Bayesian Multiscale Modeling for Aggregated Disease Mapping Data*

Mehreteab Aregay
Department of Public Heath Sciences
Division of Biostatistics and Bioinformatics
MUSC, Charleston USA
aregay@musc.edu

Andrew B. Lawson
Department of Public Heath Sciences
Division of Biostatistics and Bioinformatics
MUSC, Charleston USA
lawsonab@musc.edu

Christel Faes
Department of Mathematics and Statistics
Hasselt University
Hasselt, Belgium

Russell Kirby
Department of Community and Family Health
University of Southern Florida
Lakeland, USA

ABSTRACT

In spatial epidemiology, 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 three 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 separate convolution models at the finer and coarser levels. All these models were compared based on deviance information criterion (DIC), Watanabe-Akaike or widely applicable information criterion (WAIC) and predictive accuracy applied on real and simulated data. The results indicate that the models with a shared random effect outperform the other models.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Statistical computing, Spatial data analysis;

General Terms

Theory, Measurement, Algorithm

Keywords

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

1. INTRODUCTION

In disease mapping (spatial epidemiology), the main goal is to study the distribution of disease spatially. Often, public health officers are interested to identify areas which have a higher risk for a certain infection. Several authors have studied this risk using a standardized mortality/morbidity ratio (SMR). However, this is a very simple approach and it does not accommodate the correlation between neighbors. To overcome the limitation with SMR, Besag *et al.* (1991) [1] proposed a convolution model that allows the relative risk to be statistically modeled by including spatially structured and unstructured random effects into the model. Even though the convolution model has been widely used in spatial epidemiology, it does not accommodate the spatial scaling effect associated to the aggregations of the data space.

Scaling is a special interest in disease mapping. To encompass this effect, Kolaczyk and Haung (2001) [3] have developed a multiscale modeling approach by factorizing the likelihood into individual components of local information. The model, which they developed, assumes that the hierarchical partitions correspond to successive aggregation of an initial data space. Nevertheless, their approach assumes the effects at the higher level are fixed not random. In addition, it is not flexible enough to make inference both at the higher and lower level at the same time. To overcome such issues, in this paper, we propose a multiscale modeling that can be used to make inferences both at the higher (areas) and lower levels (subareas) simultaneously using a Bayesian convolution model. We also evaluate the performance of the different multiscale models based on a simulation study.

2. GEORGIA ORAL CANCER DATA

These data consist of the number of persons discharged from non-Federal acute-care inpatient facilities for illness in 2008 at both the county and public health (PH) level. There are 159 counties nested 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 grouping effect, i.e., counties located within the same PH district may behave similarly. Figure 1 shows the Georgia map which consists of 18 PH districts and 159 counties. We can clearly see from the figure that the 159 counties are divided 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 for providing health services.

*

HealthGIS'14 November 04-07 2014, Dallas/Fort Worth, TX, USA

http://dx.doi.org/10.1145/2676629.2676640.

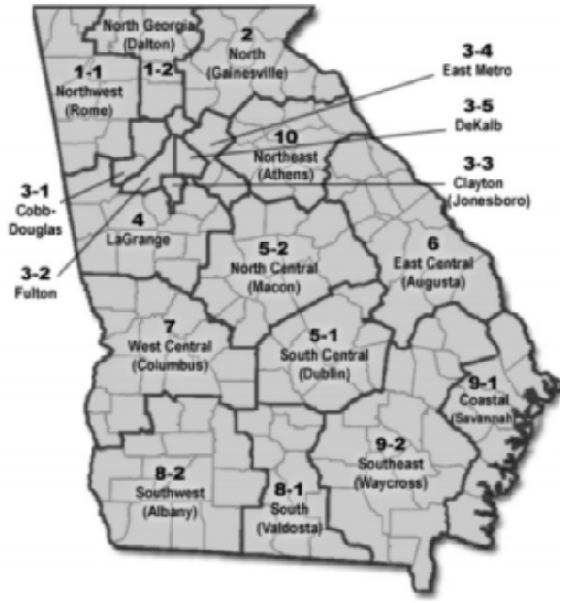

Figure 1: State of Georgia, USA: County and PH district boundary map.

3. MULTISCALE MODELING

In disease mapping, the information conveyed by the maps varies with scale. Hence, this scale effect should be accommodated during modeling. Louie and Kolaczyk (2006) [4] proposed to factorize the likelihood which contains the information of the scaling effect in a multiscale fashion under the assumed Poisson model. They assumed a multinomial distribution for the data at the finer level conditioning on the coarser level. This approach is limited to the assumption of having fixed coarser level effect. In this work, we incorporate the scale effect using a convolution multiscale modeling approach. We have proposed three different models that encompass the scaling effect. We discuss each of these models in the following subsections:

3.1 Model 1

The data set considered for all the models is the Georgia oral cancer study. This data set is aggregated at the county and public health district levels. Let $y_i^c, i = 1, \ldots, N$, is the county level count of disease and $y_j^{ph} = \sum_{i \epsilon j} y_i^c, j = 1, \ldots, n$, is the j^{th} public health (PH) district level count of disease aggregated at the county level. In this model, we considered a joint convolution model at the county and public health district levels. The linkage between these two levels was incorporated in the model by including a shared spatial structured random effect, u_j^{ph}. The model is given by:

$$\begin{aligned} & y_i^c \sim \text{Poisson}(e_i^c \theta_i^c), \\ & \log(\theta_i^c) = \alpha_0^c + v_i^c + u_j^{ph}, \\ & y_j^{ph} \sim \text{Poisson}(e_j^{ph} \theta_j^{ph}), \\ & \log(\theta_j^{ph}) = \alpha_0^{ph} + v_j^{ph} + u_j^{ph}. \end{aligned} \tag{1}$$

Here, e_i^c and e_j^{ph} are the expected rate at the county and PH level, respectively. For this model and for the other subsequent models below, we have assumed a flat prior for the intercept parameters, α_0^c and α_0^{ph}. Further, the uncorrelated heterogeneity (UH) random effects, v_j^{ph} and v_i^c, were assumed to be normally distributed, i.e., $v_j^{ph} \sim N(0, sd_{vph}^2)$ and $v_i^c \sim N(0, sd_v^2)$, whereas the correlated heterogeneity random effect, u_j^{ph}, was assumed to have a conditional autoregressive (CAR) distribution. For the hyperparameter, sd_v, sd_{vph}, and sd_{uph}, we considered a uniform prior distribution, $U(0, 100)$ [2].

3.2 Model 2

Model 2 is similar to Model 1 except now the spatial structured random effect at the county level, u_i^c, is added to the model. The model can be written as:

$$\begin{aligned} & y_i^c \sim \text{Poisson}(e_i^c \theta_i^c), \\ & \log(\theta_i^c) = \alpha_0^c + v_i^c + u_i^c + u_j^{ph}, \\ & y_j^{ph} \sim \text{Poisson}(e_j^{ph} \theta_j^{ph}), \\ & \log(\theta_j^{ph}) = \alpha_0^{ph} + v_j^{ph} + u_j^{ph}. \end{aligned} \tag{2}$$

3.3 Model 3

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

$$\begin{aligned} & y_i^c \sim \text{Poisson}(e_i^c \theta_i^c), \\ & \log(\theta_i^c) = \alpha_0^c + v_i^c + u_i^c, \\ & y_j^{ph} \sim \text{Poisson}(e_j^{ph} \theta_j^{ph}), \\ & \log(\theta_j^{ph}) = \alpha_0^{ph} + v_j^{ph} + u_j^{ph}. \end{aligned} \tag{3}$$

3.4 Model Assessment and Goodness of Fit

To investigate how the models fit the data well, a deviance information criterion (DIC) was applied. We have also considered other criteria for model selection such as WAIC (Watanabe-Akaike or widely applicable information criterion [5]). For a predictive accuracy assessment, mean absolute prediction error (MAPE) and mean square prediction error (MSPE) were used.

3.5 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. Gamma distributions will be considered as a simple case but these will be extended later to spatially structured simulations for more realistic scenarios. In practice, we may have data at census tract, county, and public health district levels. Hence, we simulated data from a three level hypothetical grid which is divided into 2^4x2^4=256 smaller areas at the finest (lower) level, 2^3x2^3=64 areas (pixels) at the second (medium) level, and 2^2x2^2=16 areas at the coarser (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 second level. Similarly, the 16 samples were obtained by aggregating the observations at the second level nested within the coarser level. Mathematically,

this can be expressed as follows:

$$\begin{aligned} y_i^f &\sim \text{Poisson}(e_i^f \theta_i^f), \\ y_j^s &= \textstyle\sum_{i \epsilon j} y_i^f, \\ y_k^h &= \textstyle\sum_{j \epsilon k} y_j^s, \end{aligned} \tag{4}$$

where e_i^f is the expected rate, θ_i^f denotes the relative risk at the finest level, $y_i^f, i = 1, \ldots, 256$, $y_j^s, j = 1, \ldots, 64$, and $y_k^h, k = 1, \ldots, 16$, are the samples generated at the finest, second, and higher level, respectively. On the other hand, we sampled data that mimic the Georgia oral cancer study with two levels.

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 risk were calculated.

To evaluate the predictive ability of the models, MSPE and MAPE were computed at each level and averaged over the 200 data sets. Besides, the DIC was calculated at each level to compare model performance. Finally, the computation time was extracted to compare the execution time for the models.

4. RESULTS

4.1 Simulation Results

The results obtained from the data generated under a hypothetical grid scenario are shown in Table 1 and Figure 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 Model 3, 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 Model 3. However, the three models produce similar MAPE and MSPE at all levels except there is a slight difference at the medium level. Model 1 converges faster than the other models.

Table 2 displays the results of the fitted model to the simulated data that mimic the Georgia oral cancer study. We have also found here that Models 1 and 2 provide smaller DIC value than Model 3. The bias, MSE, MAPE, and MSPE are similar for all the three models. However, the MSPE at the public health district for Models 1 and 2 are smaller than the MSPE of Model 3.

4.2 Application to Data

To assess the benefit of including a shared random effect to handle the scale aggregation, we have applied the models discussed above to the Georgia oral cancer study. We can clearly see that, there is again in terms of model fit (DIC and WAIC), especially at the PH level (Table 3). Moreover, the parsimony model (Model 1) fits the data slightly better than Model 2. Hence, including the structural random effect at the county level (Model 2) does not improve the model fit for this example 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 quite different among the models. If we do not take into account for model complexity, Model 3 provides better model fit (Deviance) at the PH level compared to the shared random effect models (Models 1 and 2).

5. CONCLUSION

This paper was aimed at handling properly the scaling effect, which is common in disease mapping, using a multiscale modeling framework. We have shown that the scaling effect can be accommodated using a shared random effect. This model provides not only a better fit to the data, but also it produces a better predictive accuracy compared to the other models, especially at the coarser level. This is an expected result because the effect of the coarser level is inherited into the finer level through this shared random effect. Furthermore, we obtained an unbiased estimate of the relative risk. It also converges faster than the other models. In this paper, we introduced the shared random effect through the spatially structured random effect. Currently, we are investigating using a 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.

6. ACKNOWLEDGMENTS

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

7. REFERENCES

[1] J. Besag, J. York, and A. Mollié. Bayesian image restoration with applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43(1):1–59, June 1990.

[2] A. Gelman. Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, 3(1):515–533, 2006.

[3] E. Kolaczyk and H. Haung. Multiscale statistical models for hierarchical spatial aggregation. *Geographical Analysis*, 33(2):95–118, April 2001.

[4] M. M. Louie and E. Kolaczyk. A multiscale method for disease mapping in spatial epidemiology. *Statistics in Medicine*, 25:1287–1306, October 2006.

[5] S. Watanabe. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:3571–3594, October 2010.

Table 1: Simulation results for data generated from a hypothetical grid. Lower, medium, and higher represent the three levels and PD is the number of effective parameters. θ and CT denotes the relative risk and the average computation time in seconds over the 200 data sets, respectively. The expected rates of the 256 areas were assumed to be equal to one.

θ	Models	PD			DIC			MAPE			MSPE			θ		CT
		lower	medium	higher	lower	medium	higher	lower	medium	higher	lower	medium	higher	bias	MSE	elapsed
gamma(1,1)	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

Table 2: Simulation results for data generated similar in structure to Georgia oral cancer data. There are 159 counties and 18 public health (PH) districts. The expected rates obtained from Georgia oral cancer data were used to generate data from Poisson distribution with mean $e_i^c\theta_i^c$ and the relative risks θ_i^c were obtained from Model 3 fitted to the Georgia oral cancer data.

Models	PD		DIC		MAPE		MSPE		θ	
	county	PH district	county	PH district	county	PH district	county	PH district	bias	MSE
Model 1	18.41	8.36	466.39	104.41	1.34	4.56	4.69	36.25	-0.02	0.20
Model 2	15.88	8.08	466.21	103.55	1.35	4.51	4.73	35.44	-0.02	0.19
Model 3	26.12	10.53	470.43	110.99	1.34	4.89	4.79	41.78	-0.03	0.24

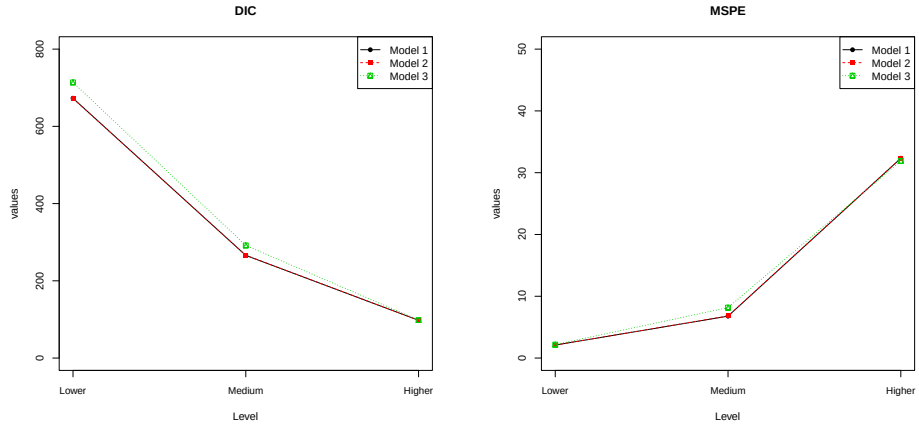

Figure 2: ***DIC and MSPE for the data generated from a hypothetical grid.***

Table 3: Model fit and predictive accuracy results for Georgia oral cancer data.

Models	PD		DIC		PDWAIC		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