A Multivariate Spatial Mixture Model for Areal Data: Examining Regional Differences in Standardized Test Scores

Brian Neelon†

Duke University, Durham, USA

Alan E. Gelfand

Duke University, Durham, USA

Marie Lynn Miranda

University of Michigan, Ann Arbor, USA

Summary. Motivated by a study exploring geographic disparities in test scores among fourth graders in North Carolina, we develop a multivariate mixture model for the spatial analysis of correlated continuous outcomes. The responses are modeled as a finite mixture of multivariate normals, which accommodates a wide range of marginal response distributions and allows investigators to examine covariate effects within subpopulations of interest. The model has a hierarchical structure incorporating both individual- and areal-level predictors as well as spatial random effects for each mixture component. Conditional autoregressive priors on the random effects provide spatial smoothing and allow the shape of the multivariate distribution to vary flexibly across geographic regions. By integrating over this distribution, we obtain region-specific joint, marginal, and conditional inferences of interest. We adopt a Bayesian modeling approach and develop an efficient posterior sampling algorithm that relies primarily on closed-form full conditionals. Our results show that students in the central and coastal counties of North Carolina demonstrate higher achievement on average than students in the other parts of the state. These findings can be used to guide county-level initiatives, such as school-based literacy programs, to improve elementary education.

1. Introduction

In 2002, the United States (U.S.) Congress enacted the No Child Left Behind (NCLB) Act requiring states to administer annual standardized tests to all students in federally funded schools (No Child Left Behind Act, 2002). In North Carolina, these tests are known as end-of-grade (EOG) tests. The EOG tests measure student performance on grade-based goals, objectives, and competencies as set forth by the state's education department (North Carolina Department of Public Instruction, 2006). In particular, the mathematics tests measure competency in areas such as arithmetic operations, measurement, and geometry, while the reading tests measure competency in areas such as vocabulary and reading comprehension. The raw EOG scores are subsequently categorized into four achievement levels: 1) insufficient mastery; 2) inconsistent mastery; 3) consistent mastery;

†Department of Biostatistics and Bioinformatics, Box 2721 Duke University Medical Center, Durham, NC 27710-2721, USA. E-mail: brian.neelon@duke.edu

and 4) superior performance (North Carolina Department of Public Instruction, 2007, 2008). Results of EOG tests have important implications for both individual schools and school districts, as they may affect state and federal funding levels.

Because scores can vary across geographic areas, there has been growing interest in examining regional differences in test scores, both at the national and state level. North Carolina, like many other states, is working to close the gap between low-performing schools and those meeting NCLB standards. Despite this goal, relatively few studies have examined geographic disparities in EOG performance in an effort to identify high- and low-performing schools and school districts. In fact, we found only one related study examining gender differences in test performance across large national Census divisions (Pope and Sydnor, 2010). Thus, there remains a need for a comprehensive study of varying test performance across a refined geographic scale. By pinpointing schools that fail to meet adequate yearly standards set forth by NCLB, state and local education officials can develop targeted interventions to improve school performance in the areas of most need. Directed efforts such as these provide new opportunities to close the achievement gap in EOG test scores.

With these goals in mind, researchers from Duke University and the University of Michigan recently conducted a study to better understand factors influencing variation in EOG scores among elementary school children from across North Carolina. As a first step, the investigators obtained math and reading test scores for fourth graders from all 100 countries in the state following completion of the 2008 school year, the most recent year for which such data were available. The data were then geo-referenced by residential address and subsequently linked at the county level to data from the 2005–2009 American Community Survey (U.S. Census Bureau, 2010). The aims of the study were to examine statewide variation in EOG test scores and to identify individual- and county-level predictors of EOG performance.

From an analytic perspective, the EOG data posed several unique challenges. First, math and reading scores tend to be highly correlated measures; we therefore needed a flexible spatial model that could account for both within-subject and within-county associations. We also wanted a model that could yield accurate predictions of average student performance for each county while inducing spatial smoothing among sparsely populated counties where predictions may be less reliable. And finally, we wanted a model that was robust to region-specific departures from normality in light of the skewness observed in the data. This paper describes a new multivariate spatial mixture model specifically designed to address these multiple aims.

Our proposed model capitalizes on recent developments in spatial modeling of multivariate, areal-referenced data, i.e., data in which the spatial units consist of discrete regions of space such as counties. Modeling of such data typically proceeds by introducing a set of region-specific random effects, which are then linked via a multivariate conditionally autoregressive (MCAR) prior distribution (Mardia, 1988). Previous applications of joint spatial models for areal data have focused on normal responses (Gelfand and Vounatsou, 2003), count responses for disease mapping (Carlin and Banerjee, 2002; Jin *et al.*, 2005; Zhang *et al.*, 2009; Congdon, 2010), and categorical responses (Gelfand and Vounatsou, 2003; Wall and Liu, 2009).

In many applications, the response variables are continuous but non-normally distributed, either due to multimodality, heavy skewness, or both. In such cases, mixture models can provide a flexible framework for modeling the response distribution and can improve model fit. There is a well-established literature on mixture models for non-spatial data (Mclachlan and Peel, 2000; Frühwirth-Schnatter, 2006). In the spatial setting, several authors have proposed mixture models for point-referenced data — i.e., data indexed by a set of specific geographic coordinates. Gelfand *et al.* (2005) used a Dirichlet process mixture model to examine precipitation measurements at fixed locations in southern France. Kottas and Sansó (2007) extended the approach by allowing the point locations to be random. Ji *et al.* (2009) used a similar Poisson point-process mixture model to identify cell abundance patterns from fluorescent intensity images of lymphatic tissue. For multivariate point-referenced data, Reich and Fuentes (2007) proposed a semiparametric mixture model specified through a stick-breaking process. In the areal setting, Green and Richardson (2002) and Lawson and Clark (2002) proposed univariate mixture models for mapping disease relative risks. More recently, Wall and Liu (2009) developed a spatial latent class model for multivariate binary data and modeled the latent class indicators using a multinomial probit model with spatially correlated error terms.

We extend this work by developing a multivariate spatial finite-mixture model for continuous, areal-referenced data. We introduce spatial random effects for each mixture component, as well as for the mixing weights, to allow the shape of the multivariate response distribution to vary in flexible ways across geographic regions and covariate profiles. As such, our model provides a practical approach to multivariate spatial density estimation. By integrating across this mixture density, one can obtain region-specific inferences and model-based predictions of interest. We adopt a Bayesian inferential approach, and for posterior computation develop an efficient Markov chain Monte Carlo (MCMC) algorithm that combines closed-form Gibbs and Metropolis steps.

The remainder of the paper is organized as follows: Section 2 describes the EOG testing data; Section 3 outlines the proposed model and discusses prior specification, posterior computation, and model selection; Section 4 presents results from two simulation studies highlighting important features of the model; Section 5 applies the method to the EOG data; and the final section provides a discussion and directions for future work.

2. The EOG Data

Table 1 provides a summary of the EOG data. For the purposes of our analysis, we restricted the sample to non-Hispanic white and non-Hispanic black students due to small sample sizes in other race and ethnicity groups and to the impact of English as a second language on early school performance. Of the 78380 students, roughly half were male, about one third were non-Hispanic black, and just over 43% received free or reduced-price lunch at school through a federal subsidy program. The math scores ranged from 319 to 373 with a median of 352, and the reading scores ranged from 313 to 370 with a median of 346. Approximately three quarters of the students achieved consistent mastery or higher on the math exam, and nearly 63% achieved consistent mastery or better on reading.

Table 1. Summary statistics for the EOG data ($N = 78380$).

Variable	Median (IQR)
Math Score	352 (345, 358)
Reading Score	346 (339, 353)
County Median Household Income ($)	44319 (39676, 51110)
County Sample Size	443 (203, 874)
	n (%)
Male	39555 (50.47)
Non-Hispanic Black Race†	25219 (32.18)
Enrolled in Free or Reduced-Price Lunch Program†	33959 (43.33)
Math Achievement Level	
Insufficient Mastery (Score $\leq$ 335)	4470 (5.70)
Inconsistent Mastery (Score 336–344)	15040 (19.19)
Consistent Mastery (Score 345–357)	37956 (48.43)
Superior Performance (Score $\geq$ 358)	20914 (26.68)
Reading Achievement Level	
Insufficient Mastery (Score $\leq$ 334)	11439 (14.59)
Inconsistent Mastery (Score 335–342)	17618 (22.48)
Consistent Mastery (Score 343–353)	30879 (39.40)
Superior Performance (Score $\geq$ 354)	18444 (23.53)

†Non-Hispanic black race and free lunch enrollment coded as binary (yes/no) variables.

Figure 1 presents a bivariate histogram of the raw math and reading scores (panel a), as well as a histogram of the standardized residuals based on an ordinary least squares (OLS) regression that included as predictors gender, race, enrollment in a free- or reduced-price lunch program, and county median household income (panel b). The distribution of the residuals is skewed toward lower values, particularly for reading, and the kurtoses in both directions are slightly negative. The univariate Kolmogorov-Smirnov tests on the residuals rejected the null hypothesis of normality ($p < 0.01$ for both outcomes), suggesting that the bivariate response distribution might be better modeled as a low-dimensional finite mixture of normals rather than as a single bivariate normal distribution.

There is also substantial variation in test scores across the state. Figure 2 shows the studentized OLS residuals averaged by county for both math and reading. The spatial pattern is similar for both math and reading, with negative residuals clustering in the interior northeast, along the eastern southern border, and in the western-most counties, while pockets of positive residuals appear in the center of the state and along the southern border in the west. This pattern suggests positive spatial autocorrelation in the residuals, violating the OLS assumption of independently distributed errors. This points to the need for a model that explicitly accounts for spatial dependence, since ignoring such spatial structure could lead to biased inferences and inaccurate assessments of parameter uncertainty.

(a) (b)

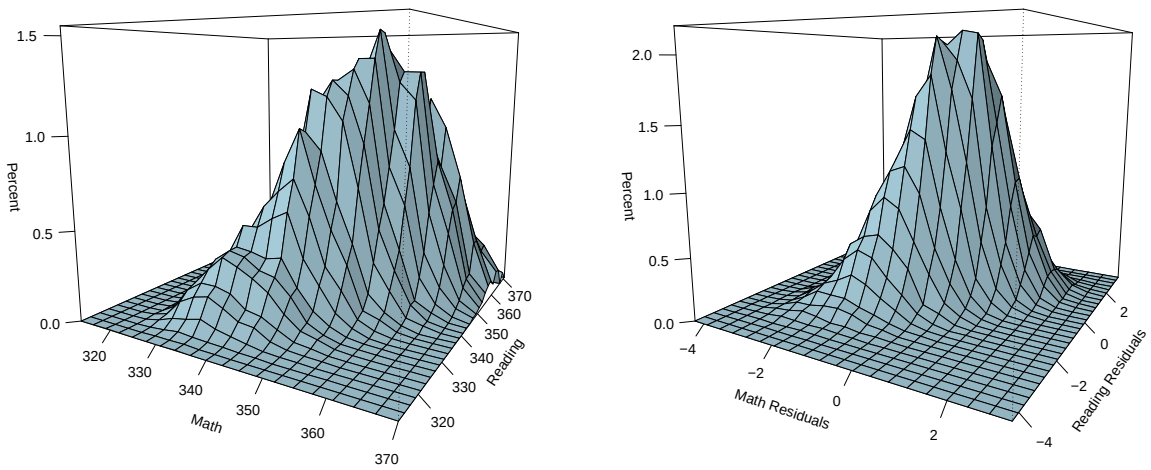

Fig. 1. Bivariate histogram of (a) the raw math and reading scores and (b) OLS standardized residuals for the 2008 fourth-grade EOG test scores.

3. Spatial Mixture Model

3.1. Model Specification

To develop the multivariate spatial mixture model, we focus on the bivariate case. It is conceptually straightforward to extend the approach to three or more outcomes.

A very general specification of the bivariate spatial mixture model can be expressed as:

$$\begin{aligned}
\boldsymbol{y}_{ij}|\boldsymbol{\phi}_i, \boldsymbol{\psi}_i &\sim \sum_{k=1}^{K} \pi_{ijk} \mathrm{N}_2(\boldsymbol{\eta}_{ijk}, \boldsymbol{\Sigma}_k) \\
\boldsymbol{\eta}_{ijk} &= \begin{pmatrix} \eta_{1ijk} \\ \eta_{2ijk} \end{pmatrix} = \boldsymbol{X}_{ij}\boldsymbol{\beta}_k + \boldsymbol{V}_i\boldsymbol{\alpha}_k + \boldsymbol{\phi}_{ik} \\
\pi_{ijk} &= \frac{\mathrm{e}^{\boldsymbol{x}'_{ij}\boldsymbol{\gamma}_k + \boldsymbol{v}'_i\boldsymbol{\delta}_k + \psi_{ik}}}{\sum_{h=1}^{K} \mathrm{e}^{\boldsymbol{x}'_{ij}\boldsymbol{\gamma}_h + \boldsymbol{v}'_i\boldsymbol{\delta}_h + \psi_{ih}}}, \qquad i = 1, \ldots, n;\ j = 1, \ldots, n_i;\ k = 1, \ldots, K, \qquad (1)
\end{aligned}$$

where $\boldsymbol{y}_{ij} = (y_{1ij}, y_{2ij})'$ denotes a 2×1 vector of math and reading scores for the j-th student in the i-th county; $\boldsymbol{X}_{ij} = \begin{pmatrix} \boldsymbol{x}'_{ij} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{x}'_{ij} \end{pmatrix}$ is a $2 \times 2p$ matrix of subject-level covariates with corresponding $2p \times 1$ component-specific fixed effects $\boldsymbol{\beta}_k = (\boldsymbol{\beta}'_{1k}, \boldsymbol{\beta}'_{2k})'$; $\boldsymbol{V}_i = \begin{pmatrix} \boldsymbol{v}'_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{v}'_i \end{pmatrix}$ is a $2 \times 2r$ matrix of county-level covariates with corresponding $2r \times 1$ component-specific fixed effects $\boldsymbol{\alpha}_k = (\boldsymbol{\alpha}'_{1k}, \boldsymbol{\alpha}'_{2k})'$; $\boldsymbol{\phi}_{ik} = (\phi_{1ik}, \phi_{2ik})'$ is a 2×1 vector of component-specific spatial random effects for the i-th county, with $\boldsymbol{\phi}_i = (\boldsymbol{\phi}'_{i1}, \ldots, \boldsymbol{\phi}'_{iK})'$; $\boldsymbol{\Sigma}_k$ is the component-k 2×2 variance-covariance matrix of $\boldsymbol{y}_{ij}$, conditional on $\boldsymbol{\phi}_{ik}$; $\boldsymbol{\gamma}_k$ and $\boldsymbol{\delta}_k$ are $p \times 1$ and $r \times 1$ vectors of mixing-weight regression

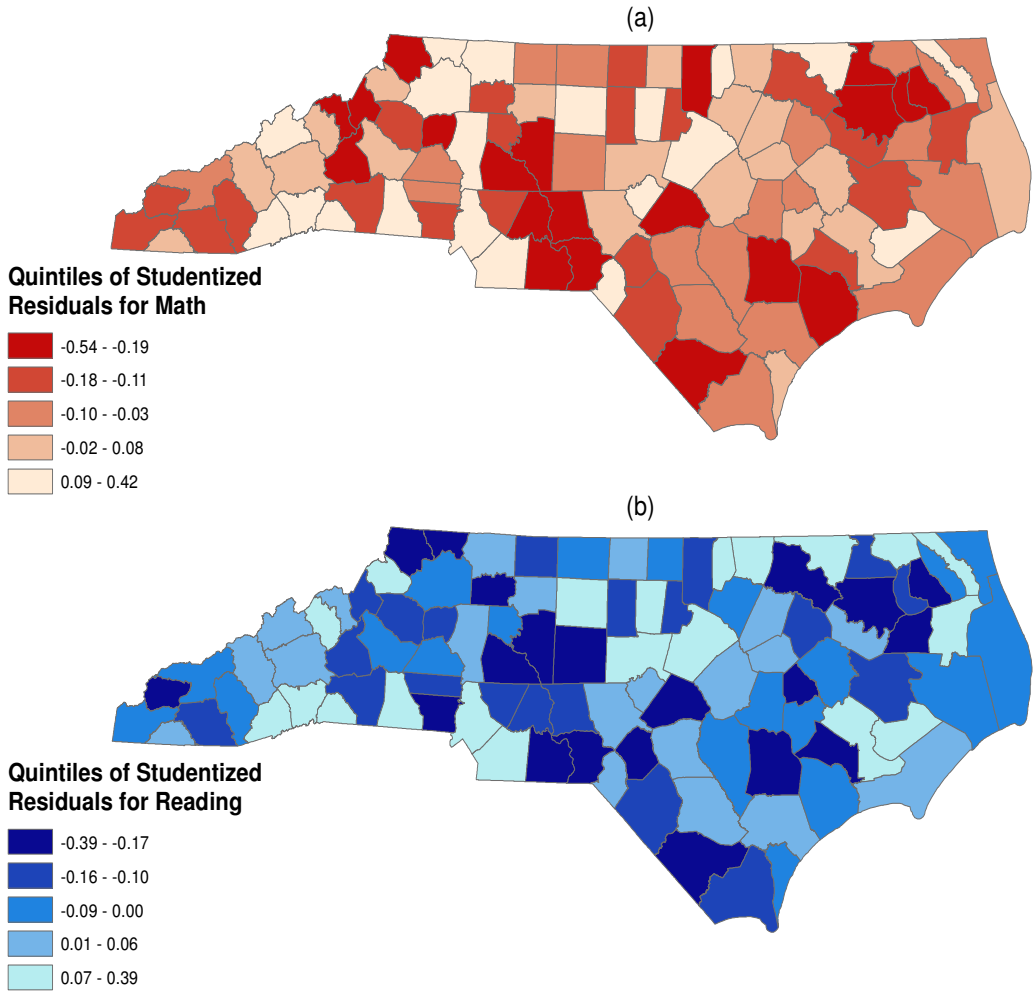

Fig. 2. Quintiles of county-averaged (a) math and (b) reading studentized residuals for the 2008 fourth-grade EOG scores.

parameters with $\boldsymbol{\gamma}_1 \equiv \mathbf{0}$ and $\boldsymbol{\delta}_1 \equiv \mathbf{0}$ for identifiability; and ψ_{ik} is a spatial random effect for county i and mixing weight k, where $\psi_{i1} \equiv 0$ and $\boldsymbol{\psi}_i = (\psi_{i1}, \ldots, \psi_{iK})'$. Throughout, we assume the same set of covariates for the component means and the mixing weights, although in general this restriction is not necessary.

Model (1) is appealing because it allows the shape of the joint response distribution to change flexibly across spatial units and covariate levels. In particular, the within-component linear predictors (the $\boldsymbol{\eta}_{ijk}$'s) permit the locations of the mixture components to vary throughout the population, while the mixing-weight parameters allow the mass of the response distribution to shift in unique ways between individuals and counties. Together, these features produce distinct response distributions for each covariate profile and areal unit, thus allowing us to obtain county-level predictions of EOG performance.

3.2. Prior Distributions

For parameter estimation, we adopt a fully Bayesian approach, assuming prior distributions for all model parameters. First, to allow for spatial smoothing and borrowing of information across counties, for each k, we assign component-specific conditionally autoregressive (CAR) priors (Besag, 1974; Besag *et al.*, 1991) to the spatial random effects — a bivariate CAR prior for $\boldsymbol{\phi}_{ik}$ and a univariate CAR prior for ψ_{ik}:

$$\boldsymbol{\phi}_{ik}|\boldsymbol{\phi}_{(-ik)}, \boldsymbol{\Lambda}_{ik} \sim \mathrm{N}_2\left(\xi_k \sum_{l\in\partial_i} \frac{w_{il}}{w_{i+}}\boldsymbol{\phi}_{lk}, \boldsymbol{\Lambda}_{ik}\right), \; k = 1, \ldots, K \tag{2}$$

$$\psi_{ik}|\psi_{(-ik)}, \tau_{ik}^2 \sim \mathrm{N}\left(\zeta_k \sum_{l\in\partial_i} \frac{w_{il}}{w_{i+}}\psi_{lk}, \tau_{ik}^2\right), \; k = 2, \ldots, K, \tag{3}$$

where ∂_i denotes the set of neighbors for county i, ξ_k and ζ_k are spatial smoothing parameters, w_{il} is an unnormalized proximity measure, $w_{i+} = \sum_{l\in\partial_i} w_{il}$, $\boldsymbol{\Lambda}_{ik} = \boldsymbol{\Lambda}_k/w_{i+}$ is a component-specific scaled variance-covariance matrix for $\boldsymbol{\phi}_{ik}$ conditional on $\boldsymbol{\phi}_{(-ik)}$, and $\tau_{ik}^2 = \tau_k^2/w_{i+}$ is a component-specific scaled variance parameter for ψ_{ik}. For the EOG study, we adopt intrinsic CAR (ICAR) priors to provide maximal smoothing of sparsely populated regions:

$$\boldsymbol{\phi}_{ik}|\boldsymbol{\phi}_{(-ik)}, \boldsymbol{\Lambda}_k \sim \mathrm{N}_2\left(\frac{1}{m_i}\sum_{l\in\partial_i}\boldsymbol{\phi}_{lk}, \frac{1}{m_i}\boldsymbol{\Lambda}_k\right) \tag{4}$$

$$\psi_{ik}|\psi_{(-ik)}, \tau_k^2 \sim \mathrm{N}\left(\frac{1}{m_i}\sum_{l\in\partial_i}\psi_{lk}, \tau_k^2/m_i\right), \tag{5}$$

where m_i denotes the number of neighbors sharing a geographic border with county i. Following Brook's Lemma (c.f., Banerjee *et al.*, 2004), priors (4) and (5) give rise to improper joint distributions for $\boldsymbol{\phi}_k$ and $\boldsymbol{\psi}_k$:

$$\boldsymbol{\phi}_k|\boldsymbol{\Lambda}_k \propto \exp\left(-\frac{1}{2}\boldsymbol{\phi}_k'\left[(\boldsymbol{M}-\boldsymbol{A})\otimes\boldsymbol{\Lambda}_k^{-1}\right]\boldsymbol{\phi}_k\right) \tag{6}$$

$$\boldsymbol{\psi}_k|\tau_k^2 \propto \exp\left(-\frac{1}{2\tau_k^2}\boldsymbol{\psi}_k'(\boldsymbol{M}-\boldsymbol{A})\boldsymbol{\psi}_k\right), \tag{7}$$

where $\boldsymbol{\phi}_k = (\boldsymbol{\phi}_{1k}', \ldots, \boldsymbol{\phi}_{nk}')'$, $\boldsymbol{\psi}_k = (\psi_{1k}, \ldots, \psi_{nk})'$, $\boldsymbol{M} = \mathrm{diag}(m_1, \ldots, m_n)$, and $\boldsymbol{A}$ is an $n \times n$ adjacency matrix with $a_{ii} = 0$, $a_{il} = 1$ if counties i and l are neighbors, and $a_{il} = 0$ otherwise. Because $(\boldsymbol{M} - \boldsymbol{A})$ is singular, the joint distributions in (6) and (7) are over-parameterized and thus improper, although the conditional prior distributions given by equations (4) and (5) are themselves proper. Propriety of the posterior, when a fixed effect intercept is included in the model, is achieved using a sum-to-zero constraint on the spatial random effects (Banerjee *et al.*, 2004).

To ensure a well-identified model, we assign weakly informative proper priors to the remaining model parameters. For the within-component fixed effects, we assume exchangeable normal priors: $\boldsymbol{\beta}_{1k}$ and $\boldsymbol{\beta}_{2k} \sim \mathrm{N}_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, $\boldsymbol{\alpha}_{1k}$ and $\boldsymbol{\alpha}_{2k} \sim \mathrm{N}_r(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha)$ for $k = 1, \ldots, K$. For the fixed effects within the mixing weights, $\boldsymbol{\gamma}_k$ and $\boldsymbol{\delta}_k$, we assign $\mathrm{N}_p(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma)$ and $\mathrm{N}_r(\boldsymbol{\mu}_\delta, \boldsymbol{\Sigma}_\delta)$ priors, respectively, for $k = 2, \ldots, K$. Throughout, we assume that the prior hyperparameters ($\boldsymbol{\mu}_\beta$, $\boldsymbol{\Sigma}_\beta$, etc.) are identical across components, but in general this is not required. To complete

the prior specification, we assign conjugate inverse-Wishart $\text{IW}(\kappa_0, \boldsymbol{S}_0)$ and $\text{IW}(\nu_0, \boldsymbol{D}_0)$ priors respectively to $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Lambda}_k$, and a conjugate inverse-gamma $\text{IG}(g, s)$ prior to τ_k^2 $(k = 2, \ldots, K)$.

Note that the proposed model accommodates a wide range of dependence structures. First, $\boldsymbol{\Sigma}_{12k}$, the off-diagonal element of $\boldsymbol{\Sigma}_k$, controls the component-specific within-subject association between outcomes. In the EOG study, for example, a positive value for $\boldsymbol{\Sigma}_{12k}$ implies that, for component k, students with high math scores also tend to have high reading scores conditional on the county-level random effects. Similarly, $\boldsymbol{\Lambda}_{12k}$, the off-diagonal element of $\boldsymbol{\Lambda}_k$, accounts for the component-specific, between-subject/within-region association between outcomes. In the EOG study, $\boldsymbol{\Lambda}_{12k} > 0$ implies that, for component k, counties with higher mean math scores tend to have higher mean reading scores, adjusting for observed covariates. And finally, the CAR priors on $\boldsymbol{\phi}_{ik}$ and ψ_{ik} capture associations between counties, implying that adjoining counties behave similarly with respect to their response distributions. Numerous submodels can be obtained by setting one or more of these association parameters to zero. For example, setting $\boldsymbol{\Lambda}_{12k} = 0\ \forall k$ implies no between-subject/within-county association in responses. This is tantamount to assigning separate univariate CAR priors to ϕ_{1ik} and ϕ_{2ik}. Further restricting $\boldsymbol{\Sigma}_{12k}$ to 0 for all k would imply that there is no within-subject association between responses, and hence the outcomes are uncorrelated at all levels of the model.

3.3. Posterior Computation and Model Comparison

Posterior inference proceeds via data augmentation by introducing a discrete latent labeling variable, C_{ij}, that takes the value $k\,(k = 1, \ldots, K)$ with probability π_{ijk} defined in equation (1). Letting $\boldsymbol{\theta}_k = \{\boldsymbol{\beta}_k, \boldsymbol{\alpha}_k, \boldsymbol{\phi}_k, \boldsymbol{\Sigma}_k, \boldsymbol{\Lambda}_k\}$ denote the within-component parameters and $\boldsymbol{\upsilon}_k = \{\boldsymbol{\gamma}_k, \boldsymbol{\delta}_k, \boldsymbol{\psi}_k, \tau_k^2\}$ denote the mixing-weight parameters, the joint posterior is given by:

$$\begin{aligned}\pi\left(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \boldsymbol{\upsilon}_2, \ldots, \boldsymbol{\upsilon}_K | \boldsymbol{y}\right) \propto & \prod_{k=1}^{K} \left\{ \prod_{i=1}^{n} \prod_{j=1}^{n_i} \left[\pi_{ijk} \text{N}_2(\boldsymbol{y}_{ij}; \boldsymbol{\eta}_{ijk}, \boldsymbol{\Sigma}_k)\right]^{\text{I}(C_{ij}=k)} \times \right. \\ & \left. \exp\left(-\frac{1}{2}\boldsymbol{\phi}_k' \left[(\boldsymbol{M} - \boldsymbol{A}) \otimes \boldsymbol{\Lambda}_k^{-1}\right] \boldsymbol{\phi}_k\right) \pi(\boldsymbol{\beta}_k)\pi(\boldsymbol{\alpha}_k)\pi(\boldsymbol{\Sigma}_k)\pi(\boldsymbol{\Lambda}_k) \right\} \times \\ & \prod_{h=2}^{K} \exp\left(-\frac{1}{2\tau_h^2}\boldsymbol{\psi}_h'(\boldsymbol{M} - \boldsymbol{A})\boldsymbol{\psi}_h\right) \pi(\boldsymbol{\gamma}_h)\pi(\boldsymbol{\delta}_h)\pi(\tau_h^2) \end{aligned} \tag{8}$$

where $\text{I}(\cdot)$ denotes the indicator function and the $\pi(\cdot)$'s represent the prior distributions for their respective parameters, as described in the previous section.

For posterior computation, we propose an MCMC algorithm that combines draws from full conditionals with Metropolis-based updates. After assigning initial values to the model parameters, the algorithm iterates between the following steps:

(a) For $k = 2, \ldots, K$, update $\boldsymbol{\gamma}_k$ and $\boldsymbol{\delta}_k$ using random-walk Metropolis steps;

(b) For $k = 2, \ldots, K$ and $i = 1, \ldots, n$, update ψ_{ik} using random-walk Metropolis;

(c) For $k = 2, \ldots, K$, update τ_k^2 from its closed-form full conditional distribution;

(d) For all (i, j), sample the mixture component indicators from a discrete distribution taking values $\{k = 1, \ldots, K\}$ with posterior probabilities $\{p_{ij1}, \ldots, p_{ijK}\}$ as described in the Appendix;

(e) For $k = 1, \ldots, K$, sample the component-specific parameters $\boldsymbol{\beta}_{1k}$, $\boldsymbol{\beta}_{2k}$, $\boldsymbol{\alpha}_{1k}$, $\boldsymbol{\alpha}_{2k}$, $\boldsymbol{\Sigma}_k$, $\boldsymbol{\Lambda}_k$, and $\boldsymbol{\phi}_{ik}$ $(i = 1, \ldots, n)$ from their closed-form full conditionals

(f) At the end of each MCMC iteration, apply a sum-to-zero constraint to $\boldsymbol{\phi}_k$ $(k = 1, \ldots, K)$ and $\boldsymbol{\psi}_k$ $(k = 2, \ldots, K)$.

Explicit details of the algorithm are provided in the Appendix. An appealing feature of the MCMC algorithm is that the within-component regression parameters and spatial effects have convenient closed-form full conditionals, leading to straightforward and efficient posterior sampling. Only the mixing-weight parameters require Metropolis-based updates.

Convergence is monitored by running multiple chains from dispersed initial values and performing standard Bayesian diagnostics, such as trace plots and evaluation of the Brooks-Gelman-Rubin statistic (Gelman *et al.*, 2004). Careful attention to such diagnostics is especially important for complex latent variable models to ensure parameter identifiability. In our experience, the proposed MCMC algorithm is generally robust to choices of initial values, with the possible exception of the mixing parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, which can be slow to converge for poorly chosen starting values. One way to choose initial values for these parameters is to perform a K-level cluster analysis, fit a multinomial logit regression to the resulting cluster indicators, and use the ensuing parameter estimates as starting values.

A well-known computational challenge for Bayesian finite mixture models is "label switching" in which draws of component-specific parameters may be associated with different components labels during the course of the MCMC run. Consequently, component-specific posterior summaries that average across the draws will be invalid. As a solution, Stephens (2000) proposed a post-hoc relabeling algorithm based on a Kullback-Leibler divergence function. We apply this approach for the analysis of the EOG data described in Section 5.

For model comparison, we adopt the deviance information criterion (DIC) proposed by Spiegelhalter *et al.* (2002). DIC includes a goodness of fit term along with a penalty for model complexity. Models with smaller DIC are considered preferable. For the EOG application, we apply a modified version of DIC recently recommended by Celeux *et al.* (2006) for finite mixture models. This modified DIC, termed DIC_3, uses the posterior predictive density of $\boldsymbol{y}$ to estimate the penalty term, and is closely related to a measure put forward by Richardson (2002) to avoid overfitting the number of mixture components.

Several approaches can be used to determine the number of classes, K. One pragmatic approach (and the one adopted here) is to impose an upper bound (say, $K_{\max}$) on the number of classes, and then use a model selection criteria such as DIC to choose the optimal value of $K = 1, \ldots, K_{\max}$. In our application, we set $K_{\max} = 2$ because we hypothesized the existence of at most two latent classes within the student population (namely, "high" and "low" performers), and because inspection of the unimodal residual plot in Figure 1(b) suggested that a low-dimensional mixture would adequately capture important features of the data. A more formal —

albeit computationally challenging — approach would be to treat K as a random variable and employ either a dimension-switching MCMC algorithm (Green, 1995) or recast the model as an infinite mixture vis-à-vis a stick-breaking representation (Sethuraman, 1994). We consider extensions to infinite mixtures further in the Discussion section.

4. Simulation Studies

To examine the performance of our model, we conducted two simulation studies. For the first study, we generated 200 datasets from a basic two-component mixture model without covariates. Our aims here were threefold: 1) to evaluate MCMC performance, 2) to ensure that we obtained reasonable parameter estimates under the true model, and 3) to highlight the model as a practical approach to spatial density estimation whereby the shape of the response distribution is allowed to vary flexibly across spatial units. To emulate the EOG data, we used the North Carolina county-level adjacency matrix for the simulation. This matrix contains 512 adjacencies among the 100 North Carolina counties. For the purposes of the simulation, counties were labeled "1" to "100". Because the $(\boldsymbol{M}-\boldsymbol{A})$ matrix in equations (6) and (7) is singular, the spatial random effects cannot be simulated directly. We therefore introduced the spatial smoothing parameters, ξ and ζ, defined in equations (2) and (3), and set them equal to $1-$ 1E-6. We then generated spatial random effects according to joint distributions (6) and (7) augmented with the smoothing parameters. Next, for each county, we simulated 80 math and reading scores from the following two-component mixture model:

$$\begin{aligned} \boldsymbol{y}_{ij}|\boldsymbol{\phi}_i, \boldsymbol{\psi}_i &\sim \sum_{k=1}^{2} \pi_{ijk} \mathrm{N}_2 \left[\begin{pmatrix} \beta_{10k} + \phi_{1ik} \\ \beta_{20k} + \phi_{2ik} \end{pmatrix}, \boldsymbol{\Sigma}_k \right] \\ \mathrm{logit}(\pi_{ij2}) &= \gamma_0 + \psi_i, \quad i = 1, \ldots, 100;\ j = 1, \ldots, 80, \end{aligned} \tag{9}$$

where π_{ij2} denotes the weight for the second mixture component. For $k = 1, 2$ we assigned independent $\mathrm{N}(0, 1000)$ priors to β_{10k} and β_{20k}, and $\mathrm{IW}(3, \boldsymbol{I}_2)$ priors to $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Lambda}_k$; for γ_0, we assigned a $\mathrm{N}(0, 1000)$ prior; for $\boldsymbol{\phi}_{ik} = (\phi_{1ik}, \phi_{2ik})'$ and ψ_i, we assigned bivariate and univariate intrinsic CAR priors, respectively; and for τ^2, the conditional variance of ψ_i, we assigned an $\mathrm{IG}(0.01, 0.01)$ prior.

For each simulation, we ran 2000 MCMC iterations in `R` version 2.14 (R Development Core Team, 2011) using a burn-in of 1000, which was sufficient to ensure convergence based on standard diagnostics. To avoid label switching, we simulated extremely well-separated mixture components, effectively imposing the order constraints $\beta_{101} < \beta_{102}$ and $\beta_{201} < \beta_{202}$.

Table 2 presents the posterior means, averaged across the 200 simulations, along with 95% coverage rates and the true parameter values used to generate the data. The posterior estimates showed minimal bias, and the coverage rates were near the nominal values for all parameters. Figure 3 displays the bivariate densities from a randomly selected simulation study. Panel (a) shows the true (i.e., simulated) density for an "average" county in which the random effects were set to 0. Panel (b) displays the corresponding model-estimated density, and

Table 2. Average posterior estimates and 95% coverage probabilities across 200 simulated datasets.

Mixture Component	Parameter	Description	Average Posterior True Value	Mean	95% Coverage
1	β_{101}	Math Intercept	340	340.01	0.93
	β_{201}	Reading Intercept	330	330.02	0.95
	$\boldsymbol{\Sigma}_{111}$	$\text{Var}(y_{1ij}\|\phi_{1i1})$	20	20.10	0.96
	$\boldsymbol{\Sigma}_{121}$	$\text{Cov}(y_{1ij}, y_{2ij}\|\boldsymbol{\phi}_{i1})$	10	10.17	0.94
	$\boldsymbol{\Sigma}_{221}$	$\text{Var}(y_{2ij}\|\phi_{2i1})$	36	36.34	0.95
	$\boldsymbol{\Lambda}_{111}$	$\text{Var}(\phi_{1i1})$	9	8.77	0.96
	$\boldsymbol{\Lambda}_{121}$	$\text{Cov}(\phi_{1i1}, \phi_{2i1})$	3	2.80	0.95
	$\boldsymbol{\Lambda}_{221}$	$\text{Var}(\phi_{2i1})$	4	3.49	0.91
2	β_{102}	Math Intercept	360	360.01	0.97
	β_{202}	Reading Intercept	345	345.01	0.96
	$\boldsymbol{\Sigma}_{112}$	$\text{Var}(y_{1ij}\|\phi_{1i2})$	50	50.04	0.96
	$\boldsymbol{\Sigma}_{122}$	$\text{Cov}(y_{1ij}, y_{2ij}\|\boldsymbol{\phi}_{i2})$	20	19.99	0.93
	$\boldsymbol{\Sigma}_{222}$	$\text{Var}(y_{2ij}\|\phi_{2i2})$	40	40.10	0.96
	$\boldsymbol{\Lambda}_{112}$	$\text{Var}(\phi_{1i2})$	4	3.53	0.93
	$\boldsymbol{\Lambda}_{122}$	$\text{Cov}(\phi_{1i2}, \phi_{2i2})$	6	5.73	0.92
	$\boldsymbol{\Lambda}_{222}$	$\text{Var}(\phi_{2i2})$	16	15.74	0.93
Mixing Weight Parameters	γ_0	Mixing Weight Intercept	0.75	0.75	0.96
	τ^2	$\text{Var}(\psi_i)$	1	1.00	0.94

panels (c) and (d) show the estimated densities for two randomly selected counties. The white circles denote the true component means: $\boldsymbol{\beta}_{01} = (340, 330)'$ and $\boldsymbol{\beta}_{02} = (360, 345)'$. As expected, the estimated average density in panel (b) closely mirrors the true density. Panel (c) shows a county in which the mixture components diverge and there is a shift in mass toward the lower component. In panel (d), the component locations shift toward higher math and reading values and the mass is concentrated on the upper component.

While we expect the proposed mixture model to perform well for multimodal data, one might wonder whether the same holds for unimodal but skewed data such as the EOG scores. To address this question, we conducted a second simulation in which we generated data from a 5-df bivariate skew-t distribution with location $\boldsymbol{\beta}_0 = (350, 360)'$, skewness $\boldsymbol{\lambda} = (2, -2)'$, and scale $\boldsymbol{\Sigma} = \begin{pmatrix} 300 & 50 \\ 50 & 400 \end{pmatrix}$. This parameterization yielded a bivariate distribution similar to the empirical histogram of EOG scores shown in Figure 1(b). As in the first simulation study, we used the North Carolina adjacency matrix and generated 80 observations from each of the 100 counties. Using bivariate CAR random effects, we allowed the location and skewness parameters to vary across the counties. We then fitted the two-component model given in equation (9). For comparison, we also fitted a single-component bivariate normal spatial model, which might be regarded as a conventional approach to analyzing such data. We also computed DIC statistics to further compare the one- and two-component models.

Figure 4 presents the true bivariate density (panel a) along with the estimated "average" densities (random effects set to zero) from the two-component (panel b) and one-component (panel c) models. Compared to the one-component model, the two-component model was able to reproduce the true density reasonably well, although

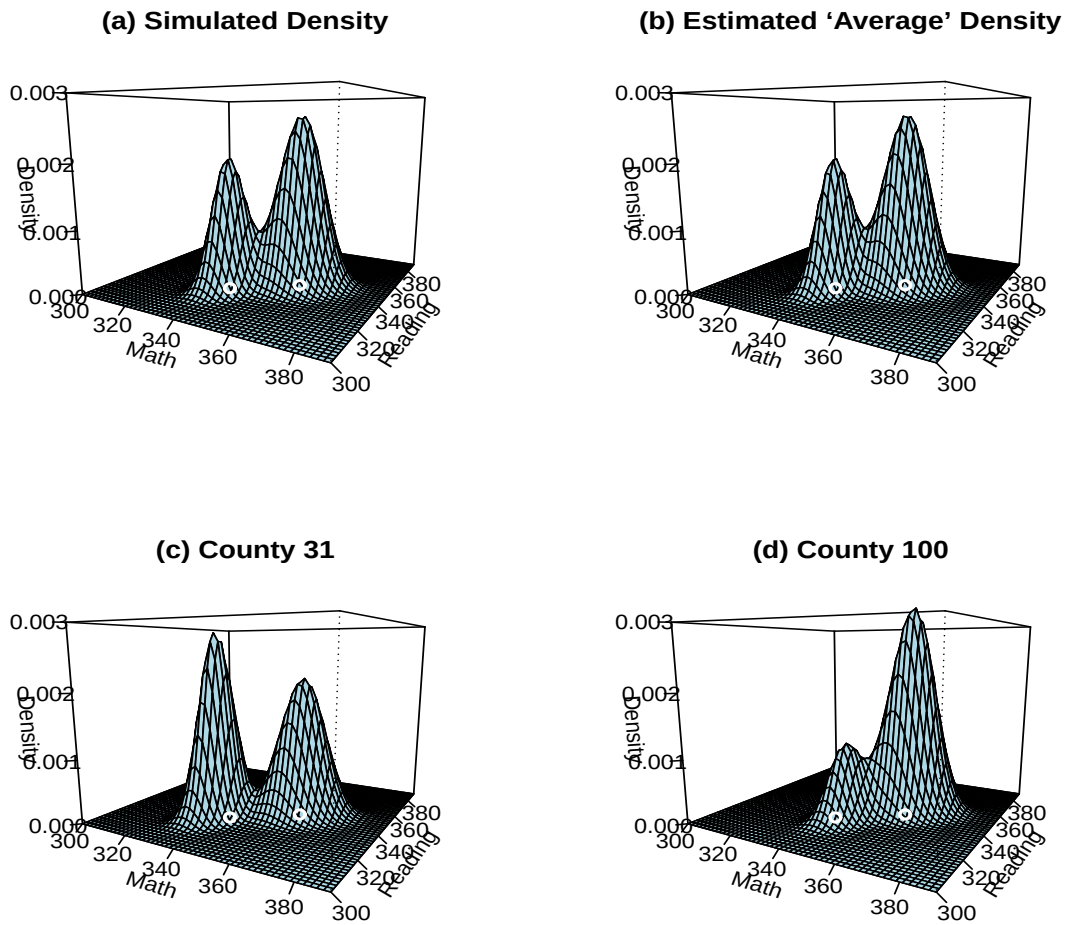

Fig. 3. True and estimated posterior densities from a randomly selected dataset from the first simulation study. Panel (a) shows the true simulated density for an "average" county (random effects set equal to zero); panel (b) depicts the corresponding model-estimated density; panel (c) presents the estimated density for a county (County 31) in which the components split apart; and panel (d) shows the estimated density for a county (County 100) in which the component locations shift toward higher values and more mass is concentrated on the upper component. White circles represent the true component means: $\boldsymbol{\beta}_{01} = (340, 330)'$ and $\boldsymbol{\beta}_{02} = (360, 345)'$.

even the two-component model failed to fully capture the kurtosis displayed in panel (a). The one-component model, on the other hand, was unduly influenced by the tails of the skew-t distribution; as a result, its mass is far more dispersed than that of the two-component model. In terms of DIC, the two-component model vastly outperformed the one-component model (139487 vs. 142566, respectively), confirming that the proposed method provides superior fit over the conventional approach even for unimodal, skewed data.

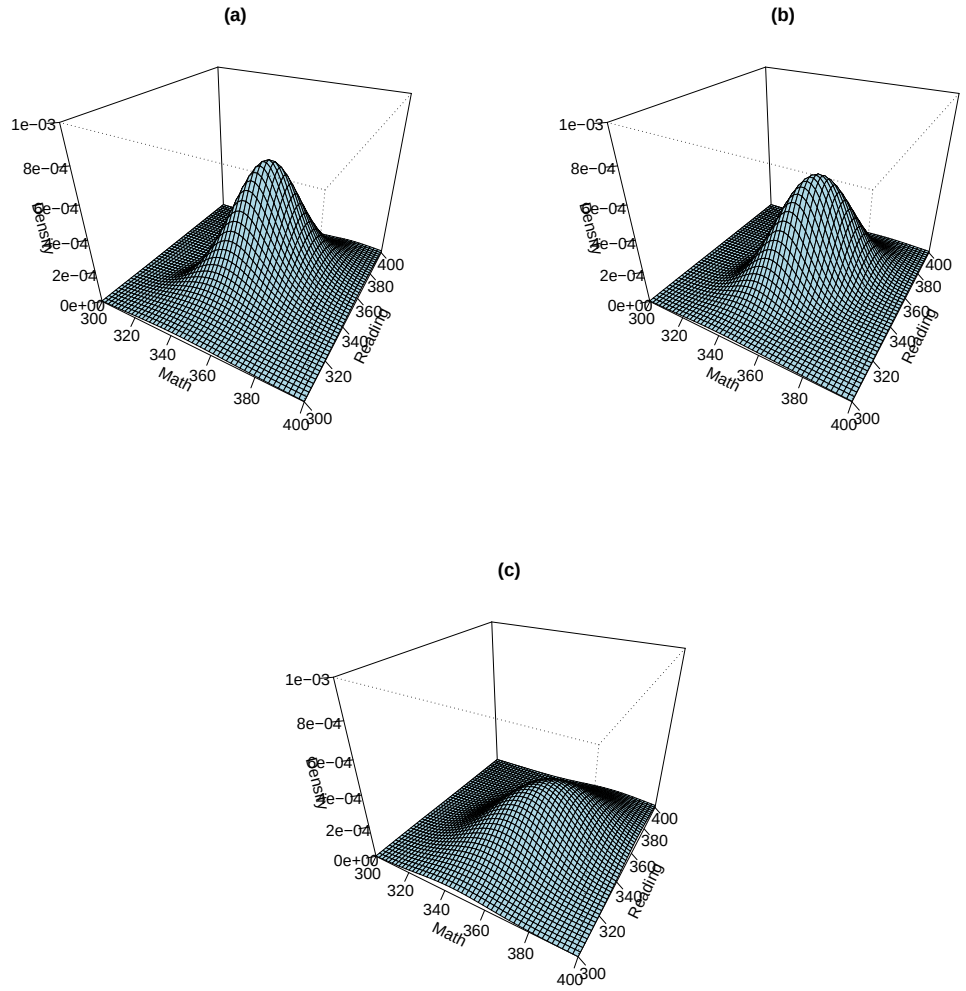

Fig. 4. True and estimated bivariate densities for simulation study 2. Panel (a) presents true 5-df bivariate skew-t distribution; panel (b) presents the estimated density for an "average county" (random effects set to zero) from the two-component mixture model; and panel (c) presents the estimated density for an average county from the one-component model.

Figure 5 presents contour plots comparing the true and two-component estimated densities, both overall and for selected counties. As panel (a) indicates, the two-component model accurately estimated the location of the true density, although with greater dispersion. In panels (b)-(d), the two-component model identified the location shifts in the county-specific densities, but was unable to model the extreme skewness exhibited in panels (c) and (d). While a higher-order mixture might provide better fit, identifiability problems could arise in attempting to fit a high-dimensional mixture to unimodal data. It is also possible that with only 80 replicates per county in the simulation study, there was insufficient sample size to capture all aspects of the county-specific densities. Ultimately, however, mixtures based on multivariate normals may lack the flexibility required to model extremely

skewed data. An alternative might be to develop a model based on mixtures of skew-elliptical distributions, such as skew-normal or skew-t densities. We consider this extension further in Section 6.

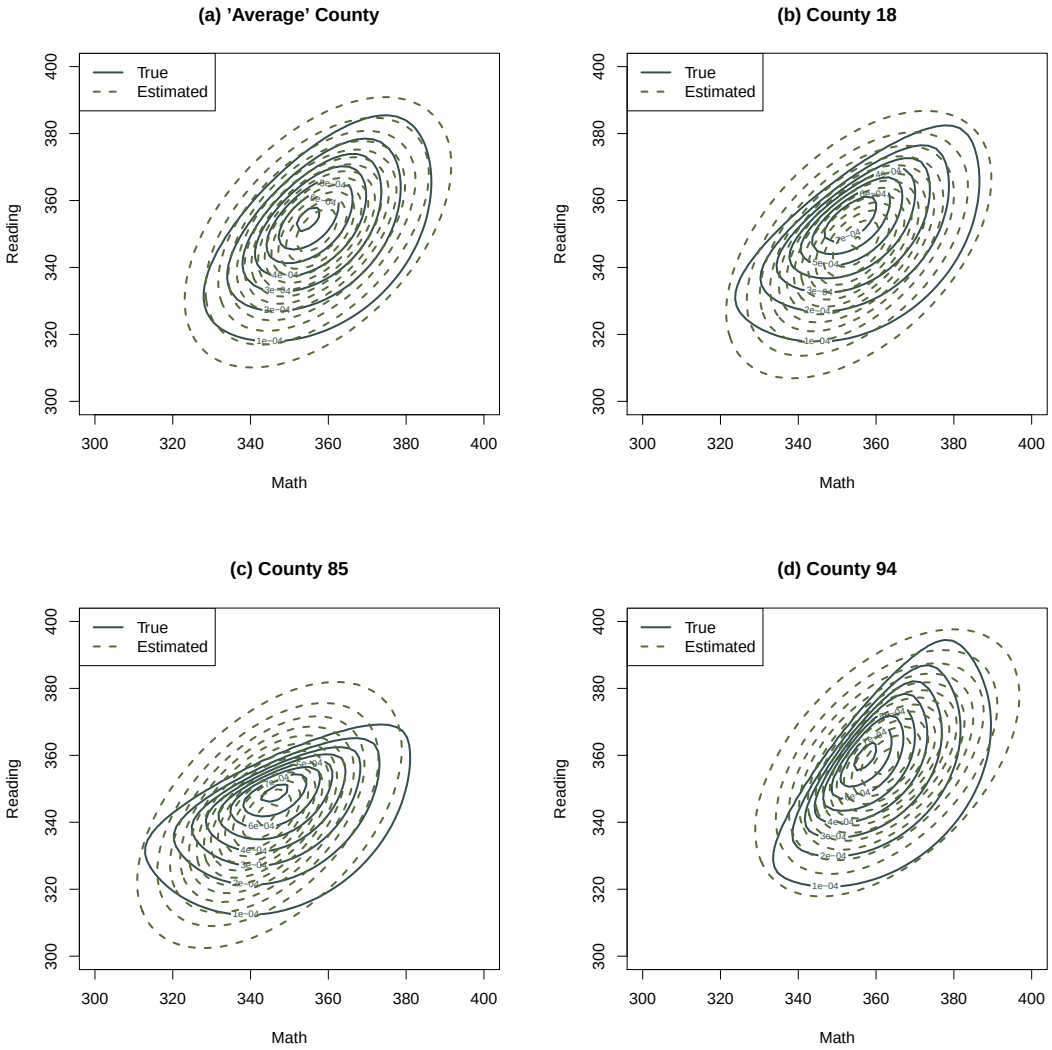

Fig. 5. Contour plots for (a) an "average" county and (b)-(d) selected simulated counties. Solid lines represent true densities and dashed lines are estimated contours from the two-component model.

5. Analysis of the EOG Data

Next, we fit the following two-component spatial mixture model to the EOG data:

$$\begin{aligned}
\boldsymbol{y}_{ij}|\boldsymbol{\phi}_i, \boldsymbol{\psi}_i &\sim \sum_{k=1}^{2} \pi_{ijk} \mathrm{N}_2 \left[\begin{pmatrix} \eta_{1ijk} \\ \eta_{2ijk} \end{pmatrix}, \boldsymbol{\Sigma}_k \right] \\
\eta_{1ijk} &= \beta_{10k} + \beta_{11k} \times \mathrm{Male}_{ij} + \beta_{12k} \times \mathrm{NHB}_{ij} + \beta_{13k} \times \mathrm{Freelunch}_{ij} + \alpha_{11k} \times \mathrm{Medinc}_i + \phi_{1ik} \\
\eta_{2ijk} &= \beta_{20k} + \beta_{21k} \times \mathrm{Male}_{ij} + \beta_{22k} \times \mathrm{NHB}_{ij} + \beta_{23k} \times \mathrm{Freelunch}_{ij} + \alpha_{21k} \times \mathrm{Medinc}_i + \phi_{2ik} \\
\mathrm{logit}(\pi_{ij2}) &= \gamma_0 + \gamma_1 \times \mathrm{Male}_{ij} + \gamma_2 \times \mathrm{NHB}_{ij} + \gamma_3 \times \mathrm{Freelunch}_{ij} + \delta_1 \times \mathrm{Medinc}_i + \psi_i, \\
&\qquad i = 1, \ldots, 100;\ j = 1, \ldots, n_i;\ k = 1, 2, \qquad (10)
\end{aligned}$$

where $\boldsymbol{y}_{ij}$ is a vector of math and reading scores for the j-th subject in the i-th county, Male is a dichotomous indicator of male gender, NHB is a dichotomous indicator taking the value 1 if the student is non-Hispanic black and 0 if non-Hispanic white, Freelunch is a dichotomous indicator taking the value 1 if the student participated in a free or reduced-price lunch program and 0 otherwise, Medinc denotes county median household income (in \$1000s), and π_{ij2} denotes the weight for the second mixture component.

As in the simulation studies, we assigned a bivariate CAR prior to $\boldsymbol{\phi}_{ik}$ and a univariate CAR prior to ψ_i. We assumed weakly informative proper priors for all other model parameters: for $\boldsymbol{\beta}_{1k} = (\beta_{10k}, \beta_{11k}, \beta_{12k}, \beta_{13k})'$, $\boldsymbol{\beta}_{2k} = (\beta_{20k}, \beta_{21k}, \beta_{22k}, \beta_{23k})'$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)'$, we assigned conjugate $\mathrm{N}_4(0, 1000\boldsymbol{I}_4)$ priors; for α_{11k}, α_{21k} and δ_1, we assigned $\mathrm{N}(0, 1000)$ priors; for $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Lambda}_k$, we assigned $\mathrm{IW}(3, \boldsymbol{I}_2)$ priors; and for τ^2 we assigned an $\mathrm{IG}(0.01, 0.01)$ prior. We ran two initially dispersed chains for 20000 iterations each, discarding the first 10000 as burn-in. To reduce autocorrelation, we retained every tenth iteration.

Model diagnostics indicated efficient mixing and rapid convergence of the chains. Figure (6) presents the post–burn-in trace plots for four selected model parameters: β_{111}, the component-1 math coefficient for male gender; γ_3, the mixing weight coefficient for Freelunch; τ^2, the variance of ψ_i; and $\boldsymbol{\Lambda}_{222}$, the variance of ϕ_{2i2}. The chains overlapped substantially, with no evidence of label switching within individual chains, and hence Stephens's (2000) relabeling algorithm converged rapidly. The component labels did require reordering across chains, but the proper labeling was easily identified so that the chains could be combined for summary purposes. For each of the four parameters, the Brooks-Gelman-Rubin upper credible interval was ≤ 1.21, indicating adequate convergence of the chains.

To evaluate the performance of our model, we compared our proposed model to three submodels:

1) A one-component version of the proposed spatial mixture model with a bivariate CAR prior for $\boldsymbol{\phi}_i$;
2) A two-component fixed-effects model excluding $\boldsymbol{\phi}_{ik}$ and ψ_i; and
3) A two-component model with component-specific $\boldsymbol{\phi}_{ik}$ but no ψ_i.

Table 3 provides the model comparison results for the various models. The two-component random effects models substantially outperformed submodels 1 and 2. Overall, the full model had the best performance,

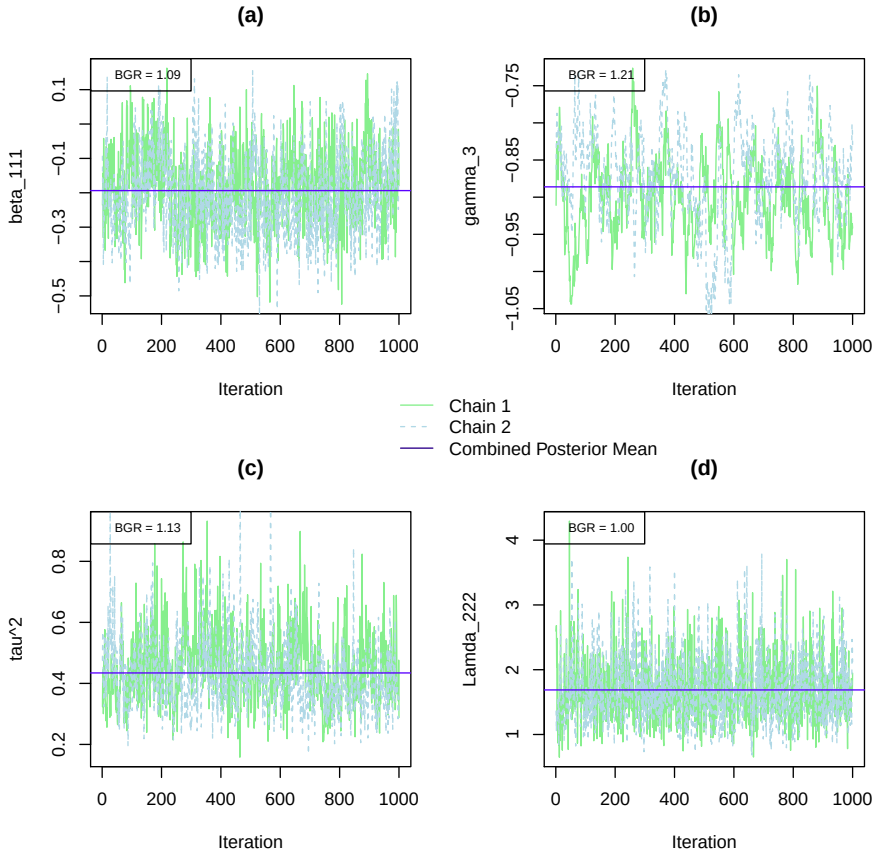

Fig. 6. Post–burn-in MCMC trace plots for four parameters from the proposed model: (a) β_{111}, the component-1 math coefficient for male gender; (b) γ_3, the mixing weight coefficient for Freelunch; (c) τ^2, the variance of ψ_i; and (d) $\mathbf{\Lambda}_{222}$, the variance of ϕ_{2i2}. Horizontal lines denote the posterior means from the combined chains. BGR = Brooks-Gelman-Rubin upper credible interval.

suggesting that incorporating the random effects ψ_i into the mixing weights provided a modest additional benefit relative to submodel 3.

Table 4 presents the posterior means and 95% credible intervals (CrIs) for the proposed model parameters. The results suggest that there are two distinct mixing components, or "latent subpopulations" of students. Subpopulation 1 contained an estimated 58% of the overall population, and was characterized by comparatively low mean math and reading scores for the reference group ($\beta_{101} = 349.25$ and $\beta_{201} = 344.50$). In addition, NHB race and free/reduced lunch enrollment were associated with lower math and reading scores. Higher county median income was associated with a slight increase in reading scores ($\alpha_{211} = 0.05$; 95% CrI=[0.00, 0.09]). Subpopulation 2 contained the remaining 42% of the overall population and was associated with roughly a 10-point increase in scores for the reference group compared to subpopulation 1 ($\beta_{102} = 358.04$ and $\beta_{202} = 354.15$).

Table 3. Model comparison statistics for analysis of the EOG data.

Model Description	$\overline{D}$	p_D	$\mathbf{DIC}_3$	$\boldsymbol{\Delta}$
One-Component Model	1068553	174	1068727	—
Two-Component Fixed Effects Model	1067255	28	1067283	1444
Two-Component Model Excluding ψ_i	1065144	295	1065439	1844
Proposed Model	1064990	328	1065318	121

In both groups, males had lower adjusted reading scores than females (e.g., $\beta_{211} = -1.74$; 95% CrI=[-1.98, -1.49]); however, in subpopulation 2, males had higher adjusted math scores ($\beta_{112} = 0.91$; 95% CrI=[0.71, 1.13]). These findings are consistent with previous research on gender disparities in standardized test scores, which has shown that boys tend to score lower than girls in reading but modestly higher in math and science, particularly in the upper tails of the test-score distribution (Pope and Sydnor, 2010). In terms of variability, subpopulation 1 displayed more individual-level and county-level heterogeneity than subpopulation 2. Not surprisingly, the within-subject correlations between math and reading (captured by ρ_1 and ρ_2) were moderately high for both subpopulations. In addition, the model accounted for almost all of the spatial association in the data: Moran's tests on the county-averaged residuals from the model yielded no evidence of residual spatial correlation ($p = 0.99$ for the math residuals and $p = 0.84$ for the reading residuals).

Figure 7 presents the predicted math and reading scores by county for reference group individuals (i.e., non-Hispanic white females not enrolled in a free- or reduced-price lunch program). Here, the between-county variation reflects both observed differences in county median household income as well as the latent heterogeneity captured by the spatial random effects. The spatial patterns of the predicted scores were similar to those for the OLS residuals in Figure 2, but with increased smoothing, particularly among the central "Piedmont" and southwest counties. This feature is expected, since the CAR priors act as spatial smoothers. In general, the northeastern and central Piedmont counties had higher predicted scores than those in the interior northeast, south central, and southwestern portions of the state. The predicted math scores for the reference group ranged from 352 to 356 across the state (panel a). Outlined on the map are Union County (along the southern border), which had the highest predicted math score at 356.4 [95% CrI=[355.5, 357.5]), and Avery County (in the northwest), which had the lowest predicted score at 352.1 [95% CrI=[350.2, 354.4]). There was also modest variation in the predicted reading scores across the state (panel b). The predicted reading scores ranged from 348 to 353, and showed a spatial pattern similar to the predicted math scores. Camden County, located in the northeastern corner of the state, had the highest predicted reading score (352.6, 95% CrI=[351.0, 354.3]), and Graham County in the southwest had the lowest predicted score (347.6, 95% CrI=[345.2, 349.7]).

Figure 8 displays four model-estimated bivariate densities for the reference cohort. Panel (a) presents the density for an "average" county with household income set at the statewide median value and random effects fixed at 0. Panels (b)-(d) display the densities for three selected counties: Camden County, which had the highest predicted reading score and a top 10% math score; Bertie County, which was in the bottom 5% for both math and reading; and Orange County, which had top 5% predicted math and reading scores. The county-

Table 4. Posterior means and 95% credible intervals (CrIs) for the proposed model.

Mixture Component (%)	Parameter	Description	Posterior Mean	95% CrI
1 (58%)	β_{101}	Math Intercept	349.25	(348.67, 349.77)
	β_{111}	Male	-0.19	(-0.41, 0.05)
	β_{121}	NHB Race	-3.12	(-3.45, -2.77)
	β_{131}	Free/Reduced Lunch	-2.50	(-2.83, -2.18)
	α_{111}	Median HH Income ($1000s)	0.03	(-0.01, 0.08)
	β_{201}	Reading Intercept	344.50	(343.84, 345.07)
	β_{211}	Male	-1.74	(-1.98, -1.49)
	β_{221}	NHB Race	-2.95	(-3.31, -2.56)
	β_{231}	Free/Reduced Lunch	-2.98	(-3.32, -2.62)
	α_{211}	Median HH Income ($1000s)	0.05	(0.00, 0.09)
	$\mathbf{\Sigma}_{111}$	$\text{Var}(y_{1ij}\|\phi_{1i1})$	57.58	(55.77, 59.34)
	$\mathbf{\Sigma}_{121}$	$\text{Cov}(y_{1ij}, y_{2ij}\|\boldsymbol{\phi}_{i1})$	35.26	(33.11, 37.19)
	$\mathbf{\Sigma}_{221}$	$\text{Var}(y_{2ij}\|\phi_{2i1})$	67.56	(65.37, 69.63)
	ρ_1	$\text{Corr}(y_{1ij}, y_{2ij}\|\boldsymbol{\phi}_{i1})$	0.57	(0.55, 0.58)
	$\mathbf{\Lambda}_{111}$	$\text{Var}(\phi_{1i1})$	7.95	(5.36, 11.31)
	$\mathbf{\Lambda}_{121}$	$\text{Cov}(\phi_{1i1}, \phi_{2i1})$	4.58	(2.58, 7.14)
	$\mathbf{\Lambda}_{221}$	$\text{Var}(\phi_{2i}\phi_{2i1})$	5.04	(3.09, 7.61)
2 (42%)	β_{102}	Math Intercept	358.04	(357.69, 358.41)
	β_{112}	Male	0.91	(0.71, 1.13)
	β_{122}	NHB Race	-3.35	(-3.77, -2.96)
	β_{132}	Free/Reduced Lunch	-2.71	(-3.03, -2.37)
	α_{112}	Median HH Income ($1000s)	0.02	(-0.02, 0.05)
	β_{202}	Reading Intercept	354.15	(353.77, 354.52)
	β_{212}	Male	-0.37	(-0.60, -0.14)
	β_{222}	NHB Race	-3.55	(-3.96, -3.09)
	β_{232}	Free/Reduced Lunch	-2.73	(-3.07, -2.37)
	α_{212}	Median HH Income ($1000s)	0.05	(0.02, 0.08)
	$\mathbf{\Sigma}_{112}$	$\text{Var}(y_{1ij}\|\phi_{1i2})$	36.08	(34.63, 37.53)
	$\mathbf{\Sigma}_{122}$	$\text{Cov}(y_{1ij}, y_{2ij}\|\boldsymbol{\phi}_{i2})$	21.76	(21.61, 22.93)
	$\mathbf{\Sigma}_{222}$	$\text{Var}(y_{2ij}\|\phi_{2i2})$	40.52	(38.95, 42.17)
	ρ_2	$\text{Corr}(y_{1ij}, y_{2ij}\|\boldsymbol{\phi}_{i2})$	0.57	(0.56, 0.58)
	$\mathbf{\Lambda}_{112}$	$\text{Var}(\phi_{1i2})$	3.35	(2.00, 5.20)
	$\mathbf{\Lambda}_{122}$	$\text{Cov}(\phi_{1i2}, \phi_{2i2})$	1.65	(0.71, 2.84)
	$\mathbf{\Lambda}_{222}$	$\text{Var}(\phi_{2i2})$	1.69	(0.91, 2.84)
Mixing Weight Parameters	γ_0	Mixing Weight Intercept	0.15	(-0.06, 0.37)
	γ_1	Male	-0.04	(-0.13, 0.04)
	γ_2	NHB Race	-1.26	(-1.39, -1.11)
	γ_3	Free/Reduced Lunch	-0.89	(-1.01, -0.77)
	δ_1	Median Household Income	0.02	(0.01, 0.04)
	τ^2	$\text{Var}(\psi_i)$	0.43	(0.25, 0.69)

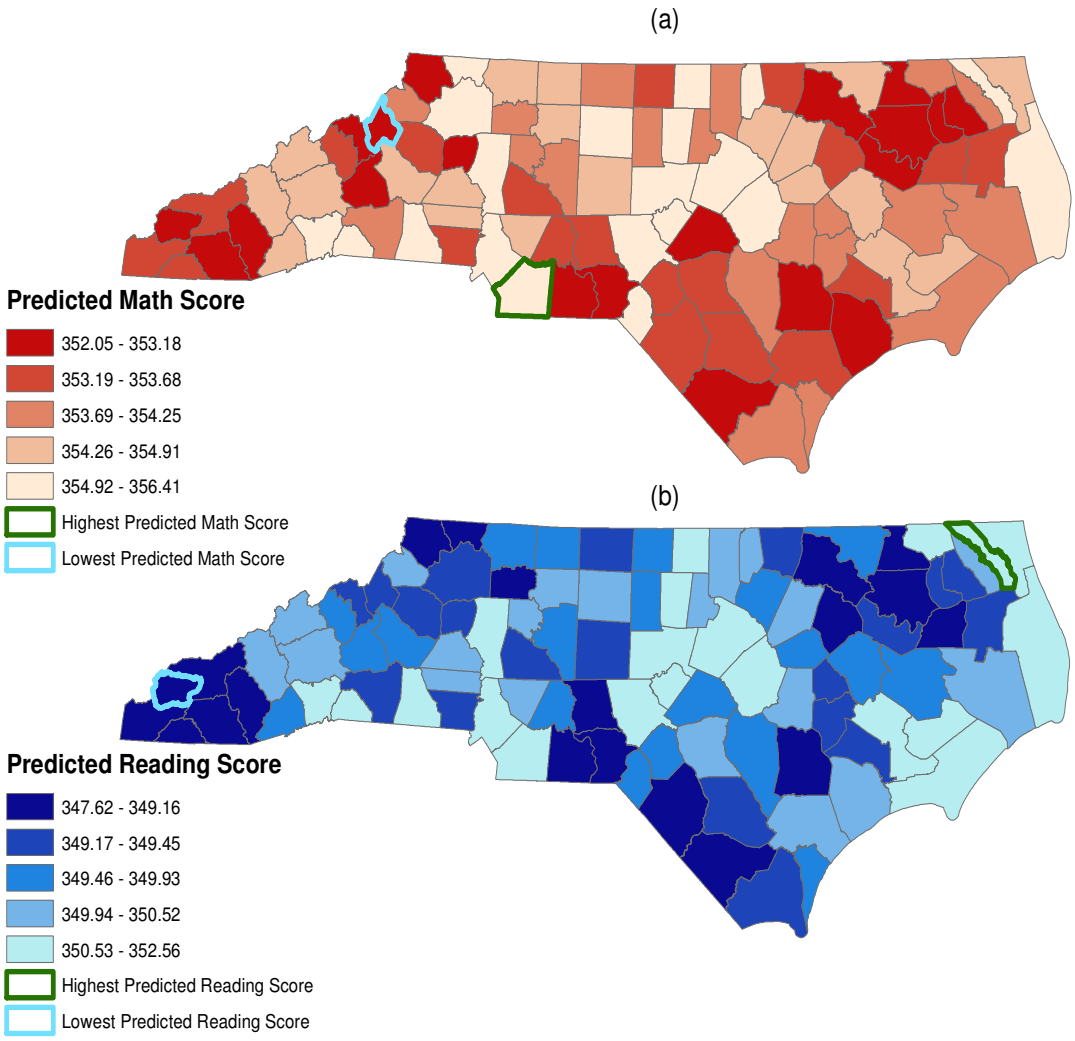

Fig. 7. Predicted (a) math and (b) reading scores for reference group, by county.

specific densities vary in both their location and distribution of mass relative to the average density presented in panel (a). In particular, Camden County shifted toward higher math and reading scores, reflecting the more favorable outcomes for this county, while Bertie county had a noticeable shift in mass toward the lower mixture component. Orange County, like Camden County, showed a shift in mass toward more favorable outcomes, but with a longer-tailed distribution, reflecting more heterogeneity for this county.

By integrating across these bivariate densities, we can obtain county-specific predictions of interest. For example, to predict an individual's joint probability of inconsistent or insufficient mastery in both the math and reading, defined as a math score < 344 and a reading score < 342 (Table 1), we evaluated the following integral:

$$(1-\pi_{ij2})\int_{-\infty}^{344}\int_{-\infty}^{342}\mathrm{N}_2(\boldsymbol{\eta}_{ij1},\boldsymbol{\Sigma}_1)\mathrm{d}y_{2ij}\mathrm{d}y_{1ij}+\pi_{ij2}\int_{-\infty}^{344}\int_{-\infty}^{342}\mathrm{N}_2(\boldsymbol{\eta}_{ij2},\boldsymbol{\Sigma}_2)\mathrm{d}y_{2ij}\mathrm{d}y_{1ij}.$$

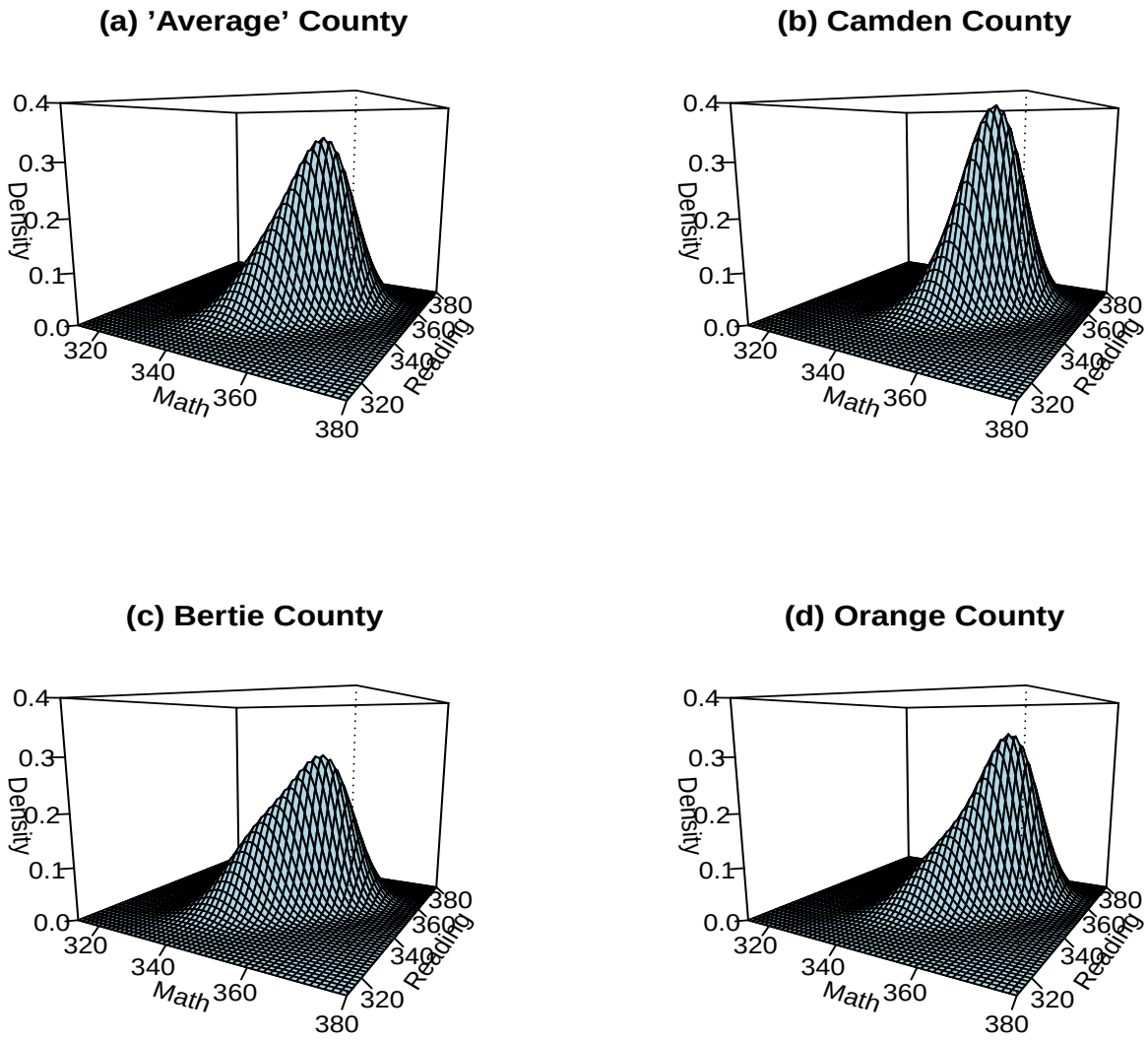

Fig. 8. Estimated county-specific densities for the reference group. Panel (a) presents the estimated density for an "average" county, and panels (b)-(d) show the estimated densities for three selected counties.

A similar approach was used to compute additional joint probabilities of interest. Using the cut-off values defined in Table 1, we calculated four joint probabilities of policy interest: (a) the probability of inconsistent or insufficient mastery in both math and reading (labeled "low math/low reading"); (b) the probability of inconsistent or insufficient math, but consistent or superior reading ("low math/high reading"); (c) the probability of consistent or superior math, but inconsistent or insufficient reading ("high math/low reading"); and (d) the probability of consistent or superior performance on both exams ("high math/high reading"). Figure 9 displays the county-

specific averages for these four joint probabilities for the reference group.

In general, the central Piedmont and northeastern coastal counties had more favorable outcomes than counties located in the interior northeast, along the southern border in the east, and in the north- and southwest corners of the state. The probability of low math/low reading achievement ranged from 0.04 to 0.13 (panel a), with Camden having the lowest (0.04, 95% CrI = [0.02, 0.06]) and Bertie County having the highest probability (0.13, 95% CrI=[0.09, 0.17]). These results are consistent with the predicted densities shown in Figures 6 for these counties. As panel (b) indicates, the combination of low math and high reading was a relatively rare event, with average probabilities ranging from 0.02 to 0.07 across the state. This result is not surprising, since students typically perform much better on math than on reading. One notable exception was Avery County along the central western border, which had the lowest predicted math score (c.f., Figure 7), but a reading score close to the state average at 349.40 (95% CrI=[347.67, 351.46]). High math/low reading performance was a more common phenomenon (panel c), ranging from 0.06 to 0.16 statewide, with Graham County having the highest probability (0.16, 95% CrI=[0.11, 0.22]). As panel (d) suggests, most students in the reference group demonstrated consistent or superior achievement on both exams. The county averages ranged from 0.69 in Bertie County (95% CrI=[0.65, 0.82]) to 0.87 in Camden County (95% CrI=[0.81, 0.93]), again supporting the results shown in Figures 6(b) and 6(c).

6. Discussion

Closing the achievement gap between high- and low-performing schools is a necessary step to help all children succeed in school (Ferguson, 2008). Although NCLB was designed to help close the gap by setting standards for evaluating schools, significant disparities remain, both within North Carolina and across the U.S. (National Center for Education Statistics, 2013). Therefore, it is an important public policy goal to examine regions of varying school performance and to identify factors associated with differences in student achievement. This will enable state education officials to direct resources to areas of greatest need. However, because few studies have examined regional differences in student achievement, there remains a need for a systematic approach to identifying geographic areas of high and low performance.

This paper has described a new spatial mixture model designed to meet these goals. The model incorporates individual- and region-level predictors, accommodates complex dependence structures, enables investigators to examine covariate effects across subgroups of the population, and supports region-specific departures from normality through the inclusion of spatial random effects for both the location parameters and the mixing weights. By integrating across this multivariate density, one can obtain region-specific joint, marginal, and conditional inferences of interest. In the EOG study, for example, we were able to compute county-specific joint probabilities of low math/low reading performance, low math, high reading performance, etc. We specified the model within a Bayesian framework and, for posterior computation, we developed a tractable MCMC algorithm that relies in large part on easy-to-sample Gibbs steps.

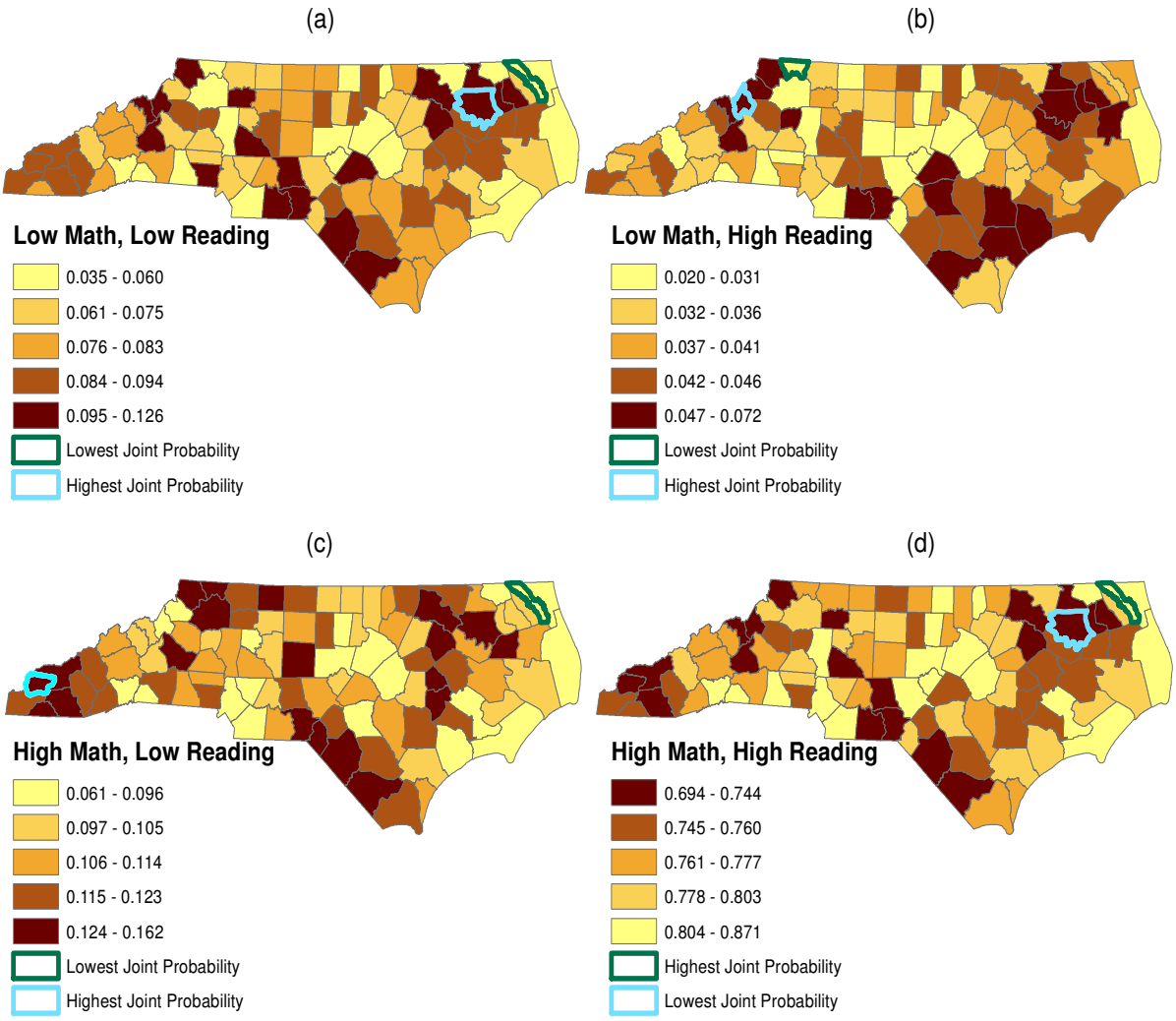

Fig. 9. Predicted joint probabilities of (a) low math, low reading; (b) low math, high reading; (c) high math, low reading; and (d) high math, high reading for the reference group.

Our exploratory analysis of the EOG data is, to our knowledge, the first to use advanced multivariate spatial modeling to examine geographic patterns in EOG scores at a refined geographic level. Through our analysis, we found that non-Hispanic black race and enrollment in subsidized lunch programs were associated with lower test scores. We also found gender gaps in EOG performance, with girls performing substantially better in reading, particularly among students with low EOG scores, and boys scoring slightly higher in math, especially among "high-achieving" students. These results are consistent with previous research on gender disparities in standardized test scores (Pope and Sydnor, 2010; National Center for Education Statistics, 2013), which has shown that boys tend to score lower than girls on standardized reading tests, but generally perform better in math and science, particularly in the upper tails of the response distribution (i.e., among high-performing

students). Together, these findings suggest the need to target two sets of gender disparities: first, in reading among students with low EOG performance and second, in math among those with high EOG scores. Future work might also examine gender-by-county interactions vis-à-vis a spatial random-coefficients model, to determine if these disparities vary by region, a finding documented at the national level in prior work (Pope and Sydnor, 2010).

Our analysis revealed similar spatial patterns for math and reading scores, with the central Piedmont and northern coastal counties displaying higher scores on average than counties in the interior northeast, south central, and western-most portions of the state. These findings could be used to guide county-level initiatives to improve elementary education. For example, by identifying counties with comparatively low reading scores, local school boards could introduce school-based literacy programs to improve vocabulary and reading comprehension. This is especially relevant in counties such as Graham County, which had the highest rate of high math/low reading performance.

Our simulations demonstrated that the proposed mixture model outperforms the conventional, single-component model, even for skewed unimodal data. However, by relying on multivariate normal distributions, we failed to fully capture the skewness observed for some of the counties in the second simulation study. An interesting extension, therefore, would be to develop models based on mixtures of skew-normal or skew-t distributions, building on recent work in the univariate non-mixture setting (Genton and Zhang, 2012; Nathoo and Ghosh, 2013; Zareifard and Khaledi, 2013). This would yield a highly flexible and computationally tractable parametric model that could accommodate both multimodality and extreme skewness.

In our application, we considered two-component mixture models, but higher-dimensional mixtures can be envisioned. More generally, one might choose to model the multivariate distribution nonparametrically via infinite mixtures, extending the recent work of Kottas *et al.* (2008) to allow for multivariate responses. Extensions to more than two outcomes are also straightforward, although as the dimension of the problem increases, it may become necessary to impose a structure on the class-specific error covariances, $\mathbf{\Sigma}_k$, to aid in identifying variance components. In high-dimensional settings, a factor analytic approach could be used to further encourage dimension reduction.

Our modeling approach should also have broad applicability to other research settings. For example, the spatial mixture model could be applied in reproductive epidemiology to explore joint spatial patterns in birth outcomes and to obtain region-specific joint probabilities of low birth weight and preterm birth. It might also find use in medical imaging as an extension of recent univariate methods (Ismail *et al.*, 2013). In short, the proposed model provides a pragmatic and flexible approach for the joint analysis of multivariate areal data. The MCMC algorithm described in Section 3 offers a computationally tractable method for fitting such models.

Acknowledgements

This work was supported by grant RD-83329301-0 from the U.S. Environmental Protection Agency, and was conducted in accordance with a human subjects research protocol approved by Duke University's Institutional Review Board. We are grateful to Claire Osgood for database management and support.

Appendix: MCMC Algorithm

1) **Update $\boldsymbol{\gamma}_k$ and $\boldsymbol{\delta}_k$**: The full conditional for p-dimensional vector $\boldsymbol{\gamma}_k$ $(k = 2, \ldots, K)$ is given by

$$\pi(\boldsymbol{\gamma}_k|\cdot) \quad \propto \quad \prod_{i=1}^{n}\prod_{j=1}^{n_i}\left(\frac{\mathrm{e}^{\boldsymbol{x}'_{ij}\boldsymbol{\gamma}_k+\boldsymbol{v}'_i\boldsymbol{\delta}_k+\psi_{ik}}}{\sum_{h=1}^{K}\mathrm{e}^{\boldsymbol{x}'_{ij}\boldsymbol{\gamma}_h+\boldsymbol{v}'_i\boldsymbol{\delta}_h+\psi_{ih}}}\right)\mathrm{N}_p(\boldsymbol{\gamma}_k;\boldsymbol{\mu}_\gamma,\boldsymbol{\Sigma}_\gamma),$$

where $\mathrm{N}_p(\boldsymbol{\gamma}_k;\cdot)$ is a p-dimensional normal distribution evaluated at $\boldsymbol{\gamma}_k$. Since this full conditional does not have a closed analytic form, we update $\boldsymbol{\gamma}_k$ using a random-walk Metropolis algorithm based on a multivariate-$t_3(s_g\boldsymbol{T}_k)$ proposal density centered at the previous value, $\boldsymbol{\gamma}_k^{\mathrm{old}}$, where the parameter s_g scales the covariance to achieve an optimal acceptance rates, and $\boldsymbol{T}_k$ is a component-specific scale matrix. To improve mixing, we apply the adaptive proposal approach developed by Haario *et al.* (2005), which uses the empirical covariance from an extended burn-in to tune $\boldsymbol{T}_k$. A similar approach can be used to update $\boldsymbol{\delta}_k$.

2) **Update ψ_{ik}**: The full conditional for ψ_{ik} $(i = 1, \ldots, n;\ k = 2, \ldots, K)$ is given by:

$$\pi(\psi_{ik}|\cdot) \quad \propto \quad \prod_{j=1}^{n_i}\left(\frac{\mathrm{e}^{\boldsymbol{x}'_{ij}\boldsymbol{\gamma}_k+\boldsymbol{v}'_i\boldsymbol{\delta}_k+\psi_{ik}}}{\sum_{h=1}^{K}\mathrm{e}^{\boldsymbol{x}'_{ij}\boldsymbol{\gamma}_h+\boldsymbol{v}'_i\boldsymbol{\delta}_h+\psi_{ih}}}\right)\mathrm{N}\left(\frac{1}{m_i}\sum_{l\in\partial_i}\psi_{lk},\tau_k^2/m_i\right),$$

where m_i and ∂_i are defined in equation (4) of Section 3. Since this full conditional does not have a closed form, we update ψ_{ik} using random-walk Metropolis.

(3) **Update τ_k^2**: Assuming an $\mathrm{IG}(g, s)$ prior, draw τ_k^2 its $\mathrm{IG}(g^*, s^*)$ full conditional, where $g^* = g + (n-\omega)/2$, $s^* = s + \boldsymbol{\psi}'_k(\boldsymbol{M}-\boldsymbol{A})\boldsymbol{\psi}_k/2$, n is the number of areal units, $\omega = \max(1, \text{number of "islands"})$, and $(\boldsymbol{M}-\boldsymbol{A})$ is defined in equation (6).

(4) **Update C_{ij}**: For all (i, j), draw C_{ij} from its full conditional

$$\begin{aligned}\pi(C_{ij}|\cdot) &= \Pr(C_{ij}=k|\cdot) = \mathrm{Cat}(p_{ijk}),\ \text{where}\\ p_{ijk} &= \frac{\pi_{ijk}\,\mathrm{N}_2(\boldsymbol{y}_{ij};\boldsymbol{\eta}_{ijk},\boldsymbol{\Sigma}_k)}{\sum_{h=1}^{K}\pi_{ijh}\,\mathrm{N}_2(\boldsymbol{y}_{ij};\boldsymbol{\eta}_{ijh},\boldsymbol{\Sigma}_h)}.\end{aligned}$$

Here, $\mathrm{Cat}(p_{ijk})$ denotes a "categorical" distribution taking the value k with probability p_{ijk}, π_{ijk} is the prior probability that $C_{ij} = k$, as given in equation (1) of Section 3, and $\mathrm{N}_2(\boldsymbol{y}_{ij};\boldsymbol{\eta}_{ijk},\boldsymbol{\Sigma}_k)$ denotes the bivariate normal density from equation (1), Section 3, evaluated at $\boldsymbol{y}_{ij}$.

(5) **Update $\boldsymbol{\Sigma}_k$:** Assuming an $\mathrm{IW}(\kappa_0, \boldsymbol{S}_0)$ prior, update $\boldsymbol{\Sigma}_k$ from its $\mathrm{IW}(\kappa^*, \boldsymbol{S}^*)$ full conditional, where $\kappa^* = \kappa_0 + N_k$, N_k is the number of observations in component k, $\boldsymbol{S}^* = \boldsymbol{S}_0 + \boldsymbol{E}_k'\boldsymbol{E}_k$, $\boldsymbol{E} = \boldsymbol{Y}_k - \boldsymbol{\eta}_k^*$, $\boldsymbol{Y}_k$ is an $N_k \times 2$ matrix consisting of y_{1ij} (first column) and y_{2ij} (second column) response values for all (i,j) in component k, and $\boldsymbol{\eta}_k^*$ is an $N_k \times 2$ matrix consisting of η_{1ij} (first column) and η_{2ij} (second column) linear predictor values for all (i,j) in component k, as defined in equation (1). Use $\boldsymbol{\Sigma}_k = \begin{pmatrix} \sigma_{1k}^2 & \rho_k\sigma_{1k}\sigma_{2k} \\ \rho_k\sigma_{1k}\sigma_{2k} & \sigma_{2k}^2 \end{pmatrix}$ to obtain $\sigma_{1|2,k}^2 = (1-\rho_k^2)\sigma_{1k}^2$, $\sigma_{2|1,k}^2 = (1-\rho_k^2)\sigma_{2k}^2$, $\beta_{1k}^* = \rho_k\sigma_{1k}/\sigma_{2k}$ and $\beta_{2k}^* = \rho_k\sigma_{2k}/\sigma_{1k}$.

(6) **Update $\boldsymbol{\beta}_{1k}$ and $\boldsymbol{\alpha}_{1k}$:** We update $\boldsymbol{\beta}_{1k}$ and $\boldsymbol{\alpha}_{1k}$ conditional on $\boldsymbol{y}_{2k}$, where $\boldsymbol{y}_{2k}$ denotes the $N_k \times 1$ vector of y_{2ij} response values for all (i,j) in component k. Specifically, assuming a $\mathrm{N}_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ prior, update $\boldsymbol{\beta}_{1k}$ from its $\mathrm{N}_p(\boldsymbol{M}_{\beta_{1k}}, \boldsymbol{V}_{\beta_{1k}})$ full conditional, where

$$\begin{aligned} \boldsymbol{V}_{\beta_{1k}} &= \left[\boldsymbol{\Sigma}_\beta^{-1} + \sigma_{1|2,k}^{-2}(\boldsymbol{X}_k'\boldsymbol{X}_k)\right]^{-1} \text{ and} \\ \boldsymbol{M}_{\beta_{1k}} &= \boldsymbol{V}_{\beta_1}\left\{\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta + \sigma_{1|2,k}^{-2}\boldsymbol{X}_k'\left[\boldsymbol{y}_{1k} - \boldsymbol{V}_k\boldsymbol{\alpha}_{1k} - \boldsymbol{\Phi}_{1k} - \beta_{1k}^*(\boldsymbol{y}_{2k} - \boldsymbol{\eta}_{2k})\right]\right\}. \end{aligned}$$

Here, $\boldsymbol{y}_{1k}$ denotes the $N_k \times 1$ vector of y_{1ij} response values for all (i,j) in component k; $\boldsymbol{X}_k$ denotes the $N_k \times p$ individual-level design matrix for component k; $\boldsymbol{V}_k$ is an $N_k \times r$ county-level design matrix for component k; $\boldsymbol{\Phi}_{1k}$ is an $N_k \times 1$ stacked vector such that ϕ_{1ik} is replicated for each observation in county i and component k; $\boldsymbol{\eta}_{2k}$ is an $N_k \times 1$ vector of η_{2ij} values for all (i,j) in component k, as defined in equation (1); and $\sigma_{1|2,k}^2$ and β_{1k}^* are defined in Step (5). A similar set of equations can be used to update the $r \times 1$ vector $\boldsymbol{\alpha}_{1k}$.

(7) **Update $\boldsymbol{\beta}_{2k}$ and $\boldsymbol{\alpha}_{2k}$:** We update $\boldsymbol{\beta}_{2k}$ and $\boldsymbol{\alpha}_{2k}$ conditional on $\boldsymbol{y}_{1k}$. Specifically, assuming a $\mathrm{N}_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ prior, update $\boldsymbol{\beta}_{2k}$ from its $\mathrm{N}_p(\boldsymbol{M}_{\beta_{2k}}, \boldsymbol{V}_{\beta_{2k}})$ full conditional, where

$$\begin{aligned} \boldsymbol{V}_{\beta_{2k}} &= \left[\boldsymbol{\Sigma}_\beta^{-1} + \sigma_{2|1,k}^{-2}(\boldsymbol{X}_k'\boldsymbol{X}_k)\right]^{-1} \text{ and} \\ \boldsymbol{M}_{\beta_{2k}} &= \boldsymbol{V}_{\beta_1}\left\{\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta + \sigma_{2|1,k}^{-2}\boldsymbol{X}_k'\left[\boldsymbol{y}_{2k} - \boldsymbol{V}_k\boldsymbol{\alpha}_{2k} - \boldsymbol{\Phi}_{2k} - \beta_{2k}^*(\boldsymbol{y}_{1k} - \boldsymbol{\eta}_{1k})\right]\right\}. \end{aligned}$$

Here, $\boldsymbol{\eta}_{1k}$ is an $N_k \times 1$ vector of η_{1ij} values for all (i,j) in component k, as defined in equation (1), $\sigma_{2|1,k}^2$ and β_{2k}^* are defined in Step (5), and all other elements are defined in a manner analogous to those in Step (6). A similar set of equations can be used to update the $r \times 1$ vector $\boldsymbol{\alpha}_{2k}$.

(8) **Update $\boldsymbol{\phi}_{ik}$:** For $i = 1, \ldots, n$, draw the 2×1 vector $\boldsymbol{\phi}_{ik}$ from its $\mathrm{N}_2(\boldsymbol{M}_\phi, \boldsymbol{V}_\phi)$ full conditional, where

$$\begin{aligned} \boldsymbol{V}_\phi &= \left(N_{ik}\boldsymbol{\Sigma}_k^{-1} + m_i\boldsymbol{\Lambda}_k^{-1}\right)^{-1} \\ \boldsymbol{M}_\phi &= \boldsymbol{V}_\phi\left[\boldsymbol{\Sigma}_k^{-1}\boldsymbol{Z}_{ik}'(\boldsymbol{y}_{ik} - \boldsymbol{X}_{ik}\boldsymbol{\beta}_k - \boldsymbol{V}_{ik}\boldsymbol{\alpha}_k) + \boldsymbol{\Lambda}_k^{-1}\sum_{l\in\partial}\boldsymbol{\phi}_{lk}\right]. \end{aligned}$$

Here, N_{ik} denotes the component-k sample size for county i; m_i is the number of neighbors of county i; $\boldsymbol{Z}_{ik}$ is a $2N_{ik} \times 2$ matrix with alternating rows of $(1,0)$ and $(0,1)$; $\boldsymbol{y}_{ik}$ is a $2N_{ik} \times 1$ stacked vector consisting of

alternating Y_1 and Y_2 response values for each observation in county i and component k; $\boldsymbol{X}_{ik}$ is a $2N_{ik} \times 2p$ design matrix for county i and component k, with rows alternating between $(\boldsymbol{x}'_{ij}, \mathbf{0})$ and $(\mathbf{0}, \boldsymbol{x}'_{ij})$ for each of the N_{ik} observations in county i and component k; $\boldsymbol{\beta}_k = (\boldsymbol{\beta}'_{1k}, \boldsymbol{\beta}'_{2k})'$ is a $2p \times 1$ vector of individual-level regression parameters; $\boldsymbol{V}_{ik}$ is a $2N_{ik} \times 2r$ design matrix of county-level predictors; and $\boldsymbol{\alpha}_k = (\boldsymbol{\alpha}'_{1k}, \boldsymbol{\alpha}'_{2k})'$ is a corresponding $2r \times 1$ vector of county-level regression coefficients.

(9) **Update $\boldsymbol{\Lambda}_k$:**
Assuming an IW$(\nu_0, \boldsymbol{D}_0)$ prior, draw $\boldsymbol{\Lambda}_k$ from its IW$(\nu^*, \boldsymbol{D}^*)$ full conditional, where $\nu^* = \nu + n - \omega$, n is the number of areal units, ω is defined in Step (3) above, and $\boldsymbol{D}^* = \boldsymbol{D}_0 + \boldsymbol{\Phi}_k^{*'}(\boldsymbol{M} - \boldsymbol{A})\boldsymbol{\Phi}_k^*$. Here, $\boldsymbol{\Phi}_k^*$ denotes an $n \times 2$ matrix with first column equal to $\boldsymbol{\phi}_{1k} = (\phi_{11k}, \ldots, \phi_{1nk})'$ and second column equal to $\boldsymbol{\phi}_{2k} = (\phi_{21k}, \ldots, \phi_{2nk})'$.

References

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

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 36*(2), 192–236.

Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics 43*(1), 1–20.

Carlin, B. P. and S. Banerjee (2002). Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). *Bayesian Statistics 7*, 45–63.

Celeux, G., F. Forbes, C. P. Robert, and D. M. Titterington (2006). Deviance information criteria for missing data models. *Bayesian Analysis 1*(4), 651–674.

Congdon, P. (2010). Random-effects models for migration attractivity and retentivity: a Bayesian methodology. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 173*(4), 755–774.

Ferguson, R. (2008). *Toward Excellence With Equity: An Emerging Vision for Closing the Achievement Gap.* Cambridge, MA: Harvard Education Press.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models.* Berlin: Springer-Verlag.

Gelfand, A. E., A. Kottas, and S. N. MacEachern (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association 100*(471), 1021–1035.

Gelfand, A. E. and P. Vounatsou (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics 4*(1), 11–15.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis* (2 ed.). Boca Raton: Chapman & Hall/CRC.

Genton, M. and H. Zhang (2012). Identifiability problems in some non-Gaussian spatial random fields. *Chilean Journal of Statistics 3*(2), 171–179.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*(4), 711–732.

Green, P. J. and S. Richardson (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Association 97*(460), 1055–1070.

Haario, H., E. Saksman, and J. Tamminen (2005). Componentwise adaptation for high dimensional MCMC. *Computational Statistics 20*, 265–273.

Ismail, S., W. Sun, F. S. Nathoo, A. Babul, A. Moiseev, M. F. Beg, and N. Virji-Babul (2013). A skew-t space-varying regression model for the spectral analysis of resting state brain activity. *Statistical Methods in Medical Research 22*(4), 424–438.

Ji, C., D. Merl, T. B. Kepler, and M. West (2009). Spatial mixture modelling for unobserved point processes: Examples in immunofluorescence histology. *Bayesian Analysis 4*(2), 297–316.

Jin, X., B. P. Carlin, and S. Banerjee (2005). Generalized hierarchical multivariate car models for areal data. *Biometrics 61*(4), 950–961.

Kottas, A., J. A. Duan, and A. E. Gelfand (2008). Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures. *Biometrical Journal 50*(1), 29–42.

Kottas, A. and B. Sansó (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference 137*, 3151–3163.

Lawson, A. B. and A. Clark (2002). Spatial mixture relative risk models applied to disease mapping. *Statistics in Medicine 21*(3), 359–370.

Mardia, K. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis 24*(2), 265–284.

Mclachlan, G. and D. Peel (2000). *Finite Mixture Models*. New York: John Wiley & Sons.

Nathoo, F. S. and P. Ghosh (2013). Skew-elliptical spatial random effect modeling for areal data with application to mapping health utilization rates. *Statistics in Medicine 32*(2), 290–306.

National Center for Education Statistics (2013). The nation's report card: Trends in academic progress 2012. *NCES Report* 2013-456, Institute of Education Sciences, U.S. Department of Education, Washington, D.C. http://nces.ed.gov/nationsreportcard/pubs/main2012/2013456.aspx.

No Child Left Behind Act (2002). Pub. l. no. 107-110, 115 stat. 1425. http://www2.ed.gov/policy/elsec/leg/esea02/index.html.

North Carolina Department of Public Instruction (2006). The North Carolina testing program 2006–2007. http://www.ncpublicschools.org/docs/accountability/NORTHCgeneralpolicies.pdf.

North Carolina Department of Public Instruction (2007). Achievement level ranges for the North Carolina end-of-grade tests: Mathematics at grades 3–8. http://www.ncpublicschools.org/docs/accountability/testing/eog/rangeseogmath.pdf.

North Carolina Department of Public Instruction (2008). Achievement level ranges for the North Carolina end-of-grade tests: Reading comprehension at grades 3–8. http://www.ncpublicschools.org/docs/accountability/testing/achievelevels/alrangesreading.pdf.

Pope, D. G. and J. R. Sydnor (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives 24*(2), 95–108.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Reich, B. J. and M. Fuentes (2007). A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Annals of Applied Statistics 1*(1), 249–264.

Richardson, S. (2002). Discussion of Spiegelhalter et al. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 626–627.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica 4*, 639–650.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 583–639.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(4), 795–809.

U.S. Census Bureau (2010). American Community Survey 2005–2009. http://www.census.gov/acs/www/.

Wall, M. M. and X. Liu (2009). Spatial latent class analysis model for spatially distributed multivariate binary data. *Computational Statistics & Data Analysis 53*(8), 3057–3069.

Zareifard, H. and M. J. Khaledi (2013). Non-Gaussian modeling of spatial data using scale mixing of a unified skew Gaussian process. *Journal of Multivariate Analysis 114* (0), 16–28.

Zhang, Y., J. S. Hodges, and S. Banerjee (2009). Smoothed ANOVA with spatial effects as a competitor to MCAR in multivariate spatial smoothing. *Annals of Applied Statistics 3*, 1805–1830.