bayesian technique for model selection and uses a posterior distribution rather than a point estimate. WAIC can be considered as computationally convenient

approximation to cross validation. To compute WAIC, define the log pointwise predictive density (lppd) as

$$\text{lppd} = \log \prod_{i=1}^n p_{\text{post}}(y_i) = \sum_{i=1}^n \log \int p(y_i | \theta) p_{\text{post}}(\theta) d\theta$$

where $p(y_i | \theta)$ is the log predictive density and $p_{\text{post}}(\theta)$ represents the posterior distribution. In practice, we can compute lppd by evaluating the expectation using draws $(\theta^s, s = 1, \dots, S)$ from the posterior distribution $p_{\text{post}}(\theta)$ as follows

$$\text{computed}_{\text{lppd}} = \sum_{i=1}^n \log \left(\frac{1}{S} \sum_{s=1}^S p(y_i | \theta^s) \right).$$

Hence, the WAIC can be calculated as

$$\begin{aligned} \text{WAIC} &= -2 * \widehat{\text{elpd}}_{\text{waic}}, \\ \widehat{\text{elpd}}_{\text{waic}} &= \text{computed}_{\text{lppd}} - \text{PD}_{\text{waic}}. \end{aligned}$$

Here, $\text{PD}_{\text{waic}} = \sum_{i=1}^n V_{s=1}^S \log(p(y_i | \theta^s))$ with $V_{s=1}^S \log(p(y_i | \theta^s))$ is the sample variance of the log predictive density for each data point y_i and $V_{s=1}^S a_s = \frac{1}{S-1} \sum_{s=1}^S (a_s - \bar{a})^2$, where $a_s = \log(p(y_i | \theta^s))$. On the other hand, $\widehat{\text{elpd}}_{\text{waic}}$, PD_{waic} , n , and S denote the expected log pointwise predictive density, effective number of parameters, the number of data points, and the number of simulation draws, respectively.

All the models were implemented using the mix of Gibbs and Metropolis-Hastings algorithms of Markov chain Monte Carlo (MCMC) in WinBUGS invoked through R with 15,000 samples after the first 15,000 samples were discarded from the analysis. We considered three chains in parallel, and hence the posterior inferences are based on 45,000 samples. For all models, convergence was assessed using an estimated potential scale reduction factor, \hat{R} , trace plots, and Brooks, Gelman and Rubin's (BGR) plots.²²

3.6 Simulation study

The aim of our simulation study is to evaluate and compare Models 1–4 using DIC, WAIC, MAPE, MSPE, bias, and MSE of the parameters obtained by fitting the models to simulated data sets under a handful of scenarios. To simulate data that could be encountered in a real life, we considered two situations. In the first situation, we simulate the relative risk ($\theta_{i,K}$) and the expected rate ($e_{i,K}$) at the finer level K from a gamma distribution and then we generate the data from a Poisson distribution with mean equal to $\theta_{i,K} e_{i,K}$. In the second situation, we simulate the relative risk using a convolution model ($\log(\theta_{i,K}) = a_{0,K} + u_{i,K} + v_{i,K}$) and then we sample the data from a Poisson's distribution with mean equal to $\theta_{i,K} e_{i,K}$. We describe these scenarios in the following sections.

3.6.1 Scenario 1: simulation from Poisson–gamma distributions—In practice, we may have data at census tract, county, and public health district levels. Taking into account this structure, we simulated data from a Poisson distribution with three levels of hypothetical grid divided into $2^4 \times 2^4 = 256$ smaller areas of the finest (lower) level, $2^3 \times 2^3 = 64$ areas (pixels) at the medium level, and $2^2 \times 2^2 = 16$ areas in the coarsest (higher) level. First, 256 samples were generated from a Poisson distribution at the finest (lower) level. To obtain the 64 samples at the next level (medium), we aggregated the samples at the finest level nested within the medium level. Similarly, the 16 samples were obtained by aggregating the observations at the medium level nested within the coarsest level. Mathematically, this could be expressed as follows

$$\begin{aligned} y_{j,1} &= \sum_{i \in ch(j,1)} y_{i,2}, \\ y_{j,2} &= \sum_{l \in ch(i,2)} y_{l,3}, \\ y_{l,3} &\sim \text{Poisson}(e_l, \theta_{l,3}) \end{aligned} \quad (9)$$

where $e_{l,3}$ is the expected rate, $\theta_{l,3}$ denotes the relative risk at the finest level, $ch(i, 2)$ is the set of subareas at the lower level nested within the i th area at the medium level, and $ch(j, 1)$ represents the set of subareas at the medium level nested within the j th area at the higher level. Further, $y_{l,3}$, $l = 1, \dots, N_1 = 256$, $y_{i,2}$, $i = 1, \dots, N_2 = 64$, and $y_{j,1}$, $j = 1, \dots, N_3 = 16$, are the samples generated at the lower, medium, and higher levels, respectively. We assumed $\theta_{l,3}$ follows a gamma distribution with shape α , and scale β parameters, i.e. $\theta_{l,3} \sim \text{gamma}(\alpha, \beta)$. For the hyperparameter of the gamma distribution, we assumed both α and β equal to one so that the mean and variance of $\theta_{l,3}$ will be one. This assumption could generate relative risks that are similar to real life examples. We assumed that the expected numbers in the 256 areas are fixed and equal to one, $e_{l,3} = 1$, $l = 1, \dots, 256$.

3.6.2 Scenario 2: simulation from a convolution model—In turn, we generated data similar to the first scenario in equation (9), but now we assumed a convolution model for the relative risk, i.e. $\log(\theta_{l,3}) = a_{0,3} + u_{l,3} + v_{l,3}$. To compute the relative risk $\theta_{l,3}$, we have to specify values for the intercept $a_{0,3}$ and generate the spatially structured ($u_{l,3}$) and unstructured ($v_{l,3}$) random effects from a ICAR and normal distributions, respectively. There are two options in order to sample the spatially structured random effects $u_{l,3}$ from the ICAR distribution. First, we simulated $u_{l,3}$ from the ICAR using the *car.normal* function through the BRugs Package. However, with this approach, we have to use the MCMC sampling method and it is computationally intensive. Hence, in the second option, we sampled these random effects directly from the ICAR (6) using a program written in R software. The steps are as follows:

1. The sampling approach is initiated using starting values for $u_{l,3}^0$ from a standard normal distribution, $u_{l,3}^0 \sim N(0, 1)$, and we assumed the overall variance equals one, $sd_{u,3}^2 = 1$. Note that we have also assumed different possible values for $sd_{u,3}^2$.

2. We sampled $u_{l,3}^t$ from a normal distribution with mean $\bar{u}_{\delta_{l,3}}^t = \frac{1}{n_{\delta_{l,3}}} \sum_{l' \in \delta_{l,3}} u_{l',3}^0$ and with variance equal to $sd_{u,3}^2/n_{\delta_{l,3}}$, i.e., $u_{l,3}^t \sim N\left(\bar{u}_{\delta_{l,3}}^t, \frac{sd_{u,3}^2}{n_{\delta_{l,3}}}\right)$, where $t = 1$.
3. Conditioning on the current simulated values of the correlated heterogeneity $(u_{l,3}^t)$, we simulated the next sample $u_{l,3}^{t+1}$ from a normal distribution with mean $\bar{u}_{\delta_{l,3}}^{t+1} = \frac{1}{n_{\delta_{l,3}}} \sum_{l' \in \delta_{l,3}} u_{l',3}^t$ and variance equal to $sd_{u,3}^2/n_{\delta_{l,3}}$, where $t = 1$.
4. We repeated step 3 for $t = 2, \dots, T$ until we obtain a clustered random effect. In our case, with $T = 10$, we obtained spatially structured random effects similar to the one generated through BRugs package using *car.normal* function.

The unstructured random effects ($v_{l,3}$) were simulated from a normal distribution with mean equal to zero and variance $sd_{v,3}^2$. Note that, the expected rates were sampled from a $\text{gamma}(1, 1)$.

4 Results

4.1 Simulation results

In this section, we describe the simulation results. First, we present the findings obtained from the simulation study in scenario 1. Thereafter, we present the results obtained from the models fitted to the simulated data in scenario 2. Table 1 displays the models fitted to 200 simulated data sets using the mix of Gibbs and Metropolis-Hastings algorithms. First, we assessed convergence for some simulated data sets, and the trace plots, BGR plots, and \hat{R} suggest that convergence has occurred for three chains of 15,000 sample size after the first 15,000 samples were discarded from the analysis. We also checked the convergence for a single chain of 15,000 samples using trace plots and here also there was a good mixing between the samples, and hence we decided to use a single chain with 15,000 sample size after convergence. To compare the models, the bias and MSE of the relative risks were calculated. To evaluate the predictive ability of the models, we calculated the MSPE and MAPE. Besides, PD, DIC, and WAIC were computed at each scale level to compare model performance. Finally, the computation time was extracted to compare the execution time for the models. To summarize the simulation results, we calculated the mean values of the bias, MSE, MAPE, MSPE, PD, DIC, and WAIC obtained from the models fitted to the 200 simulated data sets.

4.1.1 Scenario 1: results for data simulated from Poisson-gamma distribution

—The results obtained from the models fitted to sampled data within the hypothetical grid are shown in Table 2. When the relative risk is assumed to follow a gamma distribution with shape and scale parameters equal to one, Models 1 and 2 produce DIC value less than Models 3 and 4, especially at the lower and medium levels. Similarly, the bias and MSE of the relative risk computed from Models 1 and 2 are smaller than the bias and MSE of the relative risk obtained from Models 3 and 4. Further, Model 1 converges faster than the other models. However, the first three models produce similar MAPE and MSPE at all levels

except for a slight difference at the medium level. Note that Model 4 reveals MAPE and MSPE higher than the other models in the medium and higher levels. We have also simulated data within the Georgia state map and the results are presented in the Supplementary Appendix, Table 1. Note that in all tables, values that are smallest over the different models are indicated in bold.

4.1.2 Scenario 2: results for data simulated from a convolution model—Tables 3 and 4 display the results of the model fitted to the simulated data within the hypothetical grid. We can see that similar conclusion can be drawn as in scenario 1. In particular, the PD of the DIC penalizes more for model complexity as compared to the PD of the WAIC and they both tend to decrease as we aggregate the data from lower to higher level. This is not surprising because the effective number of parameters depends on the sample size. The same is true for the DIC, WAIC, and deviance. Although there is not much difference in the DIC computed from the different models at lower and higher levels, there is a large difference at the medium level. Note that the results of the DIC and WAIC are close to each other as they should be. The results of the DIC for all the 200 data sets are shown in Figure 2. Here also, Models 1 and 2 outperform the other models.

Furthermore, we have simulated data within the Georgia state map assuming the spatially structured and unstructured random effects different for all the 200 data sets and the results are presented in the Supplementary Appendix, Tables 2–5. The conclusions from these results are also similar to the conclusions drawn from the results for the simulated data within the hypothetical grid. To investigate how robust our results are with the assumed values for the random effects, we have generated data assuming the spatially structured random effect is the same for all 200 data sets (see the results in the Supplementary Appendixes, Tables 6 and 7).

4.2 Application to the Georgia oral cancer data

To assess the benefit of the shared random effect models, we applied Models 1–4 (see Section 3) to the Georgia oral cancer study. The estimated potential scale reduction factor (Table 7), trace, and BGR plots suggest that convergence has occurred for all model parameters. The results of the model fit and predictive accuracy are shown in Table 5. There is a gain in terms of model fit (DIC and WAIC), especially at the PH level. Moreover, the parsimonious model (Model 1) fits the data slightly better than Model 2. Hence, in this example, including the structural random effect at the county level (Model 2) does not improve the model fit other than adding model complexity. The predictive accuracy (MAPE and MSPE) at the PH level for Model 1 is better than the other models. However, the predictive accuracy at the county level is not different between the models. If we do not take into account model complexity, the model which introduces two separate convolution models (Model 3) provides better model fit (Deviance) at the county level as compared to the shared random effect models (Models 1 and 2). According to the CPO, which does not penalize model complexity, all the models perform similarly at the county level, but there is a slight improvement in the shared random effect models at the PH level (Figure 4). We have also calculated the computation time in seconds to check which model converges faster and the results are shown in Table 5 under the column labeled as CT, which represents the

computation time in seconds. We can clearly see that Model 1 converges much faster than the other models. These results are consistent with the results obtained from the simulation study. The implementation of Models 1–4 using WinBUGS programs is given in the Supplementary Appendix (Section 2).

The posterior summary statistics are given in Table 6. The study-wide relative risks of all the models, ($\exp(a_{0,2})$ and $\exp(a_{0,1})$), is approximately equal to one with 95% credible interval in the range of 0.85–1.2 at both the county and PH levels. Hence, there is no significant increase in the number of persons discharged from non-federal acute-care inpatient facilities for oral cancer in the Georgia state as compared to what is expected. However, this is not true for each county and PH district (Figure 3). Here, the relative risks obtained from Model 4 in the PH level range between 0.03 and 2.78, while for the other models between 0.72 and 1.72. There is some inconsistency between the relative risk obtained from Model 4 at the PH level and the other models. For instance, Model 4 indicates an elevated risk in the southwest border, whereas the other models show a lower risk in this PH district. This may be due to the correlation between the neighbors which is not accounted for in Model 4 at the PH level. Moreover, the trend of the relative risk in Models 1–3 is similar at both county and PH levels. There is an elevated risk to the southeast border at both levels. On the other hand, all the models produce similar relative risk at the county level. Note that the SMR at the county level provides crude estimates which ranges from zero to seven because it does not adjust the spatial correlation between the counties and the noise in the data (Table 8).

The variability of the unstructured random effects ($sd_{v,1}$ and $sd_{v,2}$) obtained from the models is similar. Nevertheless, the variability of the structured random effect ($sd_{u,2}$) from Model 2 is smaller than that of Model 3. This may be due to some part of this variability in Model 2 being accommodated for by the shared random effect ($sd_{u,1} = 0.474$), which is higher than the variance in Model 3 ($sd_{u,1} = 0.411$).

5 Discussion and conclusion

In this paper, we investigated the effect of scaling in disease mapping using a multiscale Bayesian modeling framework. We have shown that the scaling effect could be accommodated by using shared random effect multiscale models (Models 1 and 2). These models provide not only a better fit to the data, but also produce a better predictive accuracy as compared to the independent convolution model (Model 3) and Model 4, especially at the coarser level. This is an expected result because the effect of the coarser level is inherited by the finer level through the shared random effect. The parsimonious shared random effect model (Model 1) also converges faster than the other models.

The simulation results indicate that models with shared random effects are the best model, especially at the coarsest level, when the data are simulated from a Poisson distribution with mean equal to $e_{l,3}\theta_{l,3}$ in which $e_{l,3}$ and $\theta_{l,3}$ are generated from a gamma distribution with the shape and scale parameters equal to one. We also found that the bias and MSE of the relative risks computed from these models are lower than those of Models 3 and 4.

For data drawn from a convolution model within the Georgia state map assuming the spatially structured and unstructured random effects different for all the simulated data sets, the DIC favors the shared random effect models at the PH level, while the DIC prefers Model 4 at the county level. However, DIC supports the shared random effect models at the county level when the data are sampled from a convolution model with constant spatially correlated and uncorrelated random effects for all the simulated data sets. Here, Model 3, which uses an independent convolution model at the county and PH levels, fits the simulated data as well as the shared random effect models. On the other hand, the DIC obtained from the shared random effect models at the lower and medium levels for the data generated within a hypothetical grid is much lower than those of Models 3 and 4. In this spatially structured simulation, we have also obtained more unbiased and efficient estimates of the relative risks for the shared random effect models as compared to the independent convolution model. All the models recover well the assumed overall log relative risk ($a_{0,3}$), spatially structured ($u_{i,3}$), and unstructured ($v_{i,3}$) random effects. Nevertheless, Models 3 and 4 yield slightly better estimates of the variances of the spatially structured ($sd_{u,3}$) and unstructured ($sd_{v,3}$) random effects in most of the cases.

To investigate the results obtained from the simulation study in a real example, we applied Models 1–4 to the Georgia oral cancer data. Here also, both the DIC and WAIC tend to select the shared random effect models, especially at the PH level. Although there is not much difference at the county level, the predictive accuracy of the shared random effect models is better than that of Models 3 and 4 at the PH level. However, the deviance, which does not penalize model complexity, favors the parsimonious multiscale models, Models 3 and 4, at the county level, while it selects Models 1 and 2 at the PH level.

Multiscale modeling has been studied by several researchers.^{10–11} Our approach of multiscale modeling is different from Louie and Kolaczyk¹¹ in three respects. First, those authors assume a multinomial distribution for the outcome at the finer level conditioning on the outcomes at the coarser level. This assumption introduces a fixed coarser level effect. On the other hand, our method assumes a random effect at this level and the scaling effect of the coarser level is inherited into the finer level via the shared random effect model. Second, their method does not encompass the spatially structured random effect that handles the correlation between the neighbors, whereas our methods do so. Finally, our convolution multiscale models are easily implemented in standard software such as WinBUGS.

In summary, the shared random effect models outperform the other models both in real and simulated data. Interestingly, the parsimonious shared random effect model, i.e. the model that excludes the spatially correlated random effect at the finer level is as competitive as Model 2, which includes the correlated random effect. This may be due to the fact that the shared random effect that was inherited from the coarser level is flexible enough to handle the correlation between the neighbors. Although it is not as attractive as the shared random effect models, Model 4 is slightly better than the independent convolution model. These results indicate that there should be a linkage to account for the scaling effect between the finer and coarser levels. We conclude that sharing the random effect between these two levels improves the model fit, predictive accuracy, and estimation of the relative risks.

Though we have achieved better results by including shared random effects into the model, our paper has some limitations. First, we introduced the shared random effect through the spatially structured random effects. Currently, we are investigating sharing unstructured random effects alone and sharing both the structured and unstructured random effects simultaneously to accommodate a scale effect. Although our shared random effect model improves the model fit, especially at the coarser level, it does not quantify the scaling effect. Hence, measuring the scale effect using correlation structures between the finer and coarser level is planned. Furthermore, our multiscale model could be easily extended to account for predictor effects at the finer and coarser levels. We do not pursue this extension here.

In general, our multiscale modeling can be useful in addressing the modifiable areal unit problem when the scaling problem is an issue because it takes into account the scaling through a shared random effect. Hence, for public health applications, it would be useful to jointly model the risks at different levels to obtain more accurate risk estimates for planning purposes.

Acknowledgments

Funding

The authors would like to acknowledge support from the National Institutes of Health via grant R01CA172805. The third author also acknowledges support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy).

References

1. Broeck, J., Brestoff, JR., Kaulfuss, C. Statistical estimation. In: Broeck, VD. Brestoff, J., Jonathan, R., editors. *Epidemiology: Principles and practical guidelines*. The Netherlands: Springer; 2013. p. 417-438.
2. Besag J, York J, Mollié A. Bayesian image restoration with applications in spatial statistics (with discussion). *Ann Inst Stat Math*. 1991; 43:1-59.
3. Lawson, AB. *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology*. 2nd. New York: Chapman and Hall/CRC Press; 2013.
4. Cressie NAC, Chan NH. Spatial modeling of regional variables. *J Am Stat Assoc*. 1989; 84:393-401.
5. Stern, HS., Cressie, NA. Inference for extremes in disease mapping. In: Lawson, A. Biggeri, A. Bohning, D., et al., editors. *Disease mapping and risk assessment for public health*. Chichester: Wiley; 1999. p. 63-84.
6. Moraga P, Lawson AB. Gaussian component mixtures and CAR models in Bayesian disease mapping. *Comput Stat Data Anal*. 2012; 56:1417-1433.
7. Wong, D. The modifiable areal unit problem (MAUP). In: Fotheringham, S., Rogerson, P., editors. *The Sage handbook of spatial analysis*. New York: Sage; 2009. p. 108-123.
8. Cressie NAC. Change of support and the modifiable areal unit problem. *J Geogr Syst*. 1996; 2:83-101.
9. Banerjee, S., Carlin, BP., Gelfand, AE. *Hierarchical modeling and analysis for spatial data*. London: Chapman and Hall/CRC Press; 2004.
10. Kolaczyk ED, Huang H. Multiscale statistical models for hierarchical spatial aggregation. *Geogr Anal*. 2001; 33:95-118.
11. Louie MM, Kolaczyk ED. A multiscale method for disease mapping in spatial epidemiology. *Stat Med*. 2006; 25:1287-1306. [PubMed: 16217848]

12. Louie MM, Kolaczyk ED. Multiscale detection of localized anomalous structure in aggregate disease incidence data. *Stat Med*. 2006; 25:787–810. [PubMed: 16453373]
13. Louie MM, Kolaczyk ED. On the covariance properties of certain multiscale spatial processes. *Statist Probab Lett*. 2004; 66:407–416.
14. Lee S, Yeatts K, Serre M. A Bayesian maximum entropy approach to address the change of support problem in the spatial analysis of childhood asthma prevalence across North Carolina. *Spat Spatiotemporal Epidemiol*. 2009; 25:49–60.
15. You Y, Zhou Q. Hierarchical Bayes small area estimation under a spatial model with application to health survey data. *Surv Methodol*. 2011; 37:25–37.
16. Cressie, NAC. *Statistics for spatial data*. New York: Wiley; 1993.
17. Gelman A. Prior distribution for variance parameters in hierarchical models. *Bayesian Anal*. 2006; 3:515–533.
18. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc Series B*. 2002; 64:583–616.
19. Gelman, A., Carlin, JB., Stern, HS., et al. *Bayesian Data Analysis*. New York: Chapman and Hall; 2004.
20. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res*. 2010; 11:3571–3594.
21. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Statist Comput*. 2013; 24:997–1016.
22. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences (with discussion). *Stat Sci*. 1992; 7:457–511.

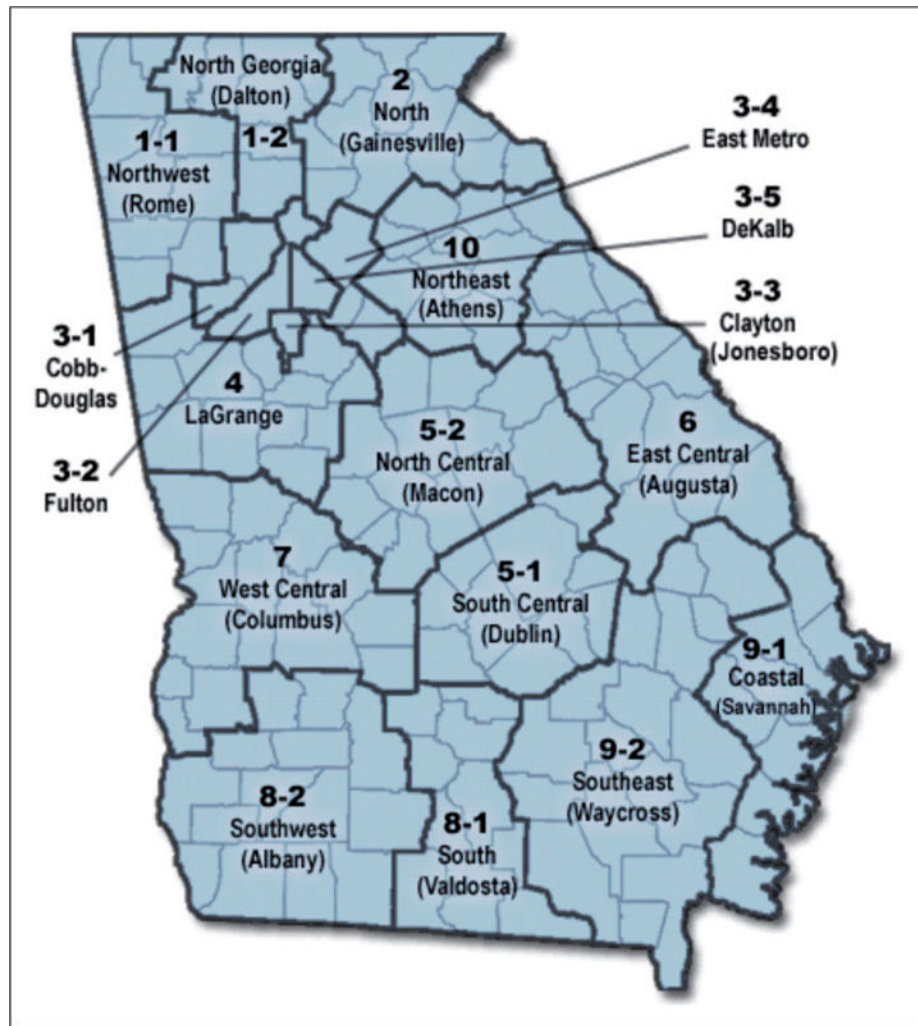


Figure 1.
State of Georgia, USA: County and PH district boundary map.

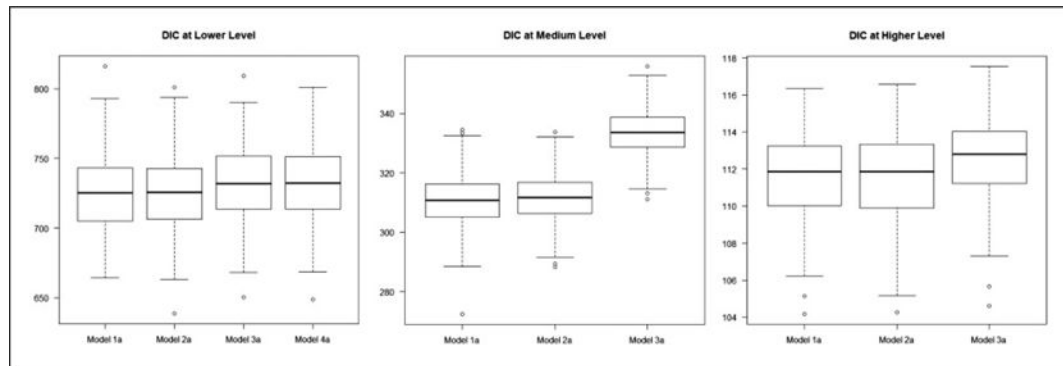


Figure 2.

Deviance information criterion (DIC) values for the 200 simulated data within a hypothetical grid with spatially structured ($u_{l,3}$) and unstructured random effects ($v_{l,3}$) assumed to be constant over all the 200 data sets (scenario 2).

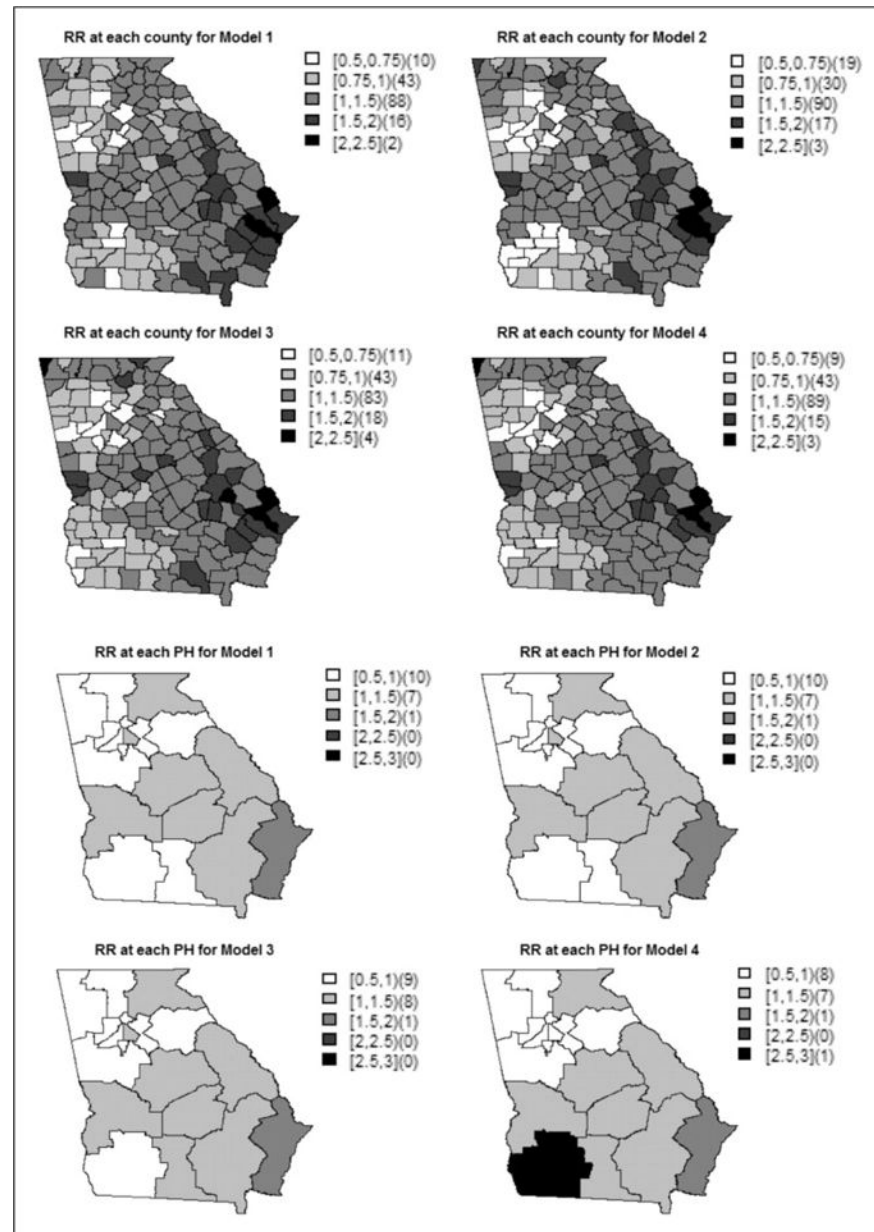


Figure 3. Georgia oral cancer data. Relative Risk (RR) at each county and public health (PH) district.

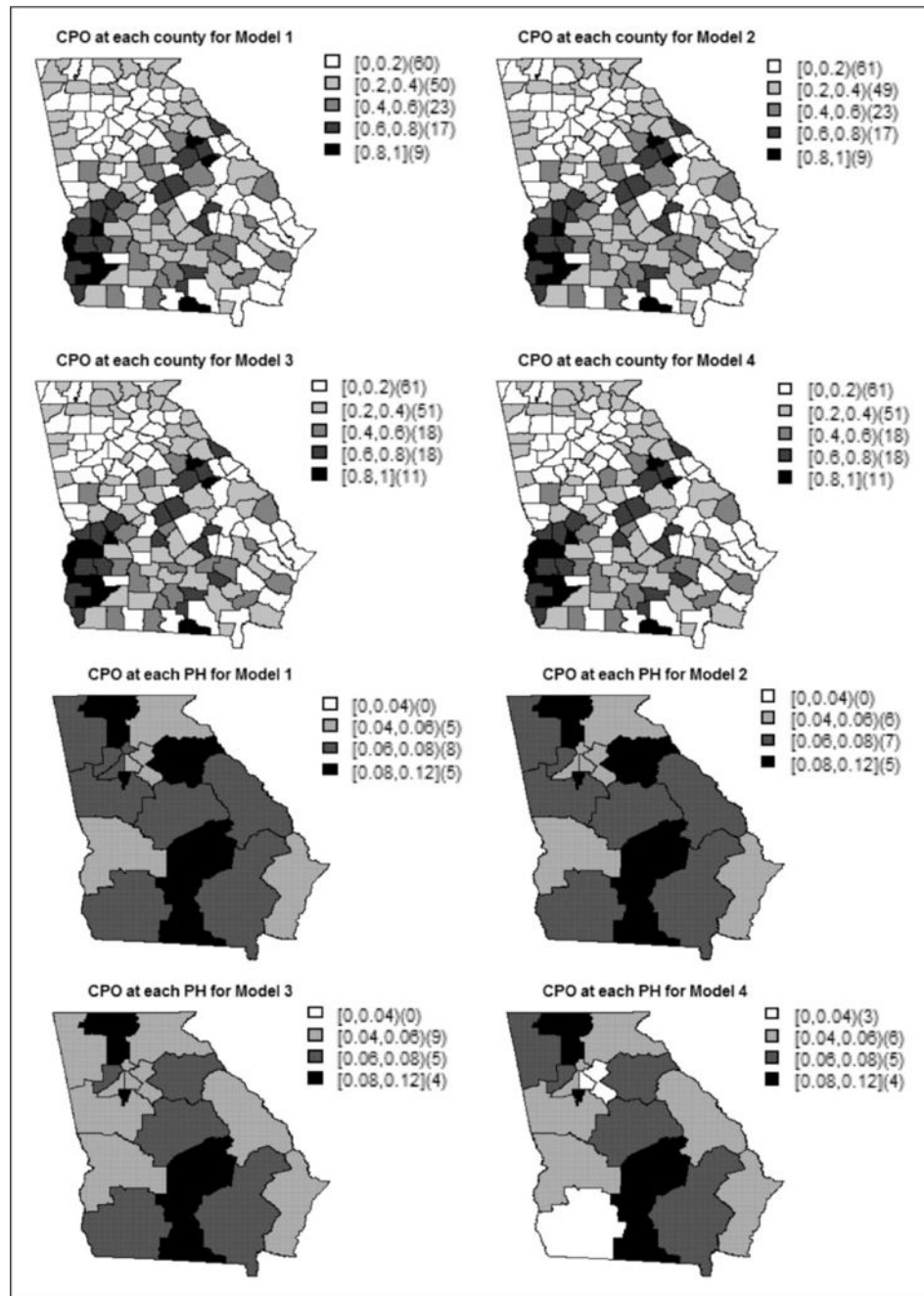


Figure 4. Georgia oral cancer study. Conditional predictive ordinates (CPO) at each county and public health (PH) district.

Table 1
Description of Models fitted to the hypothetical grid data.

Model	Description
Model 1a	$\log(\theta_{j,1}) = a_{0,1} + v_{j,1} + u_{j,1}$
	$\log(\theta_{i,2}) = a_{0,2} + v_{i,2} + u_{i,2} + u_{j,1} \text{ (} i \in j \text{)}$
	$\log(\theta_{l,3}) = a_{0,3} + v_{l,3} + u_{i,2} \text{ (} l \in i \text{)}$
Model 2a	$\log(\theta_{j,1}) = a_{0,1} + v_{j,1} + u_{j,1}$
	$\log(\theta_{i,2}) = a_{0,2} + v_{i,2} + u_{i,2} + u_{j,1} \text{ (} i \in j \text{)}$
	$\log(\theta_{l,3}) = a_{0,3} + v_{l,3} + u_{l,3} + u_{i,2} \text{ (} l \in i \text{)}$
Model 3a	$\log(\theta_{j,1}) = a_{0,1} + v_{j,1} + u_{j,1}$
	$\log(\theta_{j,2}) = a_{0,1} + v_{i,2} + u_{i,2}$
	$\log(\theta_{l,3}) = a_{0,3} + v_{l,3} + u_{l,3}$
Model 4a	$\theta_{j,1} = \sum_{i \in ch(j,1)} e_{i,2} \theta_{i,2} / e_{j,1}$
	$\theta_{i,2} = \sum_{l \in ch(i,2)} e_{l,3} \theta_{l,3} / e_{i,2}$
	$\log(\theta_{l,3}) = a_{0,3} + v_{l,3} + u_{l,3}$

Table 2

Simulation results for the data generated within a hypothetical grid (scenario 1).

Models	PD _{dic}			DIC			MAPE			MSPE			$\theta_{I,3}$			CT elapsed
	lower	medium	higher	lower	medium	higher	lower	medium	higher	lower	medium	higher	bias	MSE		
Models 1a	35.52	33.35	8.61	672.71	265.84	97.82	0.91	1.90	4.48	2.10	6.78	32.33	-0.36	0.89	69.16	
Models 2a	35.88	33.49	8.66	672.19	266.09	97.88	0.91	1.90	4.48	2.10	6.81	32.34	-0.36	0.89	213.81	
Models 3a	36.24	31.03	9.55	713.77	291.62	98.34	0.91	2.17	4.45	2.18	8.17	31.92	-0.46	0.98	207.34	
Models 4a	36.26	-	-	713.78	-	-	0.91	2.53	7.86	2.18	11.37	86.84	-0.46	0.98	484.48	

Note: Lower, medium, and higher represent the three levels; DIC, MAPE, MSPE, $\theta_{I,3} \sim \text{gamma}(\alpha, \beta)$, and *CT* represent the average deviance information criterion, mean absolute prediction error, mean square prediction error, the relative risk at the county level, and computation time in seconds over the 200 data sets, respectively.

Table 3

Simulation results for data generated within a hypothetical grid (scenario 2).

Models	PD _{dic}			DIC			PD _{waic}			WAIC			MAPE			MSPE			C/T Elapsed
	Lower	Medium	Higher	Lower	Medium	Higher	Lower	Medium	Higher	Lower	Medium	Higher	Lower	Medium	Higher	Lower	Medium	Higher	
Model 1a	93.25	36.70	12.48	703.73	304.35	110.12	69.43	18.19	7.50	700.55	291.59	107.89	1.11	2.62	5.97	3.81	14.06	60.25	667.25
Model 2a	95.09	37.12	12.45	702.25	305.49	110.05	69.93	18.62	7.47	698.06	292.92	107.80	1.11	2.63	5.97	3.78	14.12	60.21	1268.04
Model 3a	109.42	44.40	13.49	709.88	326.40	111.21	74.89	27.64	7.80	699.15	319.27	108.44	1.10	2.76	5.97	3.75	15.04	60.42	1352.94
Model 4a	109.49	–	–	709.73	–	–	74.89	26.74	7.44	698.92	318.61	107.87	1.09	2.77	5.97	3.75	14.98	59.78	1746.73

Note: The spatially structured and unstructured random effects were simulated using a programmed R software and were assumed to be different for all the 200 data sets. The assumed values were as follows: $\alpha 0,3 = 0.1$, $sd_{\alpha,3} = 1$, and $sd_{\nu,3} = 1$.

Table 4
Summary of the bias and MSE of the parameters for the data generated within a hypothetical grid (scenario 2).

Models	Assumed values						Bias						MSE												
	$a_{0,3}$	$sd_{u,3}$	$sd_{v,3}$	$\theta_{1,3}$	$a_{0,3}$	$sd_{u,3}$	$sd_{v,3}$	$u_{1,3}$	$u_{2,3}$	$sd_{u,3}$	$sd_{v,3}$	$\theta_{1,3}$	$a_{0,3}$	$sd_{u,3}$	$sd_{v,3}$	$u_{1,3}$	$u_{2,3}$	$sd_{u,3}$	$sd_{v,3}$	$\theta_{1,3}$	$a_{0,3}$	$sd_{u,3}$	$sd_{v,3}$	$u_{1,3}$	$u_{2,3}$
Model 1a	0.1	1	1	1	0.006	0.009	-	0.24	-	0.008	4.499	0.022	-	0.081	-	1.022	-	-	-	-	-	-	-	-	-
Model 2a	0.1	1	1	1	0.009	-0.018	-0.655	-0.258	-0.003	0.008	4.519	0.022	0.507	0.093	0.107	1.077	0.198	1.144	0.201	1.082	0.036	0.319	0.036	0.319	0.036
Model 3a	0.1	1	1	1	0.016	-0.012	-0.356	-0.053	-0.003	0.008	4.583	0.019	0.332	0.034	0.198	1.144	0.201	1.082	0.036	0.319	0.036	0.319	0.036	0.319	0.036
Model 4a	0.1	1	1	1	0.015	-0.014	-0.345	0.054	-0.003	0.008	4.56	0.021	0.319	0.036	0.201	1.082	0.201	1.082	0.036	0.319	0.036	0.319	0.036	0.319	0.036

Table 5

Model fit and predictive accuracy results for Georgia oral cancer data.

Models	PD _{dic}		DIC		PD _{walc}		WAIC		MAPE		MSPE		Deviance		CT
	county	PH district	county	PH district	county	PH district	county	PH district	county	PH district	county	PH district	county	PH district	
Model 1	23.85	8.74	485.09	108.11	21.68	4.12	485.95	104.65	1.39	4.72	4.96	37.01	461.25	99.34	35
Model 2	26.97	9.23	484.17	109.57	23.85	4.65	485.09	106.34	1.38	4.79	4.89	38.2	457.21	100.34	190
Model 3	33.32	11.30	485.36	114.75	27.91	7.02	485.60	112.79	1.36	5.05	4.82	42.4	452.05	103.45	188
Model 4	32.63	–	485.38	–	27.53	–	485.78	–	1.36	8.28	4.83	161.9	452.75	–	259

Table 6

Georgia oral cancer data. Posterior mean estimates, standard error, and 95% Credible Interval.

Models	Mean				sd				95% CI			
	$a_{0,2}$	$a_{0,1}$	$Sd_{u,2}$	$Sd_{u,1}$	$Sd_{v,1}$	$a_{0,2}$	$a_{0,1}$	$Sd_{u,2}$	$Sd_{u,1}$	$Sd_{v,2}$	$Sd_{u,2}$	$Sd_{v,1}$
Model 1	0.003	-2.34E-5	-	0.256	0.492	0.082	0.065	0.059	-	0.122	0.139	0.064
Model 2	-0.010	4.04E-4	0.329	0.219	0.474	0.089	0.070	0.060	0.155	0.117	0.148	0.065
Model 3	0.024	-0.004	0.517	0.259	0.411	0.139	0.074	0.068	0.166	0.121	0.177	0.107
Model 4	0.027	-	0.539	0.237	-	-	0.075	-	0.181	0.109	-	-

Table 7

Georgia oral cancer data. Estimated potential scale reduction factor (\hat{R}).

Models	Point estimate							
	$a_{0,2}$	$a_{0,1}$	$sd_{u,2}$	$sd_{r,2}$	$sd_{u,1}$	$sd_{r,1}$	Deviance	
Model 1	1.00	1.00	–	1.02	1.00	1.00	1.00	1.01
Model 2	1.01	1.04	1.00	1.10	1.00	1.11	1.11	1.02
Model 3	1.00	1.00	1.00	1.10	1.01	1.02	1.02	1.01
Model 4	1.00	1.02	1.00	–	–	–	–	1.01

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 8

Georgia oral cancer study: descriptive statistics of observed out come (O), expected rate (E) and standardized incidence ratio (SIR) at both the county and public health district.

	County			PH district		
	O	E	SIR	O	E	SIR
Minimum	0.00	0.07	0.00	7.00	6.18	0.49
1st quartile	0.00	0.47	0.00	14.55	14.50	0.77
3rd quartile	3.00	2.03	1.74	26.00	27.21	1.31
Maximum	32.00	36.00	7.11	40.00	38.99	1.79
Median	1.00	0.92	0.90	24.00	19.66	0.96
Mean	2.42	2.42	1.19	21.39	21.39	1.05
SD	4.51	5.02	1.39	9.12	9.33	0.36