Spatial and Spatio-temporal Epidemiology 22 (2017) 39–49

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Original Research

Comparing multilevel and multiscale convolution models for small area aggregated health data

Mehreteab Aregay[a,*], Andrew B Lawson[a], Christel Faes[b], Russell S Kirby[c], Rachel Carroll[a], Kevin Watjou[b]

[a] Department of Public Health, Medical University of South Carolina, Charleston, SC, USA
[b] Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium
[c] Department of Community and Family Health, University of South Florida, Tampa, FL, USA

ARTICLE INFO

Article history:
Received 15 November 2016
Revised 6 June 2017
Accepted 6 June 2017
Available online 5 July 2017

Keywords:
Multilevel model
Multiscale model
Convolution model
Shared random effects
Scaling effects
Contextual effects

ABSTRACT

In spatial epidemiology, data are often arrayed hierarchically. The classification of individuals into smaller units, which in turn are grouped into larger units, can induce contextual effects. On the other hand, a scaling effect can occur due to the aggregation of data from smaller units into larger units. In this paper, we propose a shared multilevel model to address the contextual effects. In addition, we consider a shared multiscale model to adjust for both scale and contextual effects simultaneously. We also study convolution and independent multiscale models, which are special cases of shared multilevel and shared multiscale models, respectively. We compare the performance of the models by applying them to real and simulated data sets. We found that the shared multiscale model was the best model across a range of simulated and real scenarios as measured by the deviance information criterion (DIC) and the Watanabe Akaike information criterion (WAIC).

1. Introduction

In spatial epidemiology, data are often arrayed hierarchically, i.e., individual level data are aggregated into areal units (e.g. counties) that are clustered to form larger areal units (e.g. states). The clustering of individuals into areal units, which in turn are grouped into larger areal units, can induce contextual effects (Lawson, 2013, 2016); meaning that individuals within areal units have similar characteristics. In general, contextual effects arise from the underlying spatial distribution of the individual level outcomes. Appropriate model parameters should be used to adjust for the contextual effects and when such parameters are not included then bias would be induced in the estimated relative risk, which is the parameter of interest in the disease mapping. Hence, researchers considered multilevel modeling of hierarchically available individual level data to encompass contextual effects (Bobashev and Anthony, 1998; Goldstein et al., 2002; Leyland and Goldstein, 2001; Merlo et al., 2004; Preisser et al., 2003). However, this approach only handles the correlation between the outcomes within a single areal unit; it ignores spatial correlation among neighboring areal units.

Multilevel models often assume that all spatial correlation can be reduced to within area correlation (Chaix et al., 2005); thus, there is no spatial random effect component that handles the correlation between the neighboring areas. The random effects in multilevel models only account for the correlation between the individual level outcomes within a given spatial unit. Therefore, it provides partial information on the geographical variation of health outcomes in measuring the correlation within a spatial unit but not the correlation between neighboring regions. Researchers extend multilevel models to incorporate

* Corresponding author.
E-mail address: mihrish00m@gmail.com (M. Aregay).

http://dx.doi.org/10.1016/j.sste.2017.06.001
1877-5845/

spatial interaction effects in different fields such as geography and spatial econometrics (Browne et al., 2001; Chaix et al., 2005; Dong et al., 2015; Langford et al., 1999; Tranmer et al., 2014). In this paper, we (i) extend the multilevel models to account for spatial correlations between adjacent areas using a convolution model developed by Besag et al., (1991); and (ii) develop models for aggregated unit level data by adjusting for contextual effects. The proposed multilevel model focuses on the risk variation at the fine level areal units by incorporating the contextual effects in the model. In addition, the estimation of the risk variation at larger areal units (coarse level) is possible by aggregating (e.g. averaging) the fine level estimates within the coarse level, while using only the data at the fine level.

In practice, data could be available at different geographically aligned levels. For example, the outcome of interest could be available in the form of aggregation at the census block, block group, and census tract: the responses at the census block could be summed up to obtain the responses at the block group, which in turn could be summed up to obtain the responses at the census tract. This kind of data aggregation results in losing information at the coarse level (e.g. block group and census tract). This is known as a scaling effect in geography (Wong, 2009). Scaling effects arise when data are aggregated from a lower (e.g. census tract) into a higher geographical level (e.g. county). In the literature, multiscale models have been used in different fields to solve scaling problems at multiple scale levels (Basseville et al., 1992; Berliner et al., 1999; Calder et al., 2009; Chou et al., 1994; Craigmile and Guttorp, 2011; Delouille et al., 2006; Huang and Cressie, 2000; Huang et al., 2002; Johannesson et al., 2007; Kolaczyk and Huang, 2001; Nychka et al., 2002; Vidakovic, 1999; Wikle et al., 2001; Zhu and Yue, 2004).

In spatial epidemiology, researchers have implemented multiscale models to account for scaling effect due to the aggregation of data (Banerjee et al., 2004; Cressie, 1996; Wong, 2009) by factorizing the likelihood at the coarse (high) level into the fine (low) level (Louie and Kolaczyk, 2004, 2006a, 2006b). Alternatively, we (Aregay et al., 2015a, 2015b, 2016a, 2016b) developed a shared random effect multiscale model that accommodates the aggregation (scale) effect by inheriting the coarse level effect into the fine level. However, it could be argued that the latter approach uses the data twice as the data at the coarse level are an aggregation of the data at the fine level. The objective of this paper is to describe risk variations at fine and coarse levels simultaneously by accommodating scaling and contextual effects. To achieve this goal, we applied and compared different models. First, we compare the shared multiscale model with the shared multilevel model in real and simulated data sets. Second, we study the impact of ignoring the contextual effects on the estimation of the risk variations at both the fine and coarse levels by simulating data with strong contextual effects. Note that the focus of this paper is on studying contextual effects although we touch on scaling effects as well.

The structure of the paper is as follows. In Section 2, we present the data that motivated us to conduct this research. Section 3 describes the statistical methods as well as the design of the simulation study, while Section 4 dedicates to the results obtained from fitting the models to the real and simulated data sets. Finally, in Section 5, we present the discussion and concluding remarks.

2. Georgia oral cancer data

We are motivated by the county level data available in the state of Georgia via OASIS system (http://oasis.state.da.us). We consider the number of persons discharged from non-federal acute-care inpatient facilities for oral cancer in 2008. The observed outcomes of the counties are aggregated (summed up) to the public health (PH) districts. These aggregations of data can induce a scaling effect. The state of Georgia consists of 159 counties (see left panel in Fig. 1) that are classified into 18 PH districts (see the right panel of Fig. 1). The grouping of the counties into PH districts can induce a contextual effect. Each PH district consists of one or more counties. The PH districts are used for administration of public health resources. The Georgia Department of Public Health (DPH) funds and collaborates with the 18 PH districts. The goal of modeling the risk variation at both the county and PH district levels is that it can be used for allocating of health resources at both levels in a cost-effective manner. Hence, the DPH can use the risk mapping results to legislate regulations to protect the public health in each county as well as in each PH district.

The observed standardized morbidity ratio (SMR), which is the ratio of the outcome to the expected number of cases, at both the county and PH levels are displayed in Fig. 2. We can see that the scaling effect smooths out the county level risk variation when the data are aggregated into the PH districts. To address both the contextual and scale effects, we propose different models described in Section 3. It is worth mentioning contextual and scale effects have an inverse relationship. When we have strong presence of contextual effects, we will have weak scale effects because the risk variations will be similar at the fine (e.g. county) and coarse (e.g. PH districts) levels (see Fig. 3). The scales of the relative risk (RR) ranges from 0.61 to 1.68 on both scale levels. On the other hand, when there are weak contextual effects, the scale effects will be strong (see Fig. 4). The scale of the RR in the left panel is between 0.45 and 2.23, whereas in the right panel it is between 0.97 and 1.49 indicating that the presence of a relatively strong scale effect smoothed out the risk variation at the county level when it is aggregated into the PH level. We can also see this kind of behavior in Fig. 2. The application of the models to the Georgia oral cancer data is deferred to Section 4.2.

3. Models for aggregated small area data

In the next section, we present the models most relevant to small area aggregated data. To make it clear, we abbreviated the four models considered below as M1, M2, M3, and M4 and they represent Model 1, Model 2, Model 3, and Model 4, respectively. We define the models using the two scale levels Georgia oral cancer study. Assume that y_{ij} is the outcome of interest for the $j\text{th}^{th}$ county (fine scale) at the $i\text{th}^{th}$ public health district (coarse scale)

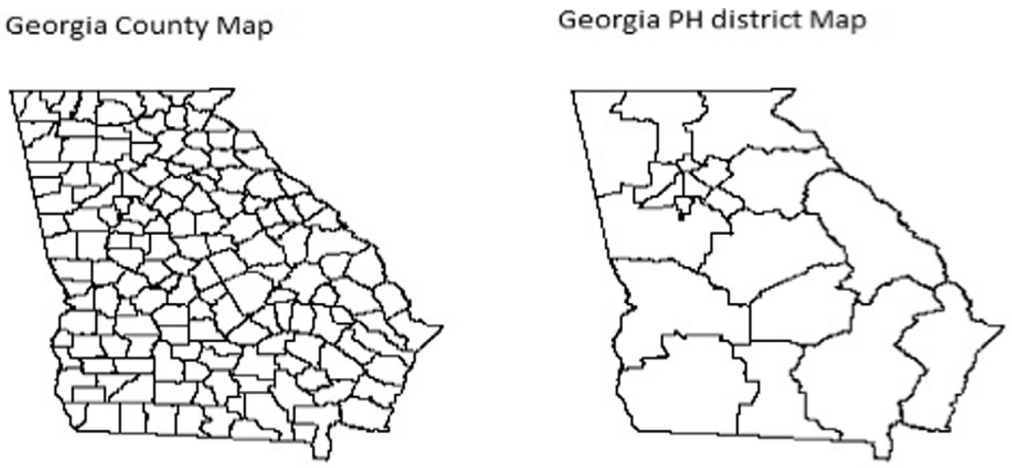

Fig. 1. State of Georgia, USA: county and PH district boundary map.

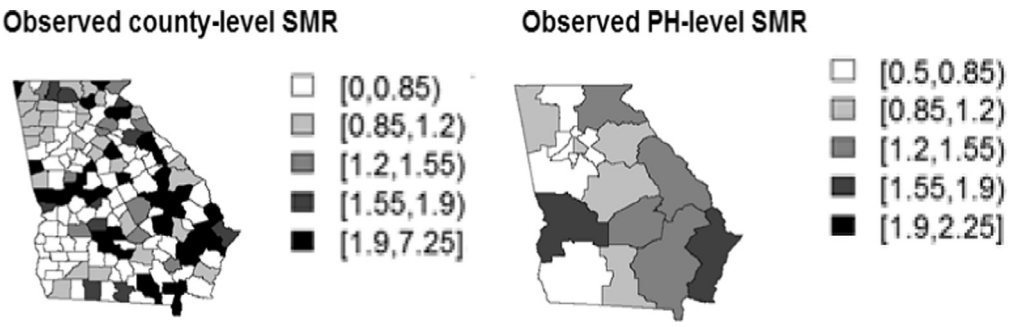

Fig. 2. Georgia oral cancer study: observed standardized morbidity ratio (SMR) pattern.

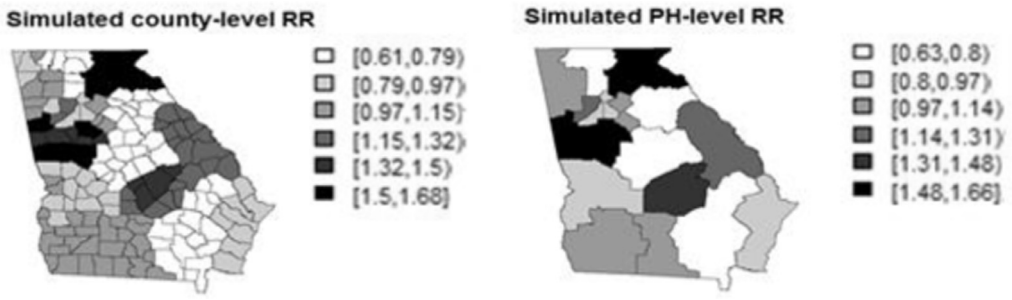

Fig. 3. Simulated relative risk (RR) at both the county and public health district levels by allowing a strong contextual effect.

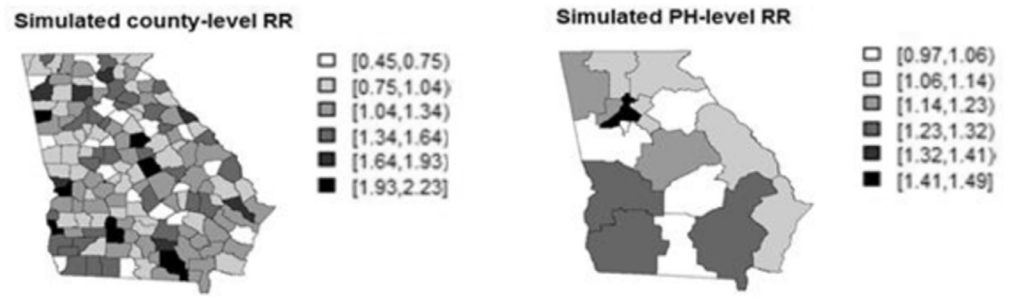

Fig. 4. Simulated relative risk (RR) at both the county and public health district levels by allowing a weak contextual effect.

where $i = 1, 2, \ldots \cdots, N$ and $j = 1, 2, \ldots \cdots, M_i$ with N representing the number of PH district, i.e., $N = 18$ and M_i denotes the number of counties within the $i$thth PH district. In the next sections, we describe the four models.

3.1. Model 1: convolution model

Besag et al. (1991) proposed a convolution model that uses correlated heterogeneity terms (CH) and uncorrelated heterogeneity terms (UH) to describe the spatial risk variation for disease(s). The CH terms measure the similarities of the risks of a certain infection between neighboring regions, whereas the UH terms uniquely quantify the risk of a certain infection for each region. In this paper, we considered the convolution model (M1) to estimate the relative risk at the county level as follows:

$$y_{ij}\big|\mu_{ij} \sim Pois(\mu_{ij} = e_{ij}\theta_{ij}), \tag{1}$$

$$\log(\theta_{ij}) = a_{01} + v_{ij} + u_{ij}, \tag{2}$$

where θ_{ij}, a_{01}, v_{ij}, u_{ij}, and e_{ij} are the relative risk, the intercept, the UH terms, the CH terms, and the expected number of cases for county$j$ within the PH district i, respectively. We computed e_{ij} as follows:

$$e_{ij} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{M_i} p_{ij}} p_{ij}$$

where p_{ij} is the county level population size in county$j$ within the PH district i. We assumed a normal distribution for the UH terms,$v_{ij} \sim N(0, {\sigma^2}_{v1})$, and an intrinsic conditional autoregressive structure (ICAR) for the CH terms, i.e.,

$$u_{ij}\left|u_{-ij} \sim N\left(\frac{1}{n_{ij}}\sum_{j\sim l} u_{il}, \frac{{\sigma^2}_{u1}}{n_{ij}}\right)\right.,$$

where $j \sim l$ indicates the two counties j and l are neighbors, $n_{ij}$is the total number of neighbors for county j of the ith PH district not just neighbors within the same PH district, u_{-ij} is the set of all county level random effects excluding the jth county within PH district i and ${\sigma^2}_{u1}$ is the variance of the CH terms. Note that we can have counties be neighbors even if they have different PH districts. Neighbors are those regions that share a common boundary and in this paper, they are defined based on the adjacency-matrix. For this and the other models below, we assumed a non-informative normal prior for the intercept $a_{01}$and uniform prior distributions, U (0100), for the standard deviations σ_{u1} and σ_{v1} (Gelman, 2006).

To estimate the relative risk at the PH district level, we aggregated the relative risk estimates from the county level, i.e., $\theta_i = \mu_i/e_i$, where $\mu_i = \sum_{j=1}^{M_i}\mu_{ij}$ and e_i is the expected number of cases at the ith PH level. Model 1 accommodates neither contextual effects nor scale effects. The next model, Model 2, will demonstrate how to adjust for the contextual effects.

3.2. Model 2: multilevel convolution model

Although the convolution model (M1) is a widely used method to estimate spatial risk variations, it is not flexible enough to address the contextual and scaling effects. The contextual effects could be accommodated by including the coarse level CH and UH terms into the fine level model in (2). The model formulation is similar as in (1) and (2), except we now include the random effects v_i' and u_i' that are shared across counties within the PH district as following:

$$\log(\theta_{ij}) = a_{0_1} + v_{ij} + u_{ij} + v_i' + u_i', \tag{3}$$

where v_i' is the PH level UH term and assumed to be normally distributed, $v_i' \sim N(0, {\sigma^2}_{v1})$, while u_i' is the PH level CH term and it has an ICAR distribution, $u_i'|u_{-i}' \sim N(\frac{1}{n_i}\sum_{i\sim l} u_l', \frac{{\sigma^2}_{u1}}{n_i})$. Here $i \sim l$ denotes the two PH districts i and l are neighbors, n_i indicates the numbers of neighbors for PH district i, and u_{-i}' is the set of all PH level random effects not including the ith. We assumed similar prior distributions as in M1 for the parameters. Here also, the relative risk at the PH level,θ_i, is estimated in a similar fashion as described in M1.

3.3. Model 3: independent multiscale model

The previous models, M1 and M2, only use the data at the county level. In our example, we have data at the county and PH levels for the state of Georgia. Hence, in this model, we use the information at each scale level; we assumed Poisson distributions for the outcomes at the county and PH levels and considered a convolution model at each scale level as follows:

$$\begin{aligned} &y_{ij}\big|\mu_{ij} \sim Pois(\mu_{ij} = e_{ij}\theta_{ij}),\\ &\log(\theta_{ij}) = a_{01} + v_{ij} + u_{ij},\\ &y_i|\mu_i \sim Pois(\mu_i = e_i\theta_i),\\ &\log(\theta_i) = a_{02} + v_i' + u_i', \end{aligned} \tag{4}$$

where y_i is the sum of the outcomes of the counties within the PH level, i.e., $y_i = \sum_{j=1}^{M_i} y_{ij}$with $y_{ij}, \theta_{ij}, a_{01}, v_{ij}, u_{ij}, v_i', u_i'$, and e_{ij} are the same as in M1 and M2. Here, θ_i, a_{02}, and e_i are the relative risk, the intercept, and the expected number of cases at the PH district level. We assumed a non-informative normal prior distribution for a_{02} and we calculated e_i as follows: $e_i = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} p_i} p_i$, where p_i is the PH level population size and N is the number of PH districts, i.e., $N = 18$. Note that Model 3 addresses neither the scale effect nor the contextual effects.

3.4. Model 4: shared multiscale model

We have seen that M3 uses the information at each scale. However, M3 does not adjust for scale and contextual effects. In this section, we present the shared multiscale model that accounts for the scale effects as well as the contextual effects using the shared CH and UH terms similar to M2 in (3). The model formulation for M4 is similar to M3, but now the shared convolution model in (3)

replaces the county level convolution model in (4) as follows:

$$
\begin{aligned}
&y_{ij}\big|\mu_{ij} \sim Pois(\mu_{ij} = e_{ij}\theta_{ij}),\\
&\log(\theta_{ij}) = a_{01} + v_{ij} + u_{ij} + v'_i + u'_i,\\
&y_i|\mu_i \sim Pois(\mu_i = e_i\theta_i),\\
&\log(\theta_i) = a_{02} + v'_i + u'_i.
\end{aligned} \tag{5}
$$

The difference between M4 and M2 is that in M4 we estimate the relative risk at each scale level using convolution models, whereas in M2, we considered only a convolution model at the fine level (county) and the coarse level (PH district) relative risks are estimated using the fine level relative risks. In addition, M2 uses only the data at the fine level, while M4 considers the aggregated data across the scale levels. Hence, the shared components in M4 could address both the contextual effect as well as the scaling effect by propagating information across the scale levels. Note that M2 only allows one to study the covariate effect at the fine level. On the other hand, M4 helps to investigate the effect of covariate on the outcome of interest across scale levels. We demonstrated this in the Supplementary Appendix. In the next section, we describe first the criteria to compare the different models and thereafter, we present a simulation study to investigate the impact of ignoring the contextual and scaling effects.

3.5. Goodness of fit

To evaluate and compare the performance of the models, we considered the deviance information criterion (DIC Spiegelhalter et al., 2002) and the Watanabe Akaike information criterion (WAIC Watanabe, 2010). While the DIC is based on a point estimate (Gelman et al., 2014; Plummer, 2008; van der Linde, 2005), the WAIC approximates cross-validation and uses a posterior predictive distribution (Gelman et al., 2014). Note that Gelman et al., (2014) discussed that it is not easy to do WAIC in some structured-data settings such as time series, spatial and network data because it depends on partitioning of data. They also discussed that DIC does not make this partition explicitly, but derivation of DIC assumes that the residuals are independent given the point estimate. However, the formulation of WAIC is based on pointwise predictive density (Gelman et al., 2014; Watanabe, 2010) and it is straightforward to implement it in standard software such as WinBUGS even for structured-data such as spatial data (Aregay et al., 2015a, 2015b, 2016a ,2016b). Moreover, as expected, we found similar results for both WAIC and DIC (see Section 4). To assess the prediction ability of the models, we employed a mean square prediction error (MSPE Lawson, 2013).

One can argue that the multilevel models (M1 and M2) should not be compared to the multiscale models (M3 and M4) using DIC and WAIC measures because they have different likelihoods. Nevertheless, we believe that the multilevel models and the multiscale models can be compared using the DIC and WAIC measures at the county level because the likelihood of the multilevel models is similar with the likelihood of the multiscale models at the county level. However, at the PH level, we can only compare M3 and M4 using DIC and WAIC because M1 and M2 only use the data (likelihood) at the county level and hence, we cannot obtain DIC and WAIC values at the PH level for M1 and M2.

We fitted all models jointly using Markov chain Monte Carlo (MCMC) posterior sampling in a single analysis as described in Section 4. We considered 30,000 iterations after we discarded the first 15,000 burn-in samples. We ran three separate chains starting from different initial values. Hence, the posterior means were calculated based on 45,000 iterations, which were sufficient for convergence. Further, we have run the MCMC algorithm for 100,000 burn-in iterations followed by 15,000 iterations and the results provided robust estimates. We assessed convergence using an estimated potential scale reduction factor ($\hat{R}$) and trace plots.

3.6. Simulation study

In this section, we aim to investigate the impact of ignoring contextual and scale effects using a simulation study under a range of scenarios. We simulated county level data within the state of Georgia from a Poisson distribution in (1) using the relative risk θ_{ij} in (3) and the expected number of cases e_{ij} which is generated from a gamma distribution with hyper-parameters equal to one. This means that the county-level data sets were simulated from M2. We fitted M1 and M3 to the simulated data sets to assess the impact of ignoring scaling and contextual effects, whereas we fitted M2 and M4 to investigate how well the models recover the simulated risks. To obtain the outcomes at the PH level, we summed up the county level simulated outcomes within the PH district, i.e., $y_i = \sum_{j=1}^{M_i} y_{ij}$. Note that we did not simulate both the county and PH level data from M3 and M4 because the county level data do not sum up to the PH level data when we simulate the PH level data from a Poisson distribution. The county level data within the PH district must sum up to obtain consistent data at the PH district level because we have hierarchically aligned counties within the PH districts. Hence, we could not address the scaling effect in our simulated model. Although we could not directly simulate the scaling effect from the models, it is naturally induced during the data aggregation from the county into the PH level. An alternative approach would be to use the multinomial simulation based on the PH district simulated count with county count probabilities a function of the county level risks. However, this approach would involve making conditions on the PH level data to simulate the county level data from a multinomial distribution, which may be unrealistic because in practice, the county level data are observed first and then aggregate to obtain PH level data (see Section 2).

Next, we will discuss the two scenarios, where we simulated the contextual effects. In the first scenario, we simulated strong contextual effects, while in the second scenario we simulated weak contextual effects. The novelty of this simulation study is that it investigates how M1 and M3 that ignore the contextual effects might badly recover the simulated risks during the presence of strong contextual effects in the simulated data sets. On the other hand,

we expect that M2 and M4 will recover well the simulated risks.

3.6.1. Scenario 1: strong contextual effects

In the first scenario, we simulated 200 data sets from a Poisson distribution in (1) by introducing a strong contextual effect in (3), i.e., we simulated 200 data sets from M2. Here, our main interest is to assess whether the models that ignore the contextual effect (M1 and M3) could recover the risk variation when there is a strong PH (contextual) effect in the data. This can be done by assuming the variances of the PH level CH and UH, ${\sigma^2}_{u2}$ and ${\sigma^2}_{v2}$, to be large relative to the variances of the county level CH and UH, ${\sigma^2}_{u1}$ and ${\sigma^2}_{v1}$, as shown in Fig. 3. We assumed the following values that reflect the practical values for the relative risk: $\sigma_{u1} = 0.01$, $\sigma_{v1} = 0.01$,$\sigma_{u2} = 0.3$,$\sigma_{v2} = 0.3$, and $a_{01} = 0.1$. Comparing the county level and the PH level risk patterns in Fig. 3, we can see that the counties within the PH district have similar risks; there is a strong grouping (contextual) effect. On the other hand, there is a weak scale effect in the simulated risks because the scale of the risks at both county and PH levels is similar; it ranges approximately from 0.6 to 1.68. We also considered other possible values with $\sigma_{u1} = 0.1$ $\sigma_{v1} = 0.1$,$\sigma_{u2} = 0.5$,$\sigma_{v2} = 0.5$, and $a_{02} = 0.1$ (see Section 1 in the Supplementary Appendix).

3.6.2. Scenario 2: weak contextual effects

In this scenario**,** our aim is to evaluate whether the complex models (M2 and M4) and the parsimonious models (M1 and M3) provide similar risk variations when there is not a strong contextual effect in the data. Thus, we simulated 200 data sets from a Poisson distribution in (1) by allowing a weak contextual effect in (3), i.e., we simulated data from M2. Here, we assumed the counties would not inherit common characteristics from their PH district; each of the counties has their own characteristics. To allow for such assumption in the model, we considered large values for the variances of the county level CH and UH relative to the variances of the PH level CH and UH terms. Hence, we assumed the following values: $\sigma_{u1} = 0.3$, $\sigma_{v1} = 0.3$,$\sigma_{u2} = 0.01$,$\sigma_{v2} = 0.01$, and $a_{01} = 0.1$. We also considered other possible values with $\sigma_{u1} = 0.5$, $\sigma_{v1} = 0.5$,$\sigma_{u2} = 0.1$,$\sigma_{v2} = 0.1$, and $a_{01} = 0.1$ (see Section 2 in the Supplementary Appendix). Fig. 4 shows the trend of the relative risk at both the county and PH levels for data simulated with a weak contextual effect. We can see that the county level risks within the PH district are different; there is no contextual (grouping) effect. Note that there is a relatively strong scale effect in Fig. 4 as compared to Fig. 3 because the scales at the county and PH levels are different; While the scale of the county level risks ranges from 0.45 to 2.23, the PH level risks were smoothed out and hence, the scale is between 0.97 and 1.49.

4. Results

We implemented the models with real and simulated data via the R2WinBUGS package. First, we present the results obtained by fitting the models to the 200 simulated data sets. Thereafter, we describe the results obtained by applying the models to the oral cancer data from the state of Georgia.

4.1. Simulation results

4.1.1. Scenario 1: strong contextual effects

Table 1 shows the model fit and prediction accuracy results for the data simulated using scenario 1, indicating that M4 is the best model at both the county and PH levels as measured by DIC and WAIC. Note that the bold-faced numbers represent the best model. The next best model at the county level is M2. M1 and M3 perform similarly at the county level. Although there is no significant difference in terms of the MSPE, M2 and M4 provide slightly smaller values of the MSPE as compared to M1 and M3 at the county level. On the other hand, at the PH level, M4 provides significantly smaller MSPE than M3, showing that M4 has better prediction accuracy than M3. The fact that M4 is better than M2 at the county level might indicate that jointly modeling the data at both levels is important to propagate information from one level to another level in both directions.

Table 2 displays the bias and MSE of the parameters obtained from the models fitted to the simulated data. The bold-faced numbers represent the best model for that particular parameter estimate in terms of bias and MSE. M1 and M3 have smaller bias and MSE estimates for the intercept at the county level a_{01} as compared to M2 and M4. This is an expected result because in M2 and M4 both the county and the shared PH level variances of the CH and UH terms affect the estimate of the county level intercept, while in M1 and M3 only the county level variances of the CH and UH terms affect the estimate of the county level intercept. On the other hand, M4 produces the least bias and MSE estimates for the county level variances of the CH and UH terms. This may be because the shared PH level CH and UH terms could capture some proportion of the county level variabilities; the PH level variances of the CH and UH terms obtained from M4 are larger than that of M3. Ultimately, M2 has smaller bias and MSE estimates for the PH level variances of the CH and UH terms as compared to M4. This could be because we simulated the data from M2. In terms of average bias and MSE of the relative risks over the 159 counties and 200 simulated data sets, M1 and M2 provide slightly smaller bias as compared to M3 and M4, while M4 produces the smallest MSE. Note that the bias measures of the relative risk estimates are smaller because they are an average of the bias of the 159 counties; some counties have negative values others have positive bias values. When the biases are averaged over the 159 counties, they canceled each other.

To investigate the models in terms of recovering the risk pattern, we calculated the average relative risk over the 200 simulated data sets for each county as shown in Fig. 5. We can see that M2 and M4 recover the simulated risk in Fig. 3, whereas M1 and M3, which ignore the contextual effects, do not recover the simulated risk well. Moreover, M2 and M4 produce consistent risk estimates at both the county and PH levels, while M1 and M3 provide risk estimates that are inconsistent across levels.

Table 1
Simulation study scenario 1: model fit and prediction accuracy for data simulated from a Poisson distribution.

Models	DIC		WAIC		MSPE	
	County	PH district	County	PH district	County	PH district
Model 1	362.90	–	362.31	–	1.969	–
Model 2	359.44	–	358.09	–	1.920	–
Model 3	362.54	88.50	362.03	87.31	1.967	19.951
Model 4	**350.09**	**79.57**	**348.28**	**75.63**	**1.877**	**14.355**

Table 2
Simulation study scenario 1: bias and MSE of the estimates obtained from Models 1–4 that are applied to the simulated data from a Poisson distribution.

Parameters	Model 1		Model 2		Model 4		Model 3	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
a_{01}	**−0.144**	**0.029**	−0.159	0.035	**−0.144**	**0.029**	−0.169	0.043
σ_{u1}	0.374	0.167	0.284	0.097	0.367	0.159	**0.202**	**0.048**
σ_{v1}	0.231	0.062	0.193	0.043	0.228	0.064	**0.159**	**0.029**
σ_{u2}	–	–	0.102	**0.022**	**0.099**	0.024	0.162	0.048
σ_{v2}	–	–	-0.017	**0.012**	**0.001**	0.014	0.081	0.029
θ_1	**−0.004**	0.073	**0.004**	0.042	−0.006	0.073	0.007	**0.040**

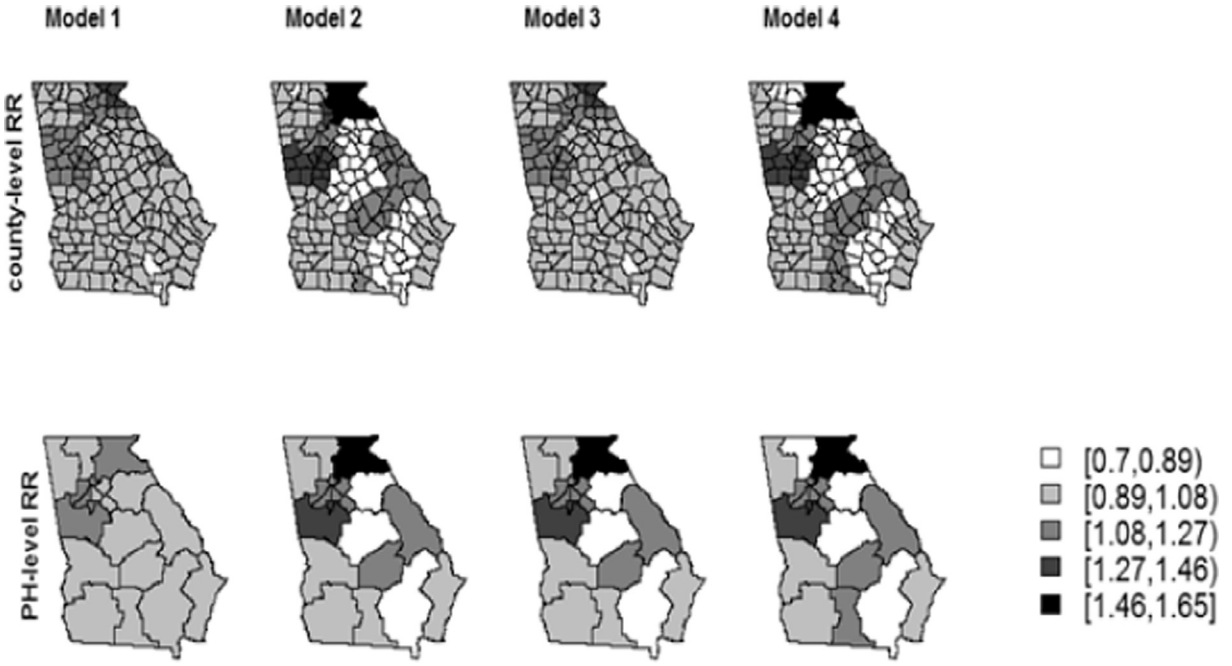

Fig. 5. Simulation study scenario 1: relative risk (RR) pattern obtained from the models fitted to the simulated data at the county and public health (PH) levels. The relative risks were obtained by averaging over all the 200 simulated data sets.

To assess how robust our results are, we assumed other possible values for the parameters with $\sigma_{u1} = 0.1$ $\sigma_{v1} = 0.1, \sigma_{u2} = 0.5, \sigma_{v2} = 0.5,$ and $a_{01} = 0.1$; the simulated relative risks are shown in the Supplementary Appendix Fig. 1A. The results obtained from the models (see Tables 1A and 2A, and Fig. 2A in Section 1 in the Supplementary Appendix) are similar with the findings of the models fitted to the simulated data assuming $\sigma_{u1} = 0.01,$ $\sigma_{v1} = 0.01, \sigma_{u2} = 0.3, \sigma_{v2} = 0.3,$ and $a_{01} = 0.1$.

4.1.2. Scenario 2: weak contextual effects

In this scenario, we obtained the results from the models fitted to the simulated data assuming a weak contextual effect as shown in Tables 3 and 4. Here also, M4 is the best model as measured by DIC, WAIC, and MSPE. This is because M4 attempts to address the scale effect due to data aggregation from the county to the PH level. However, the difference in DIC, WAIC, and MSPE values between the models in scenario 2 is not as significant as in scenario 1.

Table 3
Simulation study scenario 1: model fit and prediction accuracy for data simulated from a Poisson distribution.

Models	DIC		WAIC		MSPE	
	county	PH district	County	PH district	County	PH district
Model 1	379.06	–	378.89	–	2.171	–
Model 2	380.04	–	379.88	–	2.168	–
Model 3	378.73	86.35	378.60	84.88	2.181	18.111
Model 4	**374.84**	**80.18**	**374.39**	**77.19**	**2.146**	**15.469**

Table 4
Simulation study scenario 2: bias and MSE of the estimates obtained from Models 1–4 that are applied to the simulated data from a Poisson distribution.

Parameters	Model 1		Model 2		Model 3		Model 4	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
a_{01}	**−0.052**	**0.011**	−0.080	0.017	−0.055	0.013	−0.084	0.019
σ_{u1}	−0.006	0.016	−0.013	0.020	**−0.0001**	0.021	−0.075	**0.015**
σ_{v1}	**−0.035**	**0.011**	−0.042	0.013	−0.039	0.013	−0.068	0.016
σ_{u2}	–	–	**0.270**	**0.079**	0.274	0.083	0.333	0.125
σ_{v2}	–	–	0.151	**0.028**	**0.149**	0.031	0.197	0.053
θ_1	**-0.007**	**0.132**	**−0.007**	0.149	−0.009	0.134	−0.008	0.143

In addition, M1, M2, and M3 have similar model performance at the county level. On the other hand, M4 is still better than M3 at the PH level. This shows that even with a weak PH (contextual) effect in the counties within the PH district, M4 fits and predicts the data better than M3, especially at the PH level. This may be because the shared components in M4 can serve dual purposes; adjusting for (1) scale effects and (2) contextual effects. Aregay et al. (Louie and Kolaczyk, 2004, 2006) also found model improvement for the shared multiscale model as compared to the independent multiscale model for data simulated without contextual effects. As in scenario 1, M4 is better than M2 at the county level. This may be because M4 allows the county level data to borrow information from the PH level data and vice versa.

Similar to scenario 1, M1 and M3 reveal smaller bias and MSE estimates for the intercept as compared to M2 and M4 (Table 4). However, M2 produces the smallest bias and MSE estimates for the variances of the CH and UH terms at the PH levels. In contrast to scenario 1, the bias and MSE estimates for the county level variances ($\sigma^2{}_{u1}$ and $\sigma^2{}_{v1}$) from M2 are smaller than those of M4. These results could show that the shared components in M4 may not fully incorporate some of the county level variabilities into the PH level data when there is a weak PH effect in the county level data.

Fig. 6 displays the county and PH levels relative risks obtained from the models fitted to the simulated data sets. Here, in contrast to scenario 1, all the models provide similar results. However, M2 and M4 recover the simulated risk in Fig. 4 slightly better than M1 and M3; For example, in south-western Georgia, there is an elevated risk present in the results from M2 and M4 as well as in the simulated model in Fig. 3 at both levels. M1 and M3, however, provide relatively lower risk estimates in those areas as compared to that of M2 and M4.

As in scenario 1, here also we assumed other possible values for the parameters with $\sigma_{u1} = 0.5$, $\sigma_{v1} = 0.5$,$\sigma_{u2} = 0.1$,$\sigma_{v2} = 0.1$, and $a_{01} = 0.1$, and the simulated relative risks are displayed in Fig. 3(A) (see Section 2 in the Supplementary Appendix), indicating that the findings (see Fig. 4(A), and Tables 3(A) and 4(A) Section 2 in the Supplementary Appendix) are similar with the results assuming $\sigma_{u1} = 0.3$, $\sigma_{v1} = 0.3$,$\sigma_{u2} = 0.01$,$\sigma_{v2} = 0.01$, and $a_{01} = 0.1$.

4.2. Application to Georgia oral cancer data

The results obtained from the models applied to the Georgia oral cancer study are shown in Tables 5 and 6 and Fig. 7. In terms of DIC and WAIC, M4 slightly outperforms the other models at the county level. In addition, M4 is better than M3 at the PH level as measured by DIC, WAIC, and MSPE. On the other hand, all the models provide similar predictive accuracy at the county level. The fact that M4 is better than M3 at the PH level could indicate that the shared components in M4 attempt to recover the lost information during data aggregation. Nevertheless, there is no improvement in M2. This may be because there is not a strong contextual effect in this example (see Fig. 2). Hence, for this example, the shared components in M2 intended to account for the contextual effect could be considered unnecessary.

The results in Table 6 show that M1, M2, and M3 provide similar county level posterior parameter estimates, while M4 reveals slightly different county level estimates from those obtained using M1–M3. For example, the county level variance of the CH term ($\sigma^2{}_{u1}$) obtained from M1–M3 is almost twice as large as that of M4. This may be because some of the county level variabilities in M4 propagate into the PH level variabilities as we found in the simulated data sets. This could be seen from PH level variances ($\sigma^2{}_{u2}$ and $\sigma^2{}_{v2}$) in M4, which are higher than those of M2 and M3.

When we compare the observed SMR in Fig. 2 with the relative risk (RR) obtained from the models in Fig. 7, there is much variability in Fig. 2 at the county level as

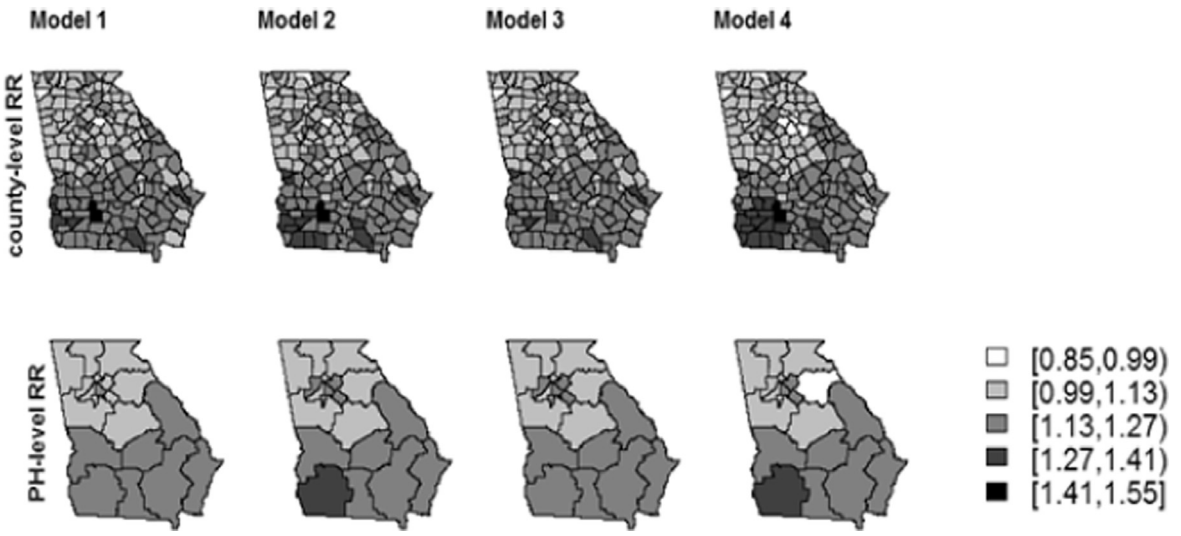

Fig. 6. Simulation study scenario 2: relative risk (RR) pattern obtained from the models fitted to the simulated data at the county and public health (PH) levels. The relative risks were obtained by averaging over all the 200 simulated data sets.

Table 5
Georgia oral cancer study: model fit and prediction accuracy results.

Models	DIC		WAIC		MSPE	
	County	PH district	County	PH district	County	PH district
Model 1	485.58	–	486.57	–	4.861	–
Model 2	487.31	–	488.17	–	4.86	–
Model 3	485.52	114.56	485.58	112.71	4.81	42.24
Model 4	**483.62**	**107.03**	**484.17**	**102.66**	**4.80**	**35.9**

Table 6
Georgia oral cancer study: the posterior mean and standard error (SE) estimates.

Parameters	Model 1		Model 2		Model 3		Model 4	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
a_{01}	0.031	0.073	0.014	0.086	0.025	0.077	−0.015	0.085
σ_{u1}	0.564	0.169	0.509	0.197	0.518	0.175	0.242	0.164
σ_{v1}	0.183	0.115	0.197	0.131	0.261	0.119	0.204	0.117
σ_{u2}	–	–	0.264	0.180	0.399	0.178	0.425	0.189
σ_{v2}	–	–	0.129	0.089	0.144	0.103	0.178	0.106

compared in Fig. 7. This is because the SMR is a crude estimate and as such, it does not adjust for neighbors' effect, whereas the Models 1–4 provide smoothed RR estimates at the county level. However, at the PH level, the observed SMR and the RR obtained from the models have similar patterns. This is because the aggregation effect smooths out the data at the PH level and we tend to see a smoothed SMR as compared to the county level: the maximum county level SMR is 7.11 but after geographical aggregations of these counties into PH district, the maximum PH level SMR is dropped into 1.78. Note that there are some differences in the RR estimates among the models. For instance, similar to the observed SMR, M4 results in higher risk estimates in the southwest part of the state of Georgia as compared to Models 1–3.

Furthermore, we extended Models 1–4 in Section 3 to include an income covariate effect on the incidence of oral cancer. The relative risks of the models with an income covariate are provided in the Supplementary Appendix, Table 5(A). Note that X_{ij} represents the median household income for the jth county at PH district i, whereas X_i denotes the average income of the counties within the $i$thth PH district, i.e., $X_i = \frac{\sum_{j=1}^{M_i} X_{ij}}{M_i}$. The observed income at the county and PH levels are shown in the Supplementary Appendix, Fig. 5A indicating that there is a high median household income at the northern Georgia.

The results of the models applied to assess the spatial impact of income on the incidence of oral cancer in the state of Georgia are provided in the Supplementary Appendix, Table 6A. Comparing these results with the findings in Table 5 above, we can see that including the income covariate in the model significantly improve neither the model fit nor the prediction accuracy. On the other hand, from Table 7A in the Supplementary Appendix, all the models indicate that income has a significant

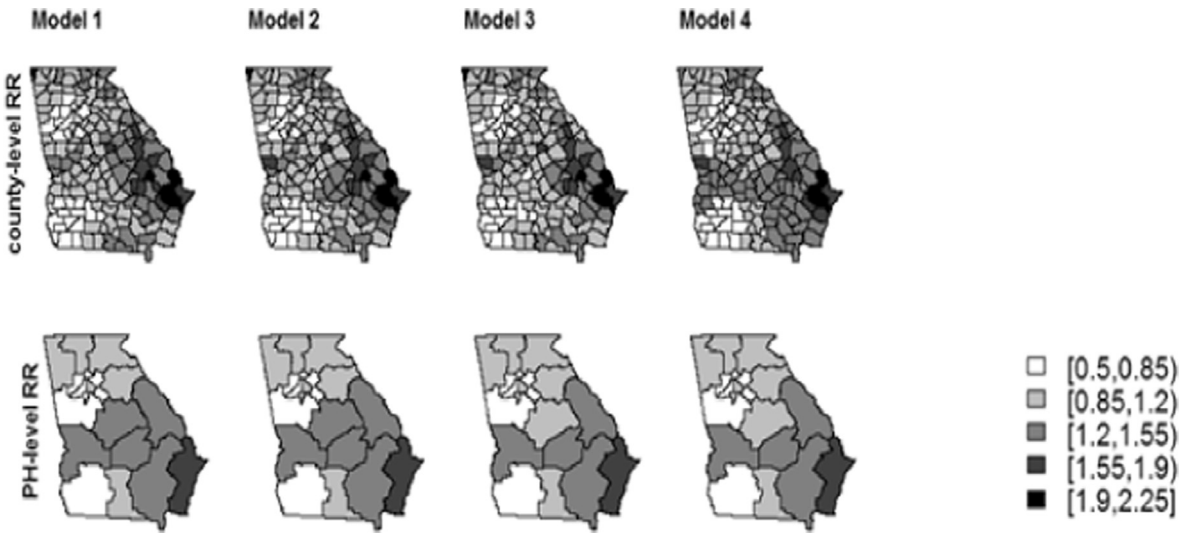

Fig. 7. Georgia oral cancer study: relative risk (RR) pattern obtained from the models fitted to the data at the county and public health (PH) levels.

spatial negative relationship with the incidence of oral cancer at the county level, whereas M3 and M4 reveal insignificant positive results at the PH level. We also computed the risk patterns obtained from these models as shown in the Supplementary Appendix, Fig. 6A. Here also, similar to Fig. 7 above, there is a high incidence of oral cancer in the eastern Georgia.

5. Discussion and conclusion

The goal of this paper is to compare a shared multilevel model (M2) and a shared multiscale model (M4) for data available at different geographical units. M2 only considers the data available at the fine level, while M4 describes the data available at both the fine and coarse levels. Furthermore, our proposed M2 can adjust for the contextual effect due to grouping of smaller areal units into larger units, whereas M4 can address both the contextual effects as well as the scaling effects. We also compared special cases of these models: the special case of the M2 is a convolution model (M1) that ignores the contextual effect, while the special case of M4 is an independent multiscale model (M3) that ignores both the contextual and scaling effects.

The models were implemented using both real and simulated data sets. When we simulated data with strong contextual effects, M2 and M4 recover the pattern of the simulated risk variations better than M1 and M3. M4 was also better than M2 as measured by DIC and WAIC. This could be because information is propagated from the fine to the coarse level and vice versa in M4. On the other hand, when there are weak contextual effects in the data, all the models provide similar risk variations. Yet, in terms of DIC and WAIC, M4 was slightly better than the other models. This shows that even with weak contextual effects, the shared components in M4 attempt to account for the scale effects due to data aggregation. For the real data set application, the M4 was the best model at both the county and PH levels.

The advantage of M2 over M4 is that it is a parsimonious model because it uses a convolution model only at the fine level to estimate the relative risk, while it estimates the relative risk at the coarse level by aggregating the relative risk estimates of the smaller areal units within the larger units. M4 provides better model fit and prediction accuracy as compared to M2. However, M4 uses the data twice at the fine and coarse levels because the data at the coarse level is an aggregation of the data at the fine level. In addition, M2 results in a more unbiased and precise estimates of the variance of the random effects at the coarse level. If someone is interested in inference of the spatial correlation between areas at the PH level, M2 is a better option than M4.

In summary, both the shared multilevel (M2) and shared multiscale models (M4) are useful to estimate risk variations at different geographically aggregated levels for public health planning purposes. We recommend using M2 if the objective is to incorporate the contextual effects in the model and if there is interest in the inference of the strength of the contextual effects, whereas we recommend using the M4 if the goal is to study both the contextual and scale effects simultaneously.

Acknowledgments

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

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sste.2017.06.001.

References

Aregay M, Lawson AB, Faes C, Kirby RS, Carroll R, Watjou K. Impact of income on small area low birth weight incidence using multiscale models. AIMS Pub Health 2015a;2(4):667–80.

Aregay M, Lawson AB, Faes C, Kirby RS, Carroll R, Watjou K. Multiscale measurement error models for aggregated small area health data. Stat Methods Med Res 2016a;25(4):1201–23.

Aregay M, Lawson AB, Faes C, Kirby RS, Carroll R, Watjou K. Spatial mixture multiscale modeling for aggregated health data. Biometrical J 2016b;58(5):1091–112. doi:10.1002/bimj.201500168.

Aregay M, Lawson AB, Faes C, Kirby R. Bayesian Multiscale modeling for aggregated disease mapping data. Stat Methods Med Res 2015b. doi:10.1177/0962280215607546.

Banerjee S, Carlin BP, Gelfand AE. Hierarchical modeling and analysis for spatial data. London: Chapman and Hall/CRC Press; 2004.

Basseville M, Benveniste A, Chou KC, Golden SA, Nikoukhah R, Willsky AS. Modeling and estimation of multiresolution stochastic processes. IEEE Trans Inf Theory 1992;38:766–84.

Berliner LM, Wikle CK, Milli RF. Multiresolution wavelet analyses in hierarchical bayesian turbulence models. In: Vidakovic B, Muller P, editors. Bayesian inference in wavelet-based models. New York: Springer; 1999. p. 341–59.

Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 1991;43(1):1–20. doi:10.1007/bf00116466.

Bobashev GV, Anthony JC. Clusters of marijuana use in the United States. Am J Epidemiol 1998;148:1168–74.

Browne WJ, Goldstein H, Rasbash J. Multiple membership multiple classification (MMMC) models. Stat Model 2001;1:103–24.

Calder C, Craigmile P, Zhang J. Regional spatial modeling of topsoil geochemistry. Biometrics 2009;65:206–15.

Chaix BJ, Merlo J, Chauvin P. Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilization in France. J Epidemiol Commun Health 2005;59:517–26.

Chou KC, Willsky AS, Nikoukhah R. Multiscale systems, kalman filters, and riccati equations. IEEE Trans Autom Control 1994;39:479–92.

Craigmile PF, Guttorp P. Space-time modelling of trends in temperature series. J Time Ser Anal 2011;32:378–95.

Cressie NAC. Change of support and the modifiable areal unit problem. J Geogr Syst 1996;2:83–101.

Delouille V, Jansen M, von Sachs R. Second generation wavelet methods for denoising irregularly spaced data in two dimensions. Signal Process 2006;86:1435–50.

Dong GP, Harris R, Jones K, Yu J. Multilevel modelling with spatial interaction effects with application to an emerging land market in Beijing, China. PLoS One 2015;10(6):e0130761. doi:10.1371/journal.pone.0130761.

Gelman A, Hwang J, Vehtari J. Understanding predictive information criteria for Bayesian models. Stat Comput 2014;24:997–1016.

Gelman A. Prior distribution for variance parameters in hierarchical models. Bayesian Anal 2006;3:515–33.

Goldstein H, Browne W, Rasbash J. Multilevel modelling of medical data. Stat Med 2002;21:3291–315.

Huang HC, Cressie N. Multiscale graphical modeling in space: applications to command and control. In: Moore M, editor. Proceedings of the spatial statistics workshop. New York: SpringerSpringer Verlag Publishers, 2000.

Huang HC, Cressie N, Gabrosek J. Fast, resolution-consistent spatial prediction of global processes from satellite data. J Comput Gr Stat 2002;11:63–88.

Johannesson G, Cressie N, Huang HC. Dynamic multi-resolution spatial models. Environ Ecolog Stat 2007;14:5–25.

Kolaczyk ED, Huang H. Multiscale statistical models for hierarchical spatial aggregation. Geogr Anal 2001;33:95–118.

Langford IH, Leyland AH, Rasbash J Goldstein. Multilevel modelling of the geographical distributions of diseases. J Roy Stat Soc C App 1999;48:253–68.

Lawson AB. Handbook of spatial epidemiology. Spatial longitudinal analysis. CRC Press; 2016 chapter 18.

Lawson AB. Bayesian disease mapping: hierarchical modeling in spatial epidemiology. 2nd ed. New York: Chapman and Hall/CRC Press; 2013.

Leyland AH, Goldstein H. Multilevel modelling of health statistics. Chichester: Wiley; 2001.

Louie MM, Kolaczyk ED. On the covariance properties of certain multiscale spatial processes. Stat Probab Lett 2004;66:407–16.

Louie MM, Kolaczyk ED. Multiscale detection of localized anomalous structure in aggregate disease incidence data. Stat Med 2006a;25:787–810.

Louie MM, Kolaczyk ED. A multiscale method for disease mapping in spatial epidemiology. Stat Med 2006b;25:1287–306.

Merlo J, Asplund K, Lynch JW, Rastam L, Dobson A. Population effects on individual systolic blood pressure-a multilevel analysis of WHO MONICA project. Am J Epidemiol 2004;159:1168–79.

Nychka D, Wikle CK, Royle JA. Multiresolution models for nonstationary spatial covariance functions. Stat Model 2002;2:315–31.

Plummer M. Penalized loss functions for Bayesian model comparison. Biostatistics 2008;9:523–39.

Preisser JS, Arcury TA, Quandt SA. Detecting patterns of occupational illness clustering with alternating logistic regressions applied to longitudinal data. Am J Epidemiol 2003;158:495–501.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). J R Stat Soc B 2002;64:583–616.

Tranmer M, Steel D, Browne WJ. Multiple-membership multiple-classification models for social network and group dependence. J R Stat Soc A Stat 2014;177:439–55.

van der Linde A. DIC in variable selection. Stat Neerl 2005;1:45–56.

Vidakovic B. Statistical modeling by wavelets. New York: John Wiley; 1999.

Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J Mach Learn Res 2010;11:3571–94.

Wikle CK, Milli R, Nychka D, Berliner LM. Spatiotemporal hierarchical bayesian modeling: tropical ocean surface winds. J Am Stat Assoc 2001;96:382–97.

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. 108–23.

Zhu J, Yue W. A multiresolution tree-structured spatial linear model. J Comput Gr Stat 2004;14(1):168–84.