Bayesian multi-scale modeling for aggregated disease mapping data

Non 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

Bayesian Multiscale Modeling for Aggregated Disease Mapping Data

Mehreteab Aregay[1] **Andrew B. Lawson**[1]

Christel Faes [2] **Russell Kirby** [3]

[1] *Department of Public Heath Sciences, Division of Biostatistics and Bioinformatics, MUSC, Charleston, USA*

[2] ~~*Department of Mathematics and Statistics*~~*, Hasselt University, Hasselt, Belgium*

[3] *Department of Community and Family Health, University of Southern Florida, Lakeland, USA*

Abstract

In disease mapping, a scaling effect due to an aggregation of data from a finer to a coarser level is a common phenomenon. This article focuses on addressing 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 one assumes two independent convolution models at the finer and coarser levels. The fourth 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 levels. All these models were compared based on predictive accuracy, deviance information criterion (DIC), and Watanabe-Akaike or widely applicable information criterion (WAIC) that are applied to real and simulated data. The results indicate that the models with shared random effect outperform the other models on a range of criteria.

Some Keywords: Deviance information criterion (DIC); Watanabe-Akaike or widely applicable information criterion (WAIC); predictive accuracy; shared random effect model; scaling effect.

1 Introduction

In spatial epidemiology (disease mapping), the main goal is to study the distribution of disease spatially. Often, public health workers are interested in identifying areas which have a higher risk for a certain infection so that resources can be allocated accordingly. Spatial epidemiology can help us to deal with such risk by taking into account population variation. Several authors have studied this risk using a

standardized mortality/morbidity ratio (SMR), which is defined as the amount of risk observed relative to what is expected under standard conditions (Broeck *et al.*, 2013). However, this is a very simple approach and SMRs can provide unstable estimate of risk due to their ratio form. Moreover, it does not accommodate the correlation between neighbors. To overcome the limitation with the SMR, most notably, Besag *et al.* (1991) proposed a convolution model that allows the relative risk to be statistically modeled by including spatially structured and unstructured random effects in the model.

A convolution model has been widely used in disease mapping by many scholars (Lawson, 2013, ch 5). Besag *et al.* (1991) first considered two additive components in their risk model: a correlated and an uncorrelated component. The correlated component (CH) was assumed to have an intrinsic conditional autoregressive (CAR) distribution. A proper CAR model has also been proposed for the CH component, and this corresponds to a proper multivariate Gaussian distribution with the full rank of covariance matrix (Cressie and Chan, 1989; Stern and Cressie, 1999). On the other hand, Moraga and Lawson (2012) have considered Gaussian Component mixture (GCM) and compared it with CAR using 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 aggregations in the data space.

Modeling scale effects is a special interest in disease mapping. When data are aggregated from finer (lower) to coarser (higher) level, obviously there will be a scaling effect. It will result in an ecological fallacy if we try to make an inference from a region to the individual located within that region without adjusting the scaling down factor. We may also find a relationship between an outcome and a covariate at the higher level, but that relationship may not retain 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 the variability of individuals' response to the diseases. In general, this scale change issue is called the modifiable areal unit problem (MAUP; Lawson, 2013, ch 9) or change of support problem in geostatistics (Cressie, 1996; Banerjee *et al.*, 2004).

To encompass scaling effects, Kolaczyk and Haung (2001) proposed a multiscale modeling approach by factorizing the likelihood into the individual components of local information. While there are other multiscale models for descriptive purpose, their method was adopted for inferential purpose. The model, which they developed, assumes that the hierarchical partitions correspond to the successive aggregation of an initial data space. Nevertheless, their approach is limited to such assumptions and it considers that the effect at the highest level is fixed and not random. In addition, it is not flexible enough to estimate the relative risks at the highest and lowest level at the same time. To overcome such issues, in this paper, we develop a multiscale modeling framework that can be used to make inference both at the higher (areas) and lower levels (subareas) simultaneously using a Bayesian models. We also evaluate the performance of the different multiscale models used via 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 choose to examine a real data set: Georgia county and public health district oral cancer incidence. This was chosen as it provides a reasonably large set of spatial unit 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 (PH) levels. There are 159 counties within 18 public health districts. These PH districts are the administrative units that provide health services. Since a public health district contains at least one county, there may be a clustering effect, i.e., counties located within the same PH district may behave similarly. The counts of disease at the public health level were created by aggregating the outcomes at the county level nested within the PH district. Hence, the data exhibit scaling effect when aggregated from the county to the PH level. This effect should be incorporated during modeling, a point on which we will elaborate in Section 3. Figure 1 depicts the Georgia map which consists of 18 PH districts and 159 counties. We can clearly see from the figure that the 159 counties are grouped into 18 public health districts. Hence, we are interested to study the grouping effect, which was occurred due to the classification of the counties into PH districts. We analyze these data in Section 4.2 and now we proceed to the multiscale modeling and a simulation study.

3 Multi-scale Modeling

In disease mapping, the information conveyed by maps varies with scale. This scale effect may need to be accommodated during modeling. Louie and Kolaczyk (2006) proposed the factorization of the likelihood that contains the information on the scaling effect in a multiscale fashion under the assumed Poisson model. They assumed a multinomial distribution for the data at finer level conditioning on coarser level. This approach is limited to the assumption of having a fixed coarser level effect. In this paper, we incorporate the scale effect using a 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.

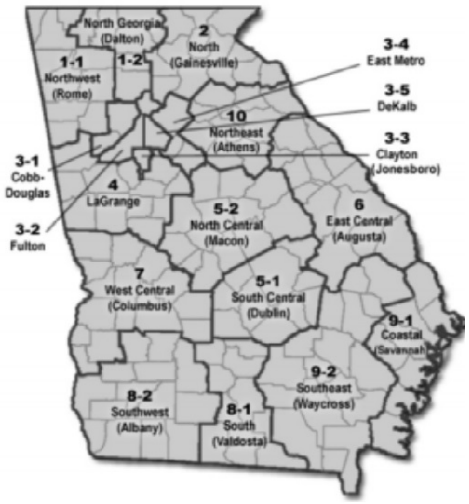

Figure 1: *State of Georgia, USA: County and PH district boundary map.*

3.1 Model 1

The framework, which was developed by Louie and Kolaczyk (2006), 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 level only. Here, we study the model both at the finer and coarser levels by including shared random effect to handle the scaling effect.

Several studies have looked at multiscale modeling of spatial data within the Bayesian framework (Louie and Kolaczyk 2006; Lee et al., 2009; Louie and Kolaczyk 2004; You and Zhou 2011). The models proposed by many of the previous studies involve complex statistical and computational techniques that may not always be easily implemented in standard softwares. 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 $\boldsymbol{y_k}$ is a vector of observed aggregated outcomes y_{ik} for unit i at the k^{th} scale level, $k = 1, \ldots, K$; $i = 1, \ldots, N_k$, where N_k is defined to be the number of units at the k^{th} level and K denotes the number of levels. Note that k is ranked from lowest to highest level. in this paper, we use the phrase lowest level to refer for the finest (e.g. census tract) while highest level to refer to the coarsest level (e.g. state). We assume that S_{ik} is the set of subunits at the $k-1$ level within the i^{th} unit at the k^{th} level. For example, for count data, we can express the aggregation as $y_{ik} = \sum_{l \epsilon S_{ik}} y_{lk}, y_{ik-1} = \sum_{l \epsilon S_{ik-1}} y_{lk-1}, \ldots$ for $k = 2, \ldots, K$. The aggregated data at each level are assumed to have a Poisson distribution, i.e., $\boldsymbol{y_k} \sim \mathrm{P}_k(\boldsymbol{\mu_k} = \boldsymbol{e_k}\boldsymbol{\theta_k})$,

where $\boldsymbol{e_k}$ is the expected rate and $\boldsymbol{\theta_k}$ is the relative risk which is given by:

$$\log(\boldsymbol{\theta_k}) = a_{0k} + v_{ik} + u_{ik}, \tag{1}$$

where a_{0k}, u_{ik}, v_{ik} are the intercept, spatially structured, and unstructured random effects for scale level k, respectively. Let $p_k(\boldsymbol{\theta_k})$ be the joint prior distribution of the p components of $\boldsymbol{\theta_k}$ for scale k, i.e., $p_k(\boldsymbol{\theta_k}) = \prod_{i=1}^{p} p_{ik}(\theta_{ik}), i = 1, \ldots, p$, and $p_k(\boldsymbol{y_k}|\boldsymbol{\theta_k}) = \prod_{i=1}^{n} p_k(y_{ik}|\theta_{ik})$ is the joint distribution of the sample which is represented as the likelihood $L(\boldsymbol{\theta_k}|\boldsymbol{y_k})$ when viewed as a function of $\boldsymbol{\theta_k}$. The posterior distribution, which is a combination of the prior distribution and the likelihood function, could be defined as:

$$P_k(\boldsymbol{\theta_k}|\boldsymbol{y_k}) \propto L_k(\boldsymbol{\theta_k}|\boldsymbol{y_k}) p_k(\boldsymbol{\theta_k}), \tag{2}$$

where $P_k(\boldsymbol{\theta_k}|\boldsymbol{y_k})$ is the posterior distribution for scale k. To obtain the posterior parameter estimates, specifically the relative risk at each scale, we can sample from the updated posterior distribution $P_k(\boldsymbol{\theta_k}|\boldsymbol{y_k})$ using an McMC sampling method. Note that the total likelihood function could be given as $L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{k=1}^{K} L_k(\boldsymbol{\theta_k}|\boldsymbol{y_k})$ (Lawson, ch 9).

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

Let $y_{i1}, i = 1, \ldots, N_1$, is the subunit level count of disease and $y_{i2} = \sum_{l \epsilon S_{i2}} y_{l1}, i = 1, \ldots, N_2$, is the i^{th} unit level count of disease aggregated at the subunits level; N_1, and N_2 are the number of subunits and units, 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 spatial structure random effect, u_{i2}. The model is given by

$$\begin{aligned} y_{i1} &\sim \text{Poisson}(e_{i1}\theta_{i1}), \\ \log(\theta_{i1}) &= a_{01} + v_{i1} + u_{i2}, \\ y_{i2} &\sim \text{Poisson}(e_{i2}\theta_{i2}), \\ \log(\theta_{i2}) &= a_{02} + v_{i2} + u_{i2}. \end{aligned} \tag{3}$$

Here, $e_{i1} = \sum y_{i1} \frac{p_{i1}}{\sum p_{i1}}$ is the expected rate at the subunit level, p_{i1} is the population size of the i^{th} subunit. The expected rate at the unit level, e_{i2}, is obtained by aggregating the expected rate at the subunit, e_{i1}, laid within the i^{th} unit. For this model and for the other subsequent models below, we have assumed a flat prior for the intercept parameters, a_{01} and a_{02}. Further, the uncorrelated heterogeneity (UH) random effects, v_{i2} and v_{i1}, were assumed to be normally distributed, i.e., $v_{i2} \sim N(0, sd_{v2}^2)$ and $v_{i1} \sim N(0, sd_{v1}^2)$, whereas the correlated heterogeneity random effect, u_{i2}, was assumed to have conditional autoregressive

(CAR) distribution of the form:

$$u_{i2} \mid u_{-i2} \sim \mathrm{N}(\bar{u}_{\delta_{i2}}, \frac{sd_{u2}^2}{n_{\delta_{i2}}}), \tag{4}$$

where

$$\bar{u}_{\delta_{i2}} = \frac{1}{n_{\delta_{i2}}} \sum_{m \epsilon \delta_{i2}} u_{m2},$$

$n_{\delta_{i2}}$ is the cardinality of δ_{i2}, which denotes the set of labels of the neighbors of unit i and u_{-i2} is the contiguous spatially structured random effect for unit i. For the hyperparameters, sd_{v1}, sd_{v2}, and sd_{u2}, we considered a uniform prior distribution, $U(0, 100)$ (Gelman *et al*, 2006).

3.2 Model 2

Model 2 is similar to Model 1 except now the spatially structured random effect at the subunit level, i.e., u_i^c is added to the model. The model can be written as

$$\begin{aligned} y_{i1} &\sim \text{Poisson}(e_{i1}\theta_{i1}), \\ \log(\theta_{i1}) &= a_{01} + u_{i1} + v_{i1} + u_{i2}, \\ y_{i2} &\sim \text{Poisson}(e_{i2}\theta_{i2}), \\ \log(\theta_{i2}) &= a_{02} + v_{i2} + u_{i2}. \end{aligned} \tag{5}$$

We assume a conditional autoregressive (CAR) distribution of the spatially structured random effect given by:

$$u_{i1} \mid u_{-i1} \sim \mathrm{N}(\bar{u}_{\delta_{i1}}, \frac{sd_{u1}^2}{n_{\delta_{i1}}}), \tag{6}$$

where

$$\bar{u}_{\delta_{i1}} = \frac{1}{n_{\delta_{i1}}} \sum_{m \epsilon \delta_{i1}} u_{m1},$$

$n_{\delta_{i1}}$ is the cardinality of δ_{i1}, which represents the set of labels of the neighbors of subunit i and $sd_{u1} \sim \mathrm{U}(0, 100)$.

3.3 Model 3

This model assumes two separate convolution models at both the subunit and unit levels. The model is of the form:

$$\begin{aligned} y_{i1} &\sim \text{Poisson}(e_{i1}\theta_{i1}), \\ \log(\theta_{i1}) &= a_{01} + u_{i1} + v_{i1}, \\ y_{i2} &\sim \text{Poisson}(e_{i2}\theta_{i2}), \\ \log(\theta_{i2}) &= a_{02} + v_{i2} + u_{i2}. \end{aligned} \tag{7}$$

Note that this model does not introduce linkage between the subunit and unit level rather it assumes they are independent.

3.4 Model 4

In this model, we assume a convolution model to calculate the relative risk at the subunit level, θ_{i1}, while the relative risk at the unit level, θ_{i2}, is simply obtained by aggregating over the subunit estimated effects. The model could be defined as:

$$\begin{aligned} y_{i1} &\sim \text{Poisson}(e_{i1}\theta_{i1}), \\ \log(\theta_{i1}) &= a_{01} + u_{i1} + v_{i1}, \\ \theta_{i2} &= \mu_{i2}/e_{i2}, \mu_{i2} = \sum_{l \epsilon S_{i2}} e_{l1}\theta_{l1}. \end{aligned} \tag{8}$$

3.5 Model Assessment and Goodness of Fit

To investigate the performance of the models, a deviance information criterion (DIC; Spiegelhalter *et al.*, 2002; Gelman *et al.*, 2004), which is a combination of the likelihood function (deviance) and model complexity (PD_{dic}; number of effective parameters), was applied to the data. We have also considered other criteria for model selection such as WAIC (Watanabe-Akaike or widely applicable information criterion; Watanabe, 2010; Gelman *et al.*, 2013) and conditional predictive ordinate (CPO; Lawson, 2013, ch 4). For a predictive accuracy assessment, mean absolute prediction error (MAPE) and mean square error prediction (MSPE) were used.

WAIC is a fully Bayesian technique for model selection and uses a posterior distribution rather than the point estimate. WAIC can be considered as computationally convenient approximations to cross validation and could be computed as:

$$\text{WAIC} = -2 * \widehat{\text{elpd}}_{\text{waic}},$$
$$\widehat{\text{elpd}}_{\text{waic}} = \widehat{\text{lpd}} - \widehat{\text{PD}}_{\text{waic}},$$

where $\widehat{\text{lpd}} = \sum_{i=1}^{n} \log(\frac{1}{S} \sum_{s=1}^{S} \text{p}(y_i|\theta^s))$, and $\widehat{\text{PD}}_{\text{waic}} = \sum_{s=1}^{S} V_{s=1}^{S} \log(\text{p}(y_i|\theta^s))$. Here, $\widehat{\text{lpd}}$, $\widehat{\text{elpd}}_{\text{waic}}$, $\widehat{\text{PD}}_{\text{waic}}$, V, n, and S denote the computed log pointwise predictive density, expected log pointwise predictive density, effective number of parameters, sample variance, the number of data points, and posterior simulations, respectively.

For all models, convergence was assessed using estimated potential scale reduction factor, $\hat{R}$, trace plots, and Brooks, Gelman and Rubin's (BGR) plots (Gelman and Rubin, 1992).

3.6 Simulation Study

A simulation study was conducted to compare the performance of the models for data generated under a hypothetical (theoretical) grid with three levels and under the real Georgia oral cancer study with two levels. Initially, gamma distributions were considered as a simple case but these were extended to spatially structured simulations for more realistic scenarios.

3.6.1 Simulation from 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 which is divided into 2^4x2^4=256 smaller areas of the finest (lower) level, 2^3x2^3=64 areas (pixels) at the medium level, and 2^2x2^2=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:

$$\begin{aligned} y_{i1} &\sim \text{Poisson}(e_{i1}\theta_{i1}), \\ y_{i2} &= \sum_{l \epsilon S_{i2}} y_{l1}, \\ y_{i3} &= \sum_{l \epsilon S_{i3}} y_{l2}, \end{aligned} \tag{9}$$

where e_{i1} is the expected rate, θ_{i1} denotes the relative risk at the finest level, S_{i2} is the set of subareas at the lower level nested within the i^{th} area at the medium level, and S_{i3} represents the set of subareas at the medium level nested within the i^{th} area at the higher level. $y_{i1}, i = 1, \ldots, N_1 = 256$, $y_{i2}, i =$

$1, \ldots, N_2 = 64$, and $y_{i3}, i = 1, \ldots, N_3 = 16$, are the samples generated at the lower, medium, and higher levels, respectively. We assumed θ_{i1} follows a gamma distribution with shape α, and scale β parameters, i.e., $\theta_{i1} \sim \text{gamma}(\alpha, \beta)$. For the hyperparameter of the gamma distribution, we assume both α and β equal to one so that the mean and variance of θ will be one. This assumption could generate relative risks that are similar to real life example. The expected rate of the 256 areas were supposed to be equal to one. Note that we have also sampled data that mimic the Georgia oral cancer study.

3.6.2 Simulation from a Convolution Model

In turn, we generated data from Model 4 (8) similar in spirit to the Georgia oral cancer study. First, the spatially structured random effects were simulated from an intrinsic conditional autoregressive (CAR) through the BRugs Package. However, with this approach, we have to use the McMC sampling method and it is computationally intensive. Hence, we sampled these random effects from CAR (6) using a program written in R software.

To sample from (6), we have to know the conditional mean and the overall variance sd_{u1}^2. To obtain the conditional mean and conditional variance, we first assumed u_{i1} follows a standard normal distribution and sd_{u1} equals to one. Then, u_{i1} was sampled from a normal distribution with conditional mean equal to the average of the contiguous u_{-i1} and conditional variance inversely proportional to $n_{\delta_{i1}}$. To cluster the random effects spatially, this process should be repeated a number of times. In our case, we repeated this process four times and obtained spatially structured random effects similar to the one generated through BRugs package. The unstructured random effects were simulated from a normal distribution with mean equal to zero and variance sd_{v1}^2. Similarly, we have simulated data from the hypothetical grid in (9), however, we used a convolution model in ((8)) to obtain the relative risk, θ_{i1}. Note that in all scenarios, the expected rates were sampled from $\gamma(1, 1)$.

The models discussed above were fitted to 200 simulated data using the Monte Carlo Markov Chain (McMC) method with 15000 samples after the first 15000 samples were discarded from the analysis. 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 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 data sets.

4 Results

4.1 Simulation Results

In this section, we describe the simulation results. First, we present the findings obtained from the simulation study that assumes the data follow a Poisson distribution with mean equal to $e_{i1}\theta_{i1}$ and the relative risk θ_{i1} itself was simulated from a gamma distribution with scale and shape parameter equal to one. Thereafter, we present the results obtained from the models fitted to the simulated data which were generated from a Poisson distribution with mean equal to $e_{i1}\theta_{i1}$, but the relative risk θ_{i1} was calculated via the convolution model in (8).

4.1.1 Results for Data Simulated from Gamma Distributions

The results obtained from the sampled data under a hypothetical grid scenario are shown in Table 1. 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 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 a slight difference at the medium level. Note that the Model 4 reveals MAPE and MSPE higher than the other models in the medium and higher levels.

Tables 2 displays the results of the data generated to mimic the Georgia oral cancer study. We have also found here that Models 1 and 2 provide smaller DIC value than Models 3 and 4, especially at the PH level. The WAIC also favors Models 1 and 2 at this level. Moreover, the MAPE, MSPE, bias, and MSE of the relative risk of these models are slightly lower than that of Models 3 and 4. However, the MAPE and MSPE at the county level are almost an identical for all the models.

4.1.2 Results for Data Simulated from a Convolution Model

The simulation results for the data generated from Model 4 using a programmed R software to sample the spatially structured random effect are shown in Tables 3-4. From Table 3, we can see that the DIC slightly opts for Model 4 at the county level, whereas it supports Model 1 at the PH level. The predictive accuracy is similar across the models except Model 4 has the smallest MSPE at the PH level. In spite of that, Model 4 provides slightly smaller MSE of the relative risk. This is an expected result because the data were generated from Model 4. On the other hand, Model 1 converges much faster than the other models. Table 4 shows the bias and MSE of the intercept a_{01} and the variance of the spatially structured

Table 1: *Simulation results for the data generated from a hypothetical grid. Lower, medium, and higher represent the three levels; DIC, MAPE, MSPE,* $\theta_{i1} \sim gamma(\alpha, \beta)$*, 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.*

Models	PD_{dic}			DIC			MAPE			MSPE			θ_1		CT
	lower	medium	higher	lower	medium	higher	lower	medium	higher	lower	medium	higher	bias	MSE	elapsed
Model 1	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**
Model 2	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
Model 3	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
Model 4	36.26	-	-	713.78	-	-	0.91	2.53	7.86	2.18	11.37	86.84	-0.46	0.98	484.48

Table 2: *Simulation results for the data generated from a Poisson distribution with mean* $e_i^c\theta_i^c$ *that mimic the Georgia oral cancer study, where* $\theta_{i1} \sim gamma(\alpha, \beta)$*. WAIC,* PD_{dic}*, and* PD_{waic} *represent the average widely applicable information criterion, and the effective number of parameters used to calculate DIC and WAIC over the 200 data sets, respectively.*

Models	PD_{dic}		DIC		PD_{waic}		WAIC		MAPE		MSPE		θ_1		CT
	county	PH district	county	PH district	county	PH district	county	PH district	county	PH district	county	PH district	bias	MSE	elapsed
Model 1	46.39	9.75	**348.91**	**87.68**	36.27	5.34	349.02	**84.94**	0.80	**2.91**	2.07	**16.94**	**0.008**	**1.41**	**69.13**
Model 2	47.44	10.06	348.92	88.27	36.97	5.58	348.98	85.57	0.79	2.92	2.06	17.12	0.009	1.44	203.18
Model 3	49.95	10.98	349.76	92.01	37.96	7.22	348.86	90.65	0.79	3.04	2.05	18.11	0.01	1.46	129.02
Model 4	49.86	-	349.63	-	37.85	6.96	**348.67**	89.31	0.79	3.00	2.05	17.86	0.01	1.44	150.79

and unstructured random effects, sd_{u1} and sd_{v1}, respectively. Model 4 yields a tiny bias of the overall relative risk ($\exp(a_{01})$), whereas Model 3 produces smallest bias of sd_{u1} and sd_{v1}. Moreover, both Models 3 and 4 have smaller MSE of a_{01} and sd_{u1}. Yet, Model 1 slightly compact MSE of sd_{v1}.

When data are generated from Model 4 with assumed values of $a_{01} = -2$, sd_{u1}=1, sd_{v1}=1, and simulated u_{i1} through BRugs package, Model 4 has the smallest DIC at the county level, whereas Models 1 and 2 have slightly smaller DIC at PH level (Table 5). The MAPE and MSPE at the PH level prefer Models 1 and 2 though there is no quite difference among the models at the county level. The bias and MSE of Models 1 to 3 are close to each other. Still, Model 1 converges faster. Model 4 has a little bias of a_{01}, sd_{u1}, and MSE of sd_{u1}, while the Model 3 has a smaller bias and MSE of sd_{v1}. Until now, we have seen the results for the simulated data assuming the spatially structured and unstructured random effects different for all the 200 data sets. To investigate the effects of these random variables, we have also generated data assuming the spatially structured random effect the same for all the 200 data sets (see the results in Appendixes A and B).

Table 3: *Simulation results for the data generated similar in structure to Georgia oral cancer data. The spatially structured random effects were simulated using programmed R software and assumed to be different for all the 200 data sets.*

Models	assumed values			PD_{dic}		DIC		MAPE		MSPE		θ_1		CT
	a_{01}	sd_{u1}	sd_{v1}	county	PH district	county	PH district	county	PH district	county	PH district	bias	MSE	CT
Model 1	0.1	1	1	70.06	13.12	448.08	**103.69**	1.14	4.24	4.31	37.69	-0.01	6.44	**50.51**
Model 2	0.1	1	1	71.12	13.37	446.67	104.18	1.13	4.25	4.25	37.85	-0.01	6.49	103.20
Model 3	0.1	1	1	72.45	14.31	446.18	106.41	1.13	4.29	4.17	38.07	-0.01	6.43	100.04
Model 4	0.1	1	1	69.50	-	**443.66**	-	**1.11**	**4.11**	**3.93**	**34.51**	-0.01	**5.31**	153.25

Table 4: *Simulation study. Summary of the bias and MSE of the parameters for the data generated similar in structure to Georgia oral cancer study. The spatially structured random effects were simulated using a programmed R software and assumed to be different for all the 200 data sets.*

Models	assumed values			bias			MSE		
	a_{01}	sd_{u1}	sd_{v1}	a_{01}	sd_{u1}	sd_{v1}	a_{01}	sd_{u1}	sd_{v1}
Model 1	0.1	1	1	-0.053	-	0.038	0.043	-	**0.029**
Model 2	0.1	1	1	-0.064	-0.451	-0.028	0.043	0.338	0.037
Model 3	0.1	1	1	-0.029	**-0.236**	**-0.011**	**0.029**	**0.246**	0.052
Model 4	0.1	1	1	**-0.014**	-0.316	-0.057	**0.029**	0.249	0.034

Table 5: *Simulation results for the data generated similar in structure to Georgia oral cancer study. The spatially structured random effects were simulated through BRugs Package and assumed to be different for all the 200 data sets. The assumed values were as follows:* a_{01}*=-2,* sd_{u1}*=1, and* sd_{v1}*=1.*

Models	PD_{dic}		DIC		MAPE		MSPE		θ_1		CT
	county	PH district	county	PH district	county	PH district	county	PH district	bias	MSE	CT
Model 1	20.45	7.22	163.11	**55.72**	0.29	**1.23**	0.45	**3.52**	0.003	0.16	**182.46**
Model 2	21.62	7.39	163.11	**55.72**	0.29	1.24	0.45	3.56	0.004	0.17	352.09
Model 3	22.87	7.73	164.55	59.26	0.29	1.31	0.45	3.82	0.003	0.17	387.68
Model 4	21.46	-	**156.45**	-	0.28	1.34	0.47	4.02	0.013	0.21	429.71

Table 6: *Simulation study. Summary of the bias and MSE of the parameters for the data generated similar in structure to Georgia oral cancer data. The spatially structured random effects were sampled through BRugs Package and assumed to be different for all the 200 data sets.*

Models	assumed values			bias			MSE		
	a_{01}	sd_{u1}	sd_{v1}	a_{01}	sd_{u1}	sd_{v1}	a_{01}	sd_{u1}	sd_{v1}
Model 1	-2	1	1	-0.199	-	-0.165	0.225	-	0.154
Model 2	-2	1	1	-0.257	-0.366	-0.217	0.287	0.292	0.163
Model 3	-2	1	1	-0.176	-0.247	**-0.087**	0.200	0.276	**0.128**
Model 4	-2	1	1	**-0.13**	**-0.14**	-0.16	**0.16**	**0.25**	0.17

Table 7 gives the results of the model fitted to the simulated data from the hypothetical grid. The DIC favors Models 1 and 2 at all levels, especially at lower and medium levels. In addition, the WAIC prefers Model 2 at the lower and higher levels, whereas the Model 1 at the medium level. Note that these results are more pronounced at the medium level. Further, using MAPE and MSPE, the predictive accuracy computed from Models 1 and 2 is slightly better than that of Models 3 and 4 at the medium level.

The posterior mean estimate and precision are displayed in Table 8. Here also, Model 1 provides better mean and precision estimates of the relative risk. On the other hand, the bias and MSE of the sd_{u1} and sd_{v1} obtained from Models 3 and 4 are smaller than that of Models 1 and 2. However, all the models produce similar bias and MSE of the structured (u_{i1}) and unstructured random effects (v_{i1}). Similar findings can be observed in the data generated by allowing the correlated and uncorrelated random effects to be constant for all the 200 data sets (Appendix B;Tables B.3 and B.4).

Table 7: *Simulation results for data generated from a hypothetical grid. 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:* a_{01}*=0.1,* sd_{u1}*=1, and* sd_{v1}*=1.*

Models	PD_{dic}			DIC			PD_{waic}			WAIC			MAPE			MSPE			CT
	lower	medium	higher	lower	medium	higher	lower	medium	higher	lower	medium	higher	lower	medium	higher	lower	medium	higher	elapsed
Model 1	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 2	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 3	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 4	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

Table 8: *Simulation study. Summary of the bias and MSE of the parameters for the data generated from a hypothetical grid. The spatially structured and unstructured random effects were generated using a programmed R software and were assumed different for all the 200 data sets.*

Models	assumed values				bias					MSE					
	a_{01}	sd_{u1}	sd_{v1}	θ_1	a_{01}	sd_{u1}	sd_{v1}	u_1	v_1	θ_1	a_{01}	sd_{u1}	sd_{v1}	u_1	v_1
Model 1	0.1	1	1	**0.006**	**0.009**	-	0.24	-	0.008	**4.499**	0.022	-	0.081	-	**1.022**
Model 2	0.1	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
Model 3	0.1	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
Model 4	0.1	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

The effective number of parameters (PD), DIC, Deviance, and MSPE at the three different levels for the simulated data from a hypothetical grid are displayed in Figure 2. We can clearly see that 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. The results of the DIC for all the 200 data sets are shown in Figure 3. Here also, Models 1 and 2 outperform the other models. Note that the results of the DIC and WAIC are close to each other as they should be. However, there is not much difference in the MSPE among the models and it tends to increase as the aggregation change from lower to higher level. Note that the MSPE at the medium level (e.g. MSPE is 15.14 for Model 1, see Appendix, Tables B.3) is approximately four times to the MSPE at the lower level (4.06). Moreover, the MSPE at the higher level (64.76) is almost four times to that of at the medium level and sixteen times to the MSPE at the lower level. This is may be due to the sample size because the sample size at the lower level (256) is four times of the sample size at the medium level (64) and sixteen times to that of at the higher level (16). Note that this is not true for the simulated data that mimic the Georgia oral cancer study because we have an irregular shape of the Georgia state.

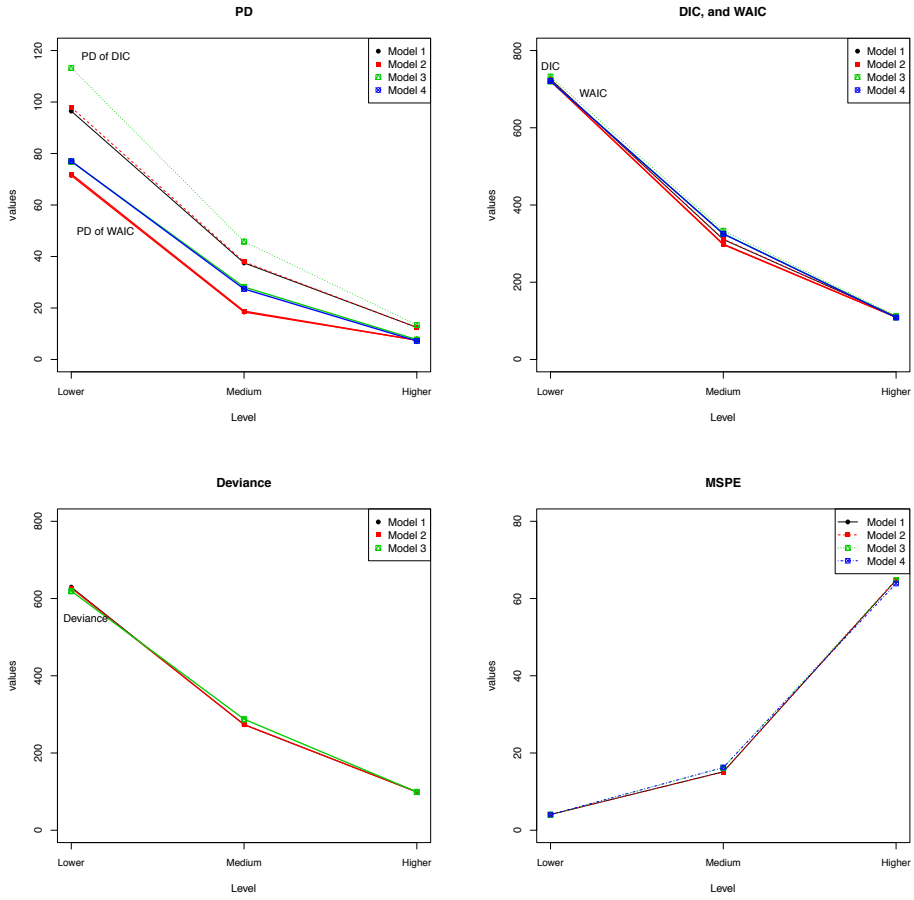

Figure 2: *PD, DIC, WAIC, Deviance, and MSPE for the data generated from a hypothetical grid with spatially structured (u_{i1}) and unstructured random effect (v_{i1}) assumed to be constant over all the 200 data sets.*

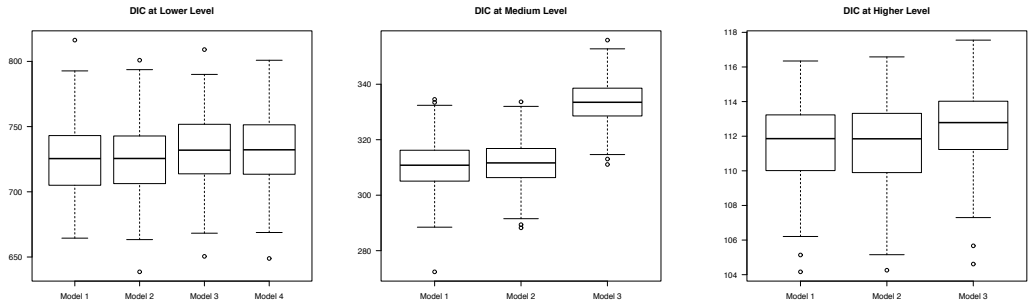

Figure 3: *Deviance information criterion (DIC) values for the 200 simulated data from a hypothetical grid with spatially structured (u_{i1}) and unstructured random effects (v_{i1}) assumed to be constant over all the 200 data sets.*

4.2 Application to Data

To assess the benefit of the shared random effect models in real life example, we have applied the models discussed above to the Georgia oral cancer study. The estimated potential scale reduction factor, trace, and BGR plots indicate good convergence for all model parameters. The results of the model fit and predictive accuracy are shown in Table 9. We can clearly see that 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 the Model 1 is better than the other models. However, the predictive accuracy at the county level is not quite different between the models. If we do not take into account for model complexity, the model which introduces two separate convolution models (Model 5) provides better model fit (Deviance) at the county level compared to the shared random effect models (Models 1 and 2). According to the CPO, which does not also 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 5).

The posterior summary statistics are given in Table 10. The overall relative risks of all the models, ($\exp(a_{01})$ and $\exp(a_{02})$), is approximately equal to one with 95 % credible interval in the range of 0.85 and 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 compared to what is expected. However, this is not true for each county and PH district (Figure 4). Here, the relative risks obtained from Model 6 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 6 at the PH level and the other models. For instance, Model 6 indicates an elevated risk in the northwest part of the Georgia state, whereas the other models show an elevated risk in the southeast part. This is may be due to the correlation between the neighbors is not accommodated in Model 6 at the PH level. 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 extra-variability in the data (Table 11.

On the other hand, the variability of the unstructured random effects (sd_{v1} and sd_{v2}) obtained from the models are similar. Nevertheless, the variability of the structured random effect (sd_{u1}) from Model 2 is smaller than that of Model 3. This is may be due to some part of this variability in Model 2 is accommodated by the shared random effect (sd_{u2}=0.474), which is higher than the variance in Model 3 (sd_{u2}=0.411).

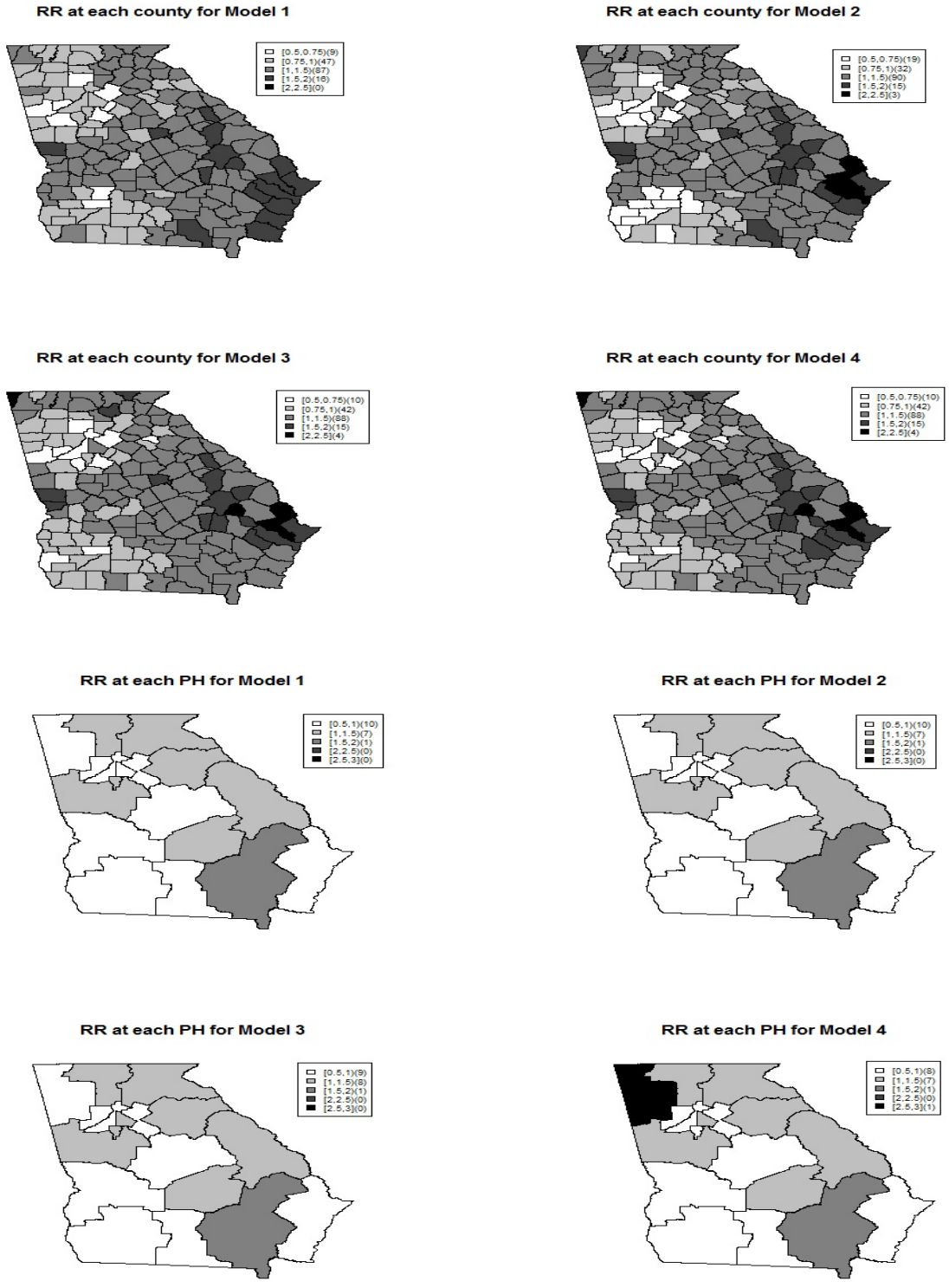

Figure 4: *Georgia oral cancer data. Relative Risk (RR) at each county and public health (PH) district.*

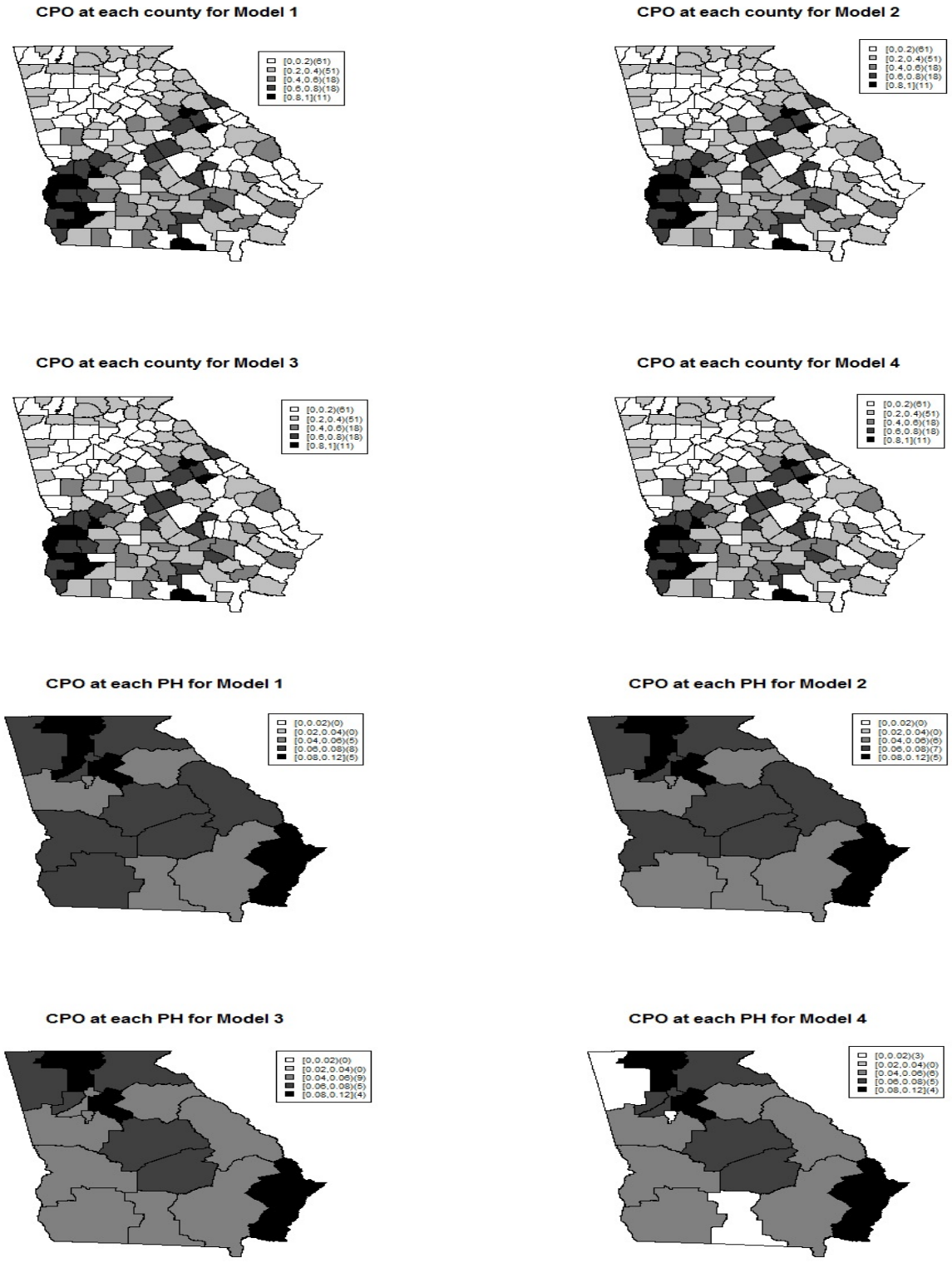

Figure 5: *Georgia oral cancer study. Conditional predictive ordinates (CPO) at each county and public health (PH) district.*

Table 9: *Model fit and predictive accuracy results for Georgia oral cancer data.*

Models	PD_{dic}		DIC		PD_{waic}		WAIC		MAPE		MSPE		Deviance	
	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**
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
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
Model 4	32.63	-	485.38	-	27.53	-	485.78	-	1.36	8.28	4.83	161.9	452.75	-

Table 10: *Georgia oral cancer data. Posterior mean estimates, standard error, and 95% Credible Interval.*

Models	Mean						sd						95%CI					
	a_{01}	a_{02}	sd_{u1}	sd_{v1}	sd_{u2}	sd_{v2}	a_{01}	a_{02}	sd_{u1}	sd_{v1}	sd_{u2}	sd_{v2}	a_{01}	a_{02}	sd_{u1}	sd_{v1}	sd_{u2}	sd_{v2}
Model 1	0.003	-2.34E-5	-	0.256	0.492	0.082	0.065	0.059	-	0.122	0.139	0.064	(-0.131,0.125)	(-0.118,0.116)	-	(0.046,0.495)	(0.265,0.808)	(0.002,0.239)
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	(-0.126,0.164)	(-0.141,0.126)	(0.083,0.674)	(0.027,0.458)	(0.226,0.810)	(0.007,0.249)
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	(-0.155,0.125)	(-0.121,0.118)	(0.219,0.865)	(0.053,0.495)	(0.099,0.806)	(0.003,0.377)
Model 4	0.027	-	0.539	0.237	-	-	0.075	-	0.181	0.109	-	-	(-0.126,0.169)	-	(0.191,0.896)	(0.057,0.472)	-	-

Table 11: *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

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 using shared random effect multiscale models (Models 1 and 2). These models provide not only a better fit to the data, but also they produce a better predictive accuracy as compared to the independent convolution model (Model 5) and Model 4, especially at the coarser level. This is an expected result because the effect of the coarsest level is inherited into the finer level through the shared random effect. Furthermore, we obtained an unbiased estimate of the relative risks. The parsimonious shared random effect model (Model 1) also converges faster than the other models.

The simulation results indicate that the models with shared random effects are the best model when data are simulated from a Poisson distribution with mean equal to $e_{i1}\theta_{i1}$ in which e_{i1} and θ_{i1} are generated from a gamma distribution with the shape and scale parameters equal to one. These results are more pronounced at the PH level of the simulated data that mimic the Georgia oral cancer study, while at the lower and medium levels for the data generated from a hypothetical grid. Moreover, at the PH and medium levels, the predictive ability of the shared random effect models is better than the other models. We have also found that the bias and MSE of the relative risks computed from these models are lower than that of Models 3 and 4.

For data drawn from Model 4 with spatially structured and unstructured random effects assuming random for all the simulated data sets, the DIC favors the shared random effect models at the PH level, while Model 4 at the county level. However, it advocates the shared random effect models at the county level when the data are sampled from Model 4 with constant spatially correlated and uncorrelated random effects for all the simulated data sets. Here, Model 5, which uses an independent model at the county and PH levels, fits the simulated data as good as the shared random effect models. On the other hand, the DIC for the shared random effect models at the lower and medium levels for the data generated from a hypothetical grid is much lower than that 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 simulated overall relative risk, spatially structured (u_{i1}), and unstructured (v_{i1}) random effects. Nevertheless, Models 3 and 4 yield more unbiased and precise estimate of the variance of the spatially structured (sd_{u1}) and unstructured (sd_{v1}) random effects in most of the cases.

To investigate the results obtained from the simulation study in a real example, we have implemented the models to the Georgia oral cancer study. 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 simple multiscale models, Models 3 and 4, at the county level, while Models 1 and 2 at the PH level.

Multiscale Modeling has been studied by several researchers (Kolaczyk and Haung, 2001; Louie and Kolaczyk, 2006). Our approach of multiscale modeling is different from Louie and Kolaczyk (2006) in many aspects. First, these authors assume a multinomial distribution at the finer level conditioning on 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 coarsest level is inherited into the finer level. Second, their method does not encompass the spatially structured random effect that handles the correlation between the neighbors, whereas our methods do so. Moreover, our convolution multiscale models could be implemented easily in standard softwares such as WinBUGS and INLA.

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 is may be due to the shared random effect that was inherited from the coarsest 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, which is the most parsimonious model, is slightly better than the independent convolution model. These results indicate that there should be a linkage to account for scaling effect between the finer and coarser level. We conclude that sharing the random effect between these two levels improves the model fit, predictive accuracy, and estimation, and efficiency of the relative risks.

Though we have achieved better results by including shared random effect into the model, our paper has some limitations. First, we introduced the shared random effect through the spatially structured random effect. Currently, we are investigating using shared unstructured random effects to accommodate a scaling 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 scaling effect using correlation structures between the finer and coarser level is planned. Furthermore, when an interest arises to measure a relation between an outcome and a covariate, say, if the relationship at the finer level will hold true at the coarser level, our multiscale model could be easily extended to account for such issues. This is also in our plan.

6 Acknowledgments

The authors would like to acknowledge support from the Nation Institutes of Health via grant R01CA172805.

References

Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. London: Chapman and Hall/CRC Press.

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

Broeck, J., Brestoff, J. R., and Kaulfuss, C. (2013). "Statistical Estimation". Epidemiology: *Principles and Practical Guidelines*. p. 417.

Cressie, N.A.C. and Chan, N.H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, **84**, 393–401.

Cressie, N.A.C. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, **2**, 83–101.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**, 457-511.

Gelman, A., Carlin, J.B., Stern, H.S, and Rubin, D.B. (2004). *Bayesian Data Analysis*. New York: Chapman and Hall.

Gelman, A., Hwang, J., and Vehtari, A. (2013). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*. Arxiv.

Kolaczyk, E.D., Huang, H. (2001). Multiscale statistical models for hierarchical spatial aggregation. *Geographical Analysis* , **33**, 95-118.

Lawson, A.B. (2013). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. New York: Chapman and Hall/CRC Press.

Lee, S., Yeatts, K., and Serre, M. (2009). A Bayesian Maximum Entropy approach to address the change of support problem in the spatial analysis of childhood asthma prevalence across North Carolina. *Spatial and Spatio-temporal Epidemiology*, **25**, 49–60.

Louie, M.M, Kolaczyk, E.D. (2004). On the covariance properties of certain multiscale spatial processes. *Statistics, Probability Letters* , **66(4)**, 407–416.

Louie, M., and Kolaczyk, E. (2006). Multiscale detection of localized anomalous structure in aggregate disease incidence data. *Statistics in Medicine*, **25**, 787–810.

Louie, M., and Kolaczyk, E. (2006) A multiscale method for disease mapping in spatial epidemiology. *Statistics in Medicine*, **25**, 1287-1306

Moraga, P. and Lawson, A.B. (2012). Gaussian component mixtures and CAR models in Bayesian disease mapping. *Computational Statistics and Data Analysis*, **56**, 1417–1433.

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 Royal Statistical Society, Series B*, **64**, 583-616.

Stern, H.S. and Cressie, N.A. (1999). Inference for extremes in disease mapping. In Disease mapping and risk assessment for public health. A. Lawson, A. Biggeri, D. Bohning, E. Lesaffre, J-F. Viel and R. Bertollini (eds.), Chichester: Wiley, p. 63–84.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, **11**, 3571-3594.

You, Y., and Zhou, Q. (2011). Hierarchical Bayes small area estimation under a spatial model with application to health survey data. *Survey Methodology* , **37**, 25–37.

APPENDIX

In this appendix, we present the supplementary material for the simulation study. Section A provides the simulation results for the data generated assuming the spatial structured random effect is the same for all 200 data sets while the unstructured random effect is allowed to be random for all the 200 data sets. Section B illustrates the results for the simulated data assuming both the structured and unstructured random effect will be constant across the 200 data sets.

A Fixing the spatially structured random effect

To compare the effects of the spatially structured random effect, we have also simulated data from Model 4 by fixing u_{i1}, i.e., assuming the same u_{i1} for all the 200 data sets. Here, the Model 2 produces slightly smaller DIC at the county level, while Model 1 at the PH level. Additionally, the predictive accuracy at the PH district level slightly advocates Model 1, whereas it is almost an identical at the county level. The bias and MSE of the relative risk at the county level is also similar for all the models except the MSE computed from Model 4 is slightly larger than the MSE obtained from the other models (Table A.1). On the other hand, Model 1 has the lowest bias, MSE of a_{01}, and sd_{u1}. But, Model 2 provides tiny bias and MSE of sd_{v1}. Note that the Model 3 has a relatively high bias and MSE of sd_{u1} (Table A.2).

Table A.1: *Simulation results for the data generated similar in structure to Georgia oral cancer data. The spatially structured random effects were generated using a programmed R software and were assumed constant, while the unstructured random effects were assumed different for all the 200 data sets. The assumed values were as follows:* a_{01}*=0.1,* sd_{u1}*=1, and* sd_{v1}*=1.*

Models	PD_{dic}		DIC		MAPE		MSPE		θ_1		CT
	county	PH district	county	PH district	county	PH district	county	PH district	bias	MSE	elapsed
Model 1	64.26	12.00	422.05	**98.37**	1.04	3.81	3.49	**30.46**	0.03	**4.16**	**184.76**
Model 2	66.81	13.32	**421.67**	101.71	1.03	3.89	**3.43**	31.18	0.03	4.22	346.65
Model 3	65.01	12.21	422.14	98.80	1.04	3.82	3.49	30.60	0.03	4.22	352.99
Model 4	67.91	-	425.11	-	1.05	3.89	3.57	31.42	0.04	4.56	251.84

Table A.2: *Simulation study. Summary of the bias and MSE of the parameters for the data generated similar in structure to Georgia oral cancer data. The spatially structured random effects were generated using a programmed R software and were assumed constant, while the unstructured random effect were assumed different for all the 200 the data sets.*

Models	assumed values			bias			MSE		
	a_{01}	sd_{u1}	sd_{v1}	a_{01}	sd_{u1}	sd_{v1}	a_{01}	sd_{u1}	sd_{v1}
Model 1	0.1	1	1	-0.226	-	**0.017**	0.083	-	**0.021**
Model 2	0.1	1	1	-0.147	-0.349	-0.036	0.043	0.256	0.034
Model 3	0.1	1	1	-0.231	-0.519	-0.075	0.084	0.368	0.038
Model 4	0.1	1	1	**-0.137**	**-0.209**	-0.053	**0.039**	**0.187**	0.039

B Fixing both the spatially structured and unstructured random effect

Furthermore, we have sampled data assuming the same u_{i1} and v_{i1} for all the 200 data sets and results are displayed in Table B.1. Here, the DIC is almost the same at the county level, while it tends to select slightly model 1 at the PH level. Nevertheless, the MAPE and MSPE slightly favor Model 4 at the PH level, whereas all the models have similar MAPE and MSPE at the county level. The bias and MSE of the relative risk are homogenous for all the models, indicating that all the models produce similar estimation and precision of the relative risk if both the structured and unstructured random effects are allowed to be constant among the data sets. Using the CPO method, there is not much difference among the models (Figures B.6). Moreover, the bias of the structured u_{i1} and unstructured v_{i1} random effects is almost an identical for all the models (Table B.2). The spatially structured and unstructured random effects obtained from all the models (Figure B.8) are similar to the simulated one (Figure B.7). In spite of that, the bias and MSE of the overall relative risk and sd_{u1} obtained from Models 3 and 4 are smaller than that of Models 1 and 2. Model 2 has the smallest bias of sd_{v1}, whereas the Model 1 reveals lower MSE of sd_{v1}.

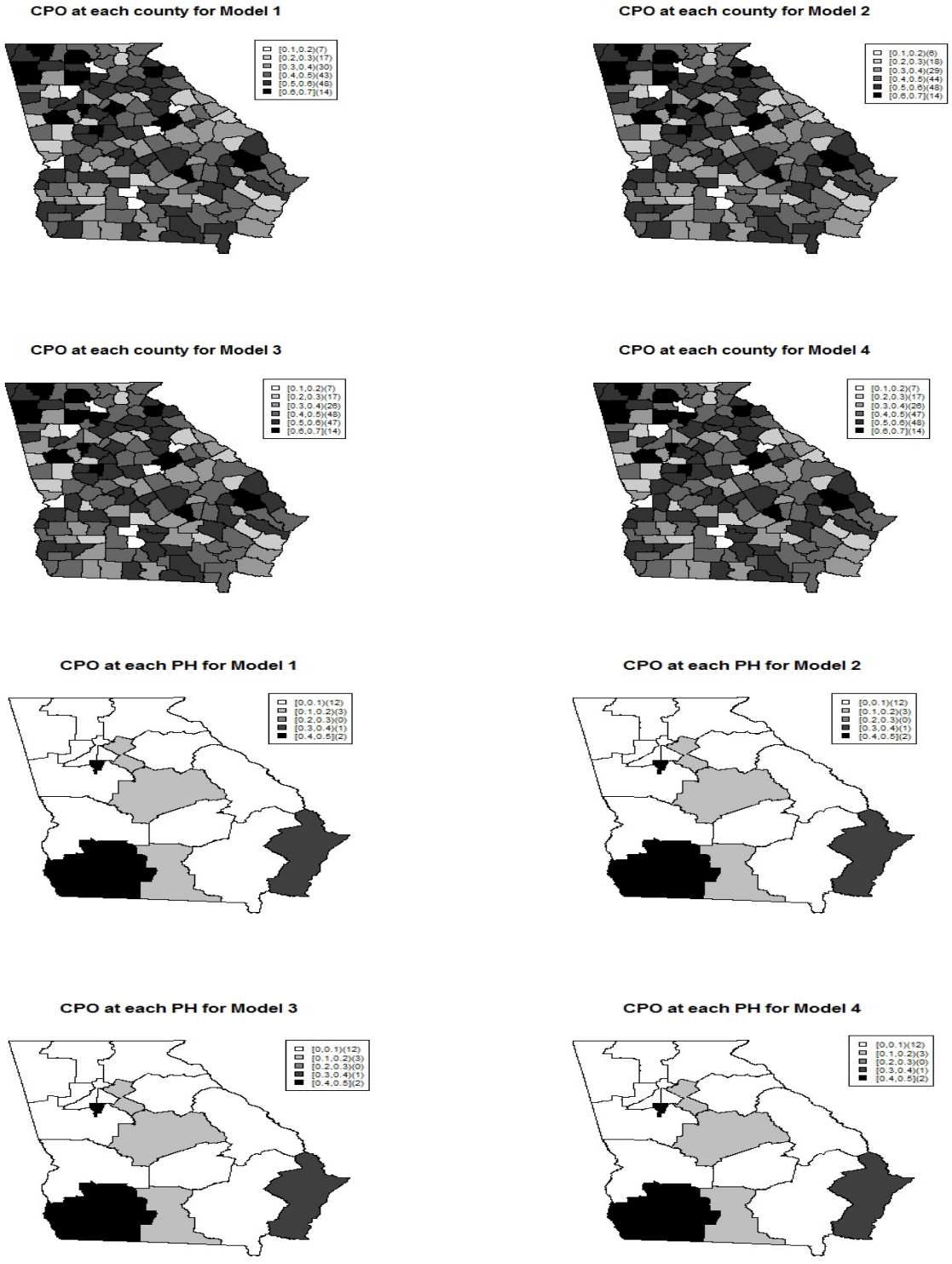

Figure B.6: *Simulation Study. Conditional predictive ordinates (CPO) at each county and public health (PH) district.*

Table B.1: *Simulation results for data generated similar in structure to Georgia oral cancer data. The spatially structured and unstructured random effects were assumed constant for all the 200 data sets. The assumed values were:* a_{01}*=0.1,* sd_{u1}*=1, and* sd_{v1}*=1.*

Models	PD_{dic}		DIC		PD_{waic}		WAIC		MAPE		MSPE		θ_1		CT
	county	PH district	county	PH district	county	PH district	county	PH district	county	PH district	county	PH district	bias	MSE	CT
Model 1	40.79	9.21	452.39	**103.64**	49.81	7.89	446.07	**101.01**	1.14	4.21	4.11	36.61	0.04	**5.59**	**203.24**
Model 2	72.86	13.14	452.48	**103.64**	49.99	8.03	446.03	101.34	1.14	4.22	4.11	36.79	0.04	5.69	279.19
Model 3	74.09	14.25	**451.54**	106.00	49.32	8.89	**442.96**	103.96	1.14	4.27	**4.06**	36.93	0.04	5.65	320.00
Model 4	74.17	-	**451.56**	-	49.33	7.77	**442.93**	100.71	1.14	4.19	**4.06**	**35.50**	0.04	5.65	338.89

Table B.2: *Simulation study. Summary of the bias and MSE of the parameters for the data generated similar in structure to Georgia oral cancer data. The spatially structured and unstructured random effects were assumed constant for all the 200 data sets.*

Models	assumed values			bias					MSE				
	a_{01}	sd_{u1}	sd_{v1}	a_{01}	sd_{u1}	sd_{v1}	u_1	v_1	a_{01}	sd_{u1}	sd_{v1}	u_1	v_1
Model 1	0.1	1	1	-0.101	-	0.053	-	**-0.001**	0.031	-	**0.014**	-	1.313
Model 2	0.1	1	1	-0.109	-0.494	**-0.004**	-0.011	**-0.001**	0.033	0.341	0.020	**0.219**	**1.311**
Model 3	0.1	1	1	**0.003**	**-0.416**	0.057	-0.011	-0.003	**0.014**	**0.263**	0.023	0.236	**1.317**
Model 4	0.1	1	1	**0.003**	-0.435	0.059	-0.011	-0.003	**0.015**	0.290	0.026	0.230	1.319

Table B.3: *Simulation results for data generated from a hypothetical grid. The spatially structured and unstructured random effects were generated using a programmed R software and were assumed constant for all the 200 data sets. The assumed values were as follows:* a_{01}*=0.1,* sd_{u1}*=1, and* sd_{v1}*=1.*

Models	PD_{dic}			DIC			PD_{waic}			WAIC			MAPE			MSPE			CT
	lower	medium	higher	lower	medium	higher	lower	medium	higher	lower	medium	higher	lower	medium	higher	lower	medium	higher	elapsed
Model 1	96.47	37.50	12.47	726.05	**310.77**	111.73	**71.45**	18.42	7.51	721.72	**297.57**	109.46	1.17	**2.73**	6.24	4.06	15.14	64.76	**277.79**
Model 2	97.95	37.89	12.42	**725.04**	311.74	111.62	72.03	18.79	7.49	720.87	298.66	109.43	1.16	2.74	6.24	4.09	**15.07**	64.91	534.11
Model 3	113.28	45.82	13.40	732.66	333.41	112.56	76.92	28.14	7.71	720.73	325.59	109.76	1.15	2.88	6.23	4.04	16.20	64.79	564.95
Model 4	113.34	-	-	733.78	-	-	77.03	27.38	7.18	720.90	325.27	108.58	1.15	2.89	6.19	4.06	16.26	**63.88**	690.84

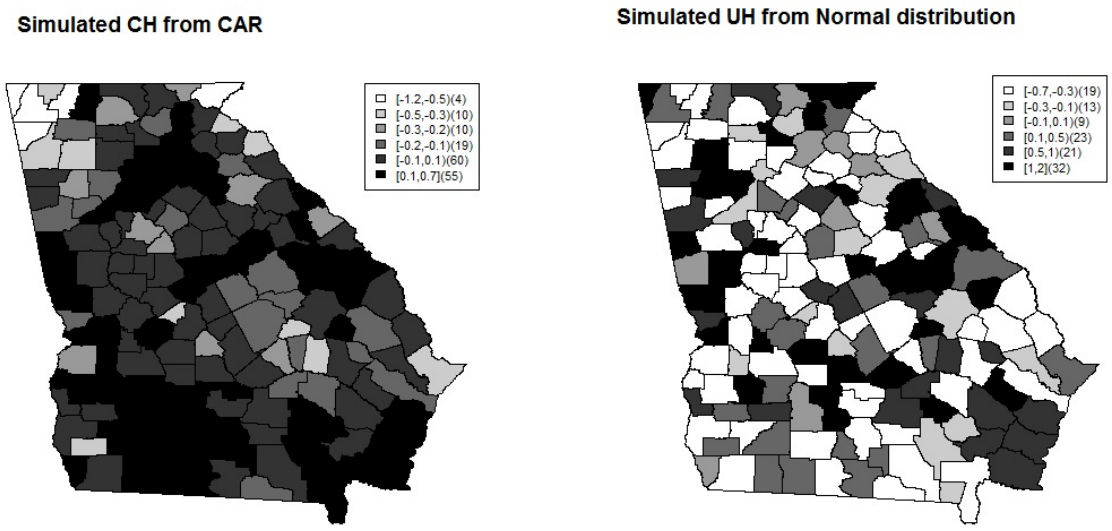

Figure B.7: *The simulated correlated heterogeneity (CH) from a conditional autoregressive (CAR) and simulated uncorrelated heterogeneity (UH) random effects from a normal distribution.*

Table B.4: *Simulation study. Summary of the bias and MSE of the parameters for the data generated from a hypothetical grid. The spatially structured and unstructured random effects were generated using a programmed R software and were assumed constant for all the 200 data sets.*

Models	assumed values				bias						MSE				
	a_{01}	sd_{u1}	sd_{v1}	θ_1	a_{01}	sd_{u1}	sd_{v1}	u_1	v_1	θ_1	a_{01}	sd_{u1}	sd_{v1}	u_1	v_1
Model 1	0.1	1	1	**0.011**	0.075	-	0.242	-	-0.107	**4.707**	0.014	-	0.068	-	**1.022**
Model 2	0.1	1	1	0.014	**0.069**	-0.679	-0.257	0.019	-0.107	4.737	0.013	0.517	0.079	**0.109**	**1.022**
Model 3	0.1	1	1	0.022	0.072	**0.216**	-0.078	0.019	-0.107	4.917	0.013	**0.240**	0.042	0.226	1.084
Model 4	0.1	1	1	0.021	0.073	-0.444	**0.041**	0.019	-0.107	4.86	0.013	0.299	**0.016**	0.234	1.082

Figure B.8: *The average spatially correlated heterogeneity (CH) and uncorrelated heterogeneity (UH) random effects obtained from the models fitted to 200 data sets.*