NISS

A Spatially Correlated Hierarchical Random Effects Model for Ohio Corn Yield

Nancy Jo McMillan and L. Mark Berliner

Technical Report Number 10
January, 1994

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

A Spatially Correlated Hierarchical Random Effects Model for Ohio Corn Yield

NANCY JO MCMILLAN and L. MARK BERLINER *

January 17, 1994

Abstract

Estimation of total county crop yields is of interest to both federal and state governments. This article focuses on estimation of corn production in the counties of Ohio; Our primary goal is producing accurate estimates of corn yield per acre stratified by county and by size of farm. These estimates can be used in conjunction with census data to produce the desired total county corn yield estimates. Our data is responses of 3,842 farms to a voluntary survey. A Bayesian hierarchical random effects model is proposed. The key idea in this formulation involves the input of prior information based on anticipated spatial dependence of corn production between neighboring counties. The model suggested is sufficiently complex to prohibit simple computations. Hence, we employ Gibbs Sampling. The resulting estimates of yield per acre show a strong spatial trend. Lower productivity in the Appalachian foothills gradually increases to higher productivity in the central-northwest region. The geography of Ohio suggests this effect is reasonable. We also estimate the posterior covariance structure of the random effects, including the spatial county effect. This is a large covariance matrix and, thus, difficult to examine carefully. Our approach to investigation of this matrix is graphical examination of one row or "slice" at a time. The "slices" examined display a desirable spatial property; Neighboring counties are generally more correlated than distant counties. Our methodology is easily adaptable to other crops and states.

KEY WORDS: Bayesian analysis; Gibbs sampling; Markov random field.

1 Introduction

The United States Department of Agriculture (USDA), in cooperation with state agricultural departments, is interested in estimating the corn yield in the state of Ohio on a county basis. They have available for this project data of two types, 1) responses (to a voluntary survey)

*Nancy McMillan is Post-Doctoral Fellow, National Institute of Statistical Sciences, P. O. Box 14162, Research Triangle Park, NC 27709-4162. Mark Berliner is Associate Professor, Department od Statistics, Ohio State University, 1958 Neil Ave., Columbus, OH 43210-1247. The authors are grateful to Prem Goel and Elizabeth Stasny for useful discussions. This research was supported in part by the United States Department of Agriculture under Cooperative Agreement No. 58-3AEU-9-80040.

of 3,842 farms in the state of Ohio which reported planting corn and 2) farm census data collected every five years reporting total farm acreages and acreages planted in particular crops.

The data for the farms responding to the voluntary survey includes an indication of the county in which the farm is located, total acres planted, acres planted in corn, acres of corn harvested, and yield in bushels per acre harvested. The census data available includes the total number of farms and total acreage of farmland for each county stratified by size of farms. There are twelve strata of farm sizes in the census data. For simplicity the twelve categories of farm size available in the census data will be condensed to three categories. The size groups were chosen to make the number of farms in each group roughly equivalent. (This implies that the total acreage in farmland in each group is not similar.) Group 1 contains farms with 0-179 acres, Group 2, 180-499 acres, and Group 3, 500+ acres.

Inspite of the fact that the voluntary survey data's size covariate was available as a quantitative variable, it was decided that size should be incorporated as a categorical effect in modeling the voluntary data. The disadvantage of possibly introducing non-existent regression effects into the model, by categorizing a quantitative variable, was outweighed by the advantage of compatibility with the census data which only reported size information as a categorical effect.

Figure 1 displays the yield per acre harvested averages for each county, group combination. (Note that counties which are not shaded have no data collected for the size group being represented.) Notice the tendency of neighboring counties to have similar yield per acre harvested averages. This is indicated by contiguous regions, larger than county size, having the same color. The data also indicates that there is an overall tendency of the yield per acre harvested to decrease from the northwest corner of Ohio to the southeast. Looking at all three size groups suggests that overall, the Group 1 averages are smaller than the Group 2 averages which in turn are smaller than the Group 3 averages. The number in each county is the sample size collected in that particular county, size strata.

The procedure for producing total county yield estimates will be as follows: 1) estimate from the voluntary survey data the yield per acre harvested for each county, size pair, call this $\hat{Y}_{i,j}$; 2) estimate (also from the voluntary survey data) the proportion of acres of corn harvested to total acres of farmland for each county, size pair, call this $\hat{P}_{i,j}$; 3) use the census information available to determine the total acreage, say $A_{i,j}$, of farmland in each county, size pair. The yield per county can then be estimated as $\hat{Y}_i = \sum_{j=1}^{3} \hat{Y}_{i,j}\hat{P}_{i,j}A_{i,j}$.

The main portion of this article is devoted to the development of the $\hat{Y}_{i,j}$. Section 2 presents a detailed formulation of a Bayesian model for the $Y_{i,j}$. The key idea in this formulation involves the input of prior information based on anticipated spatial dependence of corn production between neighboring counties. The model suggested is sufficiently complex to prohibit simple computations. Hence, the popular notions of Gibbs Sampling are employed in the analysis; The procedure used is described in Section 3. In Section 4 the results are used to produce the desired estimates of corn production on a county basis.

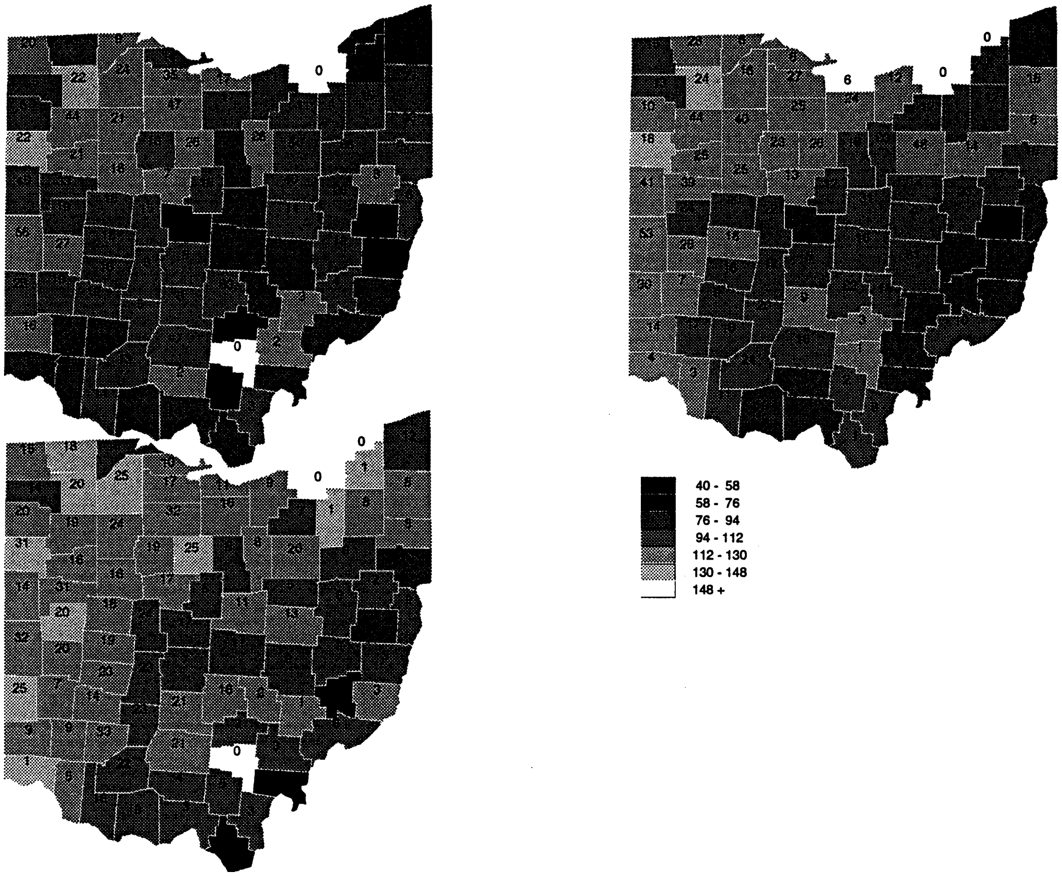

Figure 1: County maps of Ohio are shaded to reflect average yield per acre of the data collected from the voluntary survey in each county in each of the three size groups. The map on the upper left is Group 1; the upper right is Group 2; the lower is Group 3. The numbers in each county represent the sample size for that particular county, size pair.

2 Model Development

The first step in producing the desired estimates of corn yield per county is to propose a model for the yield per acre harvested. A Bayesian hierarchical random effects model is proposed. The effects will be a mean effect, μ, a county effect, $\mathbf{C}$, and a group effect, $\mathbf{G}$. Priors will be places on all three effects.

Model 2.1 Corn Yield per Acre Harvested Model
We consider the following linear model for the observed bushels of corn produced per acre of land harvested for each farm within a county, size pair:

$$Y_{i,j,k} = \mu + C_i + G_j + e_{i,j,k}, \tag{1}$$

where

$$e_{i,j,k} \sim N(0, \sigma^2) \tag{2}$$

and $i = 1, ..., N_C$, $j = 1, 2, 3$, $k = 1, ..., n_{i,j}$ *and the* $e_{i,j,k}$ *are assumed i.i.d.*

Vector notation for this model will be desirable. Let $\mathbf{Y_{i,j}} = (Y_{i,j,1}, Y_{i,j,2}, ..., Y_{i,j,n_{i,j}})^T$ and $\mathbf{Y} = (\mathbf{Y_{1,1}}^T, \mathbf{Y_{1,2}}^T, \mathbf{Y_{1,3}}^T, \mathbf{Y_{2,1}}^T, ..., \mathbf{Y_{N_C,3}}^T)^T$. Let $\beta = (\mu, \mathbf{C}^T, \mathbf{G}^T)^T$. The design matrix

X can be written as a partitioned matrix, $X = (X_1 : X_2 : X_3)$, to provide easy manipulation of the posterior distribution. Using this notation the linear model can be written as:

$$\begin{aligned} \mathbf{Y} &= X\beta + \mathbf{e} \\ \mathbf{Y} &= X_1\mu + X_2\mathbf{C} + X_3\mathbf{G} + \mathbf{e}. \end{aligned} \tag{3}$$

A key feature of our analysis involves the incorporation of prior information about the county effects. (Other prior specifications used here are more standard. A discussion of the specifications is given at the end of the section.) The county effects are meant to account for variability introduced into corn yield per acre harvested due to local growing conditions such as terrain, soil fertility, and rainfall. The county borders in Ohio were drawn arbitrarily and do not reflect the division of Ohio into areas which are internally homogeneous, but independent of each other as regards local growing conditions. It is believed that "neighboring" counties should have similar growing conditions. We therefore propose prior model specifications which incorporate spatial dependence structures.

Figure 2: County map of Ohio used to create neighborhood structure.

Let $S = \{1, ..., N_C\}$ be an arbitrary labelling of the counties of Ohio. We will define a neighborhood structure $\mathcal{N} = \{\eta_i | i \in S\}$ for the counties which mirrors their physical location. Specifically, counties i and j are neighbors if they share a common border. (County borders that are extremely short such as that between Darke and Auglaize counties in Western Ohio are not considered to be neighbors. See Figure 2. The corn yield per acre harvested estimates of interest were also produced under the model considering counties with short borders as neighbors. This produced insignificant changes in the final results.)

The prior distributional assumptions for the model components of Model 2.1 will be as follows:

Model 2.2 First Stage Prior for Corn Yield per Acre Harvested Model *Given hyperparameters $\sigma_C^2, \Sigma,$ and σ_G^2 and assuming independence of μ, $\mathbf{C}$, and $\mathbf{G}$ given the hyperparameters the first stage of the prior is:*

$$\begin{aligned} \mu &\sim N(\mu_0, \sigma_0^2) \\ \mathbf{C} &\sim N(\mathbf{0}, \sigma_C^2 \Sigma) \\ \mathbf{G} &\sim N(\mathbf{0}, \sigma_G^2 I) \\ \pi(\sigma^2) &= \frac{1}{\sigma^2}, \end{aligned} \tag{4}$$

where we set $\mu_0 = 100$ and $\sigma_0^2 = 100$.

These assumptions imply that the distribution of β given the hyperparameters is normal with mean $\mathbf{b}$, where $\mathbf{b}$ is an $(N_C + 4) \times 1$ vector whose first entry is μ_0 and remaining entries are all 0, and the following partitioned, block diagonal covariance matrix, denoted by A,

$$A = \begin{bmatrix} \sigma_0^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_C^2 \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_G^2 I \end{bmatrix}. \tag{5}$$

2.1 Modeling spatial dependence

The covariance matrix Σ will introduce the desired spatial effect into the model. Define $\Delta = ((\delta_{i,j}))$ to be the $N_C \times N_C$ symmetric matrix determined by the "shared boundary" neighborhood structure, where:

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 1 & \text{if } i \in \eta_j \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

There are at least two ways that the neighborhood matrix, Δ, can be used to specify Σ. For example, by defining

$$\Sigma = (1 - \rho) I + \rho \Delta, \tag{7}$$

special implications follow concerning the *marginal* independence of counties which are not neighbors. In this article we seek specifications motivated by the notion of a Markov random field. These models are popular in imaging and spatial statistics problems, primarily because of their ability to build in varieties of spatial dependence structures. (See for example [5], [4], or [1] for discussions.) Throughout this paper we will make use of the increasingly popular notation; Specifically, for random variables or vectors U and V, let $[U|V]$ denote the conditional distribution of U given V. The essential feature of a Markov random field model involves specifications of the *conditional* distributions of county effects as

$$[C_i | C_j, j \neq i] = [C_i | C_j, j \in \eta_i] \tag{8}$$

for arbitrary neighborhoods. In this paper we only consider the aforementioned nearest neighbor structure, though more complex structures can be analyzed by similar methods. Note that in general (7) is incompatiable with (8).

To construct Σ compatiable with (8), consider the associated precision matrix Σ^{-1}. Specifically, represent Σ^{-1} as

$$\Sigma^{-1} = (1-a)I + a\Delta. \tag{9}$$

(Note that the parameter,a, cannot be viewed as a correlation coefficient.) We now demonstrate that this leads to the desired conditional independence model in (8). The full conditional distribution of C_i is proportional to the joint distribution of $\mathbf{C}$,

$$[C_i|C_j, j \neq i] \propto [\mathbf{C}]. \tag{10}$$

By definition,

$$[\mathbf{C}] \propto (\sigma_C^2)^{-\frac{N}{2}} |\Sigma^{-1}|^{\frac{1}{2}} \exp\{-\frac{1}{2\sigma_C^2} C^T \Sigma^{-1} C\} \tag{11}$$

which, together with (9), imply that the full conditional distribution of C_i is

$$[C_i|C_j, j \neq i] \propto \exp\{-\frac{1}{2\sigma_C^2}[(1-a)C^T I C + aC^T \Delta C]\}. \tag{12}$$

Expressing the matrix multiplications in (12) as sums leads to

$$[C_i|C_j, j \neq i] \propto \exp\{-\frac{1}{2\sigma_C^2}[(1-a)\sum_{k=1}^{N_C} C_k^2 + a\sum_{k=1}^{N_C}\sum_{j=1}^{N_C} C_k \delta_{k,j} C_j]\}. \tag{13}$$

The definition of $\delta_{k,j}$ reduces the range of the double sum yielding

$$[C_i|C_j, j \neq i] \propto \exp\{-\frac{1}{2\sigma_C^2}[\sum_{k=1}^{N_C} C_k^2 + a\sum_{k=1}^{N_C}\sum_{j\in\eta_k} C_k C_j]\}. \tag{14}$$

Removing the unnecessary pieces from the above proportionality statement reduces the kernel of the full conditional to

$$[C_i|C_j, j \neq i] \propto \exp\{-\frac{1}{2\sigma_C^2}[C_i^2 + 2aC_i \sum_{j\in\eta_i} C_j]\} \tag{15}$$

which can be recognized as

$$[C_i|C_j, j \neq i] = N(-a\sum_{j\in\eta_i} C_j, \sigma_C^2). \tag{16}$$

There is a difficulty inherent in this specification of the distribution of $\mathbf{C}$. The mean of the full conditional for a particular C_i is based on the total of the neighbors, $\sum_{j\in\eta_i} C_j$, rather than the mean, $\frac{1}{|\eta_i|}\sum_{j\in\eta_i} C_j$. Since all counties don't have the same number of neighbors, the parameter a, representing the strength of the dependence on neighbors, cannot account for this. Development of a similar Markov random field model which accounts for this problem has proven to be difficult. We proceed with the model as specified.

In the specification of Σ, we will require Σ to be positive definite. The resulting constraints are considered next. Let $\lambda_l, l = 1, \ldots, N_C$ be the eigenvalues of the neighborhood matrix Δ. Each eigenvalue of Σ^{-1}, say η_l, can be expressed as a function of an eigenvalue of Δ. By definition, each eigenvalue of Σ^{-1} is a solution to the equation obtained by setting the characteristic function of Σ^{-1} equal to zero,

$$|\Sigma^{-1} - \eta I| = 0. \tag{17}$$

Expansion based on the definition of Σ^{-1} leads to,

$$|(1-a)I + a\Delta - \eta I| = 0. \tag{18}$$

Manipulating (18) yields a similar characteristic function equation for neighborhood matrix Δ,

$$\begin{aligned} a|\Delta - (\frac{\eta}{a} + 1 - \frac{1}{a})I| &= 0 \\ |\Delta - \lambda I| &= 0. \end{aligned} \tag{19}$$

Provided $a \neq 0$, in which case the analysis is trivial, the N_C solutions to (19) can be matched, even accounting for multiplicity if necesary, to yield

$$\begin{aligned} \frac{\eta_l}{a} + 1 - \frac{1}{a} &= \lambda_l \\ \eta_l &= 1 - a + a\lambda_l. \end{aligned} \tag{20}$$

In general, we know little about the eigenvalues of Δ beyond the facts that they are real and sum to N_C. This implies that at least one eigenvalue lies on each side of 1. These facts, combined with a simple algebraic analysis of (20), yield the final constraint on a insuring that each η_l is positive:

$$a^- = (1 - \max_l \lambda_l)^{-1} < a < (1 - \min_l \lambda_l)^{-1} = a^+. \tag{21}$$

Our neighborhood matrix for the counties of Ohio resulted in $a^- = -0.1817$ and $a^+ = 0.3431$.

A third stage prior is still needed to complete the hierarchical model. Some discussion of the role of a in this model is needed first.

2.2 Normalizing constants

A difficulty which frequently arises in complex hierarchical Bayesian models is that evaluation of normalizing constants becomes necessary. This difficulty is discussed in [6]. In our model, a normalizing constant makes sampling, as needed in our implementation of the Gibbs Sampler, difficult. The posterior density for Model 2.1 provided the first stage prior is specified by Model 2.2 is:

$$\begin{aligned} \pi(\mu, \mathbf{C}, \mathbf{G}, \sigma^2 | \mathbf{Y}, \sigma_C^2, a, \sigma_G^2) \propto\ & (\sigma^2)^{-1}(\sigma^2)^{-\frac{n}{2}}(\sigma_C^2)^{-\frac{N_C}{2}}|\Sigma^{-1}|^{\frac{1}{2}}(\sigma_G^2)^{-\frac{N_G}{2}} \\ & \exp\{-\frac{1}{2\sigma^2}(\mathbf{Y} - X\beta)^T(\mathbf{Y} - X\beta) \\ & -\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2 - \frac{1}{2\sigma_C^2}\mathbf{C}^T\Sigma^{-1}\mathbf{C} - \frac{1}{2\sigma_G^2}\mathbf{G}^T\mathbf{G}\}. \end{aligned} \tag{22}$$

The function $Z(a) =^{def} |\Sigma^{-1}|$ arises from the normalizing constant of $[\mathbf{C}|\sigma_C^2, a]$. In a non-hierarchical model this could be ignored. It cannot be ignored in our analysis as we intend to introduce a second stage prior on a. The definition of Σ^{-1} leads immediately to a formula for $Z(a)$:

$$Z(a) = |\Sigma^{-1}| = \prod_{l=1}^{N_C} (\lambda_l^1) = \prod_{l=1}^{N_C} (1 - a + a\lambda_l). \tag{23}$$

This function appears in both the joint posterior and the full conditional for a. Since we will implement the Gibbs Sampler, sampling the full conditional distribution of a is a necessity. This step will be handled via the *Griddy Gibbs* approach of [7]. Specifically, $Z(a)$ will be evaluated on a grid of values and a's will be drawn from the resulting discrete approximation to the correct distribution. Figure 3 displays graphs of a vs $Z(a)$ and a vs $\log Z(a)$, demonstrating the effect of $Z(a)$ on the posterior.

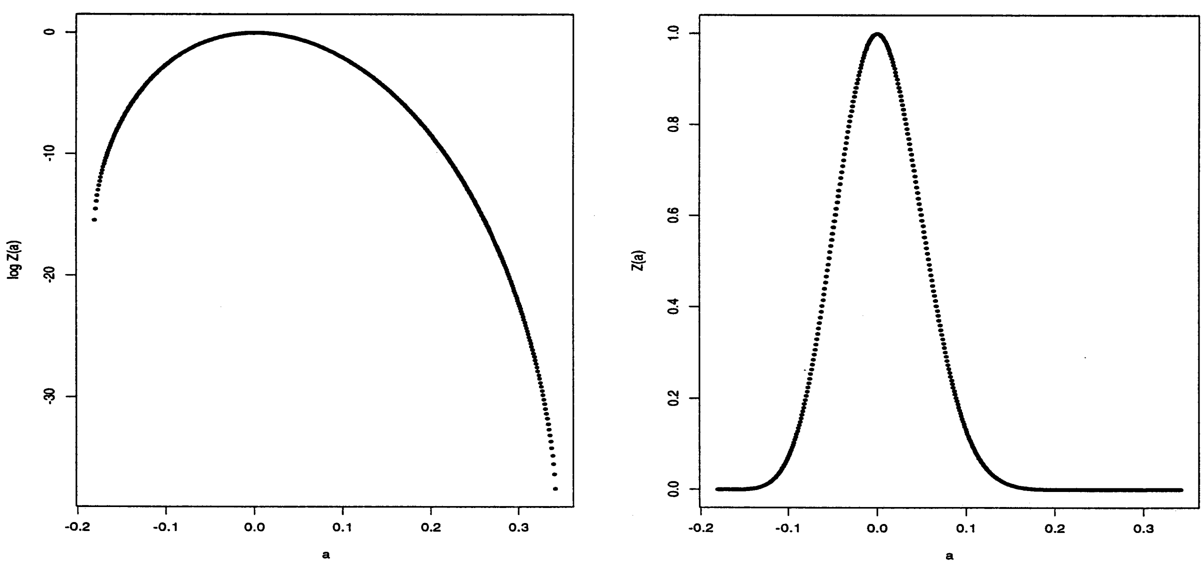

Figure 3: The left graph displays the behavior of $\log Z(a)$ as a function of a. The right shows how this affects the posterior distribution of a.

Model 2.3 Second Stage Prior for Corn Yield per Acre Harvested Model
Priors will be taken as:

$$\begin{aligned} \sigma_G^2 &\sim \text{InverseGamma}(\alpha_G, \beta_G) \\ \pi(\sigma_C^2) &= 1, \sigma_C^2 > 0 \\ a &\sim \text{Uniform}(a^-, a^+), \end{aligned} \tag{24}$$

where we set $\alpha_G = 2$ *and* $\beta_G = 0.02$.

We conclude this section with comments motivating some of the prior specifications made. First, an uninformed, though proper, uniform prior for a seems natural, though a more complete specification of prior information might be given if a is restricted to be less than or equal to 0. We did not use this extra condition. Our hope was that by not forcing a to be negative, the extra flexibility would allow the data to convincingly verify our prior beliefs. We believe this hope was realized; the results described below indicate that the posterior for a assigns very little probability to positive a. Next, the natural invariant noninformative prior $1/\sigma^2$ is used for σ^2. However, we employ a uniform, improper prior for σ_C^2. This choice was made to insure the existence of a proper posterior. This aspect is more completely discussed in the Appendix. The discussion is of general interest beyond the scope of this paper. Finally, proper, though not very sharp, priors are used for both μ and σ_G^2. It seems quite plausible that an expert in Ohio corn production would have useful prior beliefs concerning both "average" corn yields and the "average" effects of farm size.

3 Analysis via Gibbs Sampling

Estimation in this model is accomplished via Gibbs sampling. See [3], [2], or [8] for a review of Gibbs sampling. Our primary goal involves estimation of the posterior expected yield per acre harvested in each county and group. To obtain these estimates based on (3), an estimate of the posterior expected value of β is needed. An estimate of the covariance matrix of β is essential to estimating the associated variances.

Let $\theta = (\beta, \sigma^2, \sigma_C^2, a, \sigma_G^2)^T$. In order to run the Gibbs sampler, the full conditionals for each of the elements of θ are needed. For ease of notation define $\theta_{-\beta} = (\sigma^2, \sigma_C^2, a, \sigma_G^2)^T$ and make similar definitions for the other parameters. The full conditionals can be determined by direct calculation and are:

$$\beta|\mathbf{Y}, \theta_{-\beta} \sim \mathrm{N}(D^{-1}(X^T\mathbf{Y} + \sigma^2 A^{-1}\mathbf{b}), \sigma^2 D^{-1}) \tag{25}$$

where

$$D = (X^T X + \sigma^2 A^{-1}), \tag{26}$$

$$\begin{aligned} \sigma^2|\mathbf{Y}, \theta_{-\sigma^2} &\sim \text{InverseGamma}(\frac{n_{..}}{2}, \frac{2}{(\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta)}), \\ \sigma_C|\mathbf{Y}, \theta_{-\sigma_C^2} &\sim \text{InverseGamma}(\frac{N_C}{2} - 1, \frac{2}{\mathbf{c}^T\Sigma^{-1}\mathbf{c}}), \\ \sigma_G|\mathbf{Y}, \theta_{-\sigma_G^2} &\sim \text{InverseGamma}(\frac{N_G}{2} + \alpha_G, (\frac{\mathbf{g}^T\mathbf{g}}{2} + \frac{1}{\beta_G})^{-1}), \end{aligned} \tag{27}$$

and

$$a|\mathbf{Y}, \theta_{-a} \propto \exp\{-\frac{a}{2\sigma_c^2}[\mathbf{c}^T(\Delta - I)\mathbf{c}] + \frac{1}{2}\sum_{l=1}^{N_C}\log(1 - a + a\lambda_l)\}. \tag{28}$$

We performed an initial investigation of the posterior distribution using multiple, shorter runs of the Gibbs sampler. These results will not be reported here. Estimates which are reported in this work are based on one Gibbs run of 32,000 iterations, the first 10,000 of which are thrown away as a burn-in phase. The first and second moments of each parameter of the posterior are estimated as ergodic averages of the iterations not in the burn-in phase. Mean and variance estimates for the parameters are based on these moment estimates. We report our posterior mean and variance estimates in Figure 4.

	μ	G_1	G_2	G_3
Expected Value	106.774	-5.789	-0.065	7.256
Standard Deviation	4.102	2.622	2.640	2.670

	a	σ^2	σ_C^2	σ_G^2
Expected Value	-0.119	630.027	413.225	21.507
Standard Deviation	0.041	14.544	362.986	21.156

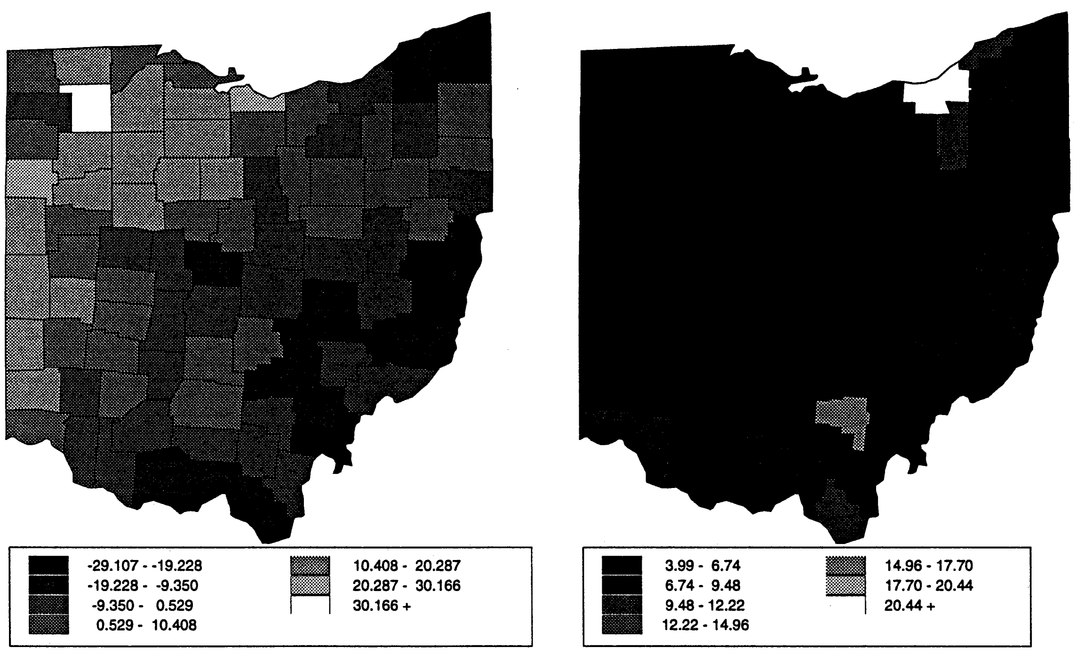

Figure 4: Estimates of $E(\beta)$ and standard deviation of β. The left image of Ohio displays the estimates of county means. The right image of Ohio displays the standard deviation estimate for the county effect.

We adopt a Rao-Blackwellized estimator for inference regarding the marginal posterior covariance matrix of β. Recall from (25) that the full conditional form for the covariance matrix of β is $\sigma^2 D^{-1}$. Based on this, we estimate the posterior covariance matrix of β with $\hat{M}$, where $\hat{M}$ is the ergodic average of $\sigma^2 D^{-1}$ evaluated at each iteration,

$$\hat{M} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 D_i^{-1}. \tag{29}$$

A fortuitous side effect of this estimator of the covariance of $\beta|\mathbf{Y}$ arises from the multivariate normal structure of the full conditional of β; Appropriate blocks of $\hat{M}$ serve as Rao-Blackwellized estimators of the covariance matrices of $\mu|\mathbf{Y}$, $\mathbf{C}|\mathbf{Y}$, and $\mathbf{G}|\mathbf{Y}$. These secondary estimators are based on their respective posterior distributions given the hyperparameters, a, σ_C^2, and σ_G^2, and σ^2.

$\hat{M}$ is an $(1 + N_C + N_G) \times (1 + N_C + N_G)$ matrix, and thus, is difficult to examine carefully. We chose to focus our attention of the $N_C \times N_C$ block of the estimated covariance matrix which corresponds to the county effect. Figure 5 graphically displays three rows of our estimate of the covariance matrix of $\mathbf{C}|\mathbf{Y}$. These rows contain the estimated posterior covariances between Cuyahoga, Lake, and Vinton counties and all the other counties. Cuyahoga, Lake, and Vinton counties were chosen due to their small sample sizes. We believe counties with small sample sizes will benefit most from the dependence structure introduced in the prior of the county effect which allows local pooling of information. Examination of Figure 5 reveals a desirable covariance structure. Spatially local counties appear to have larger covariances than distant counties.

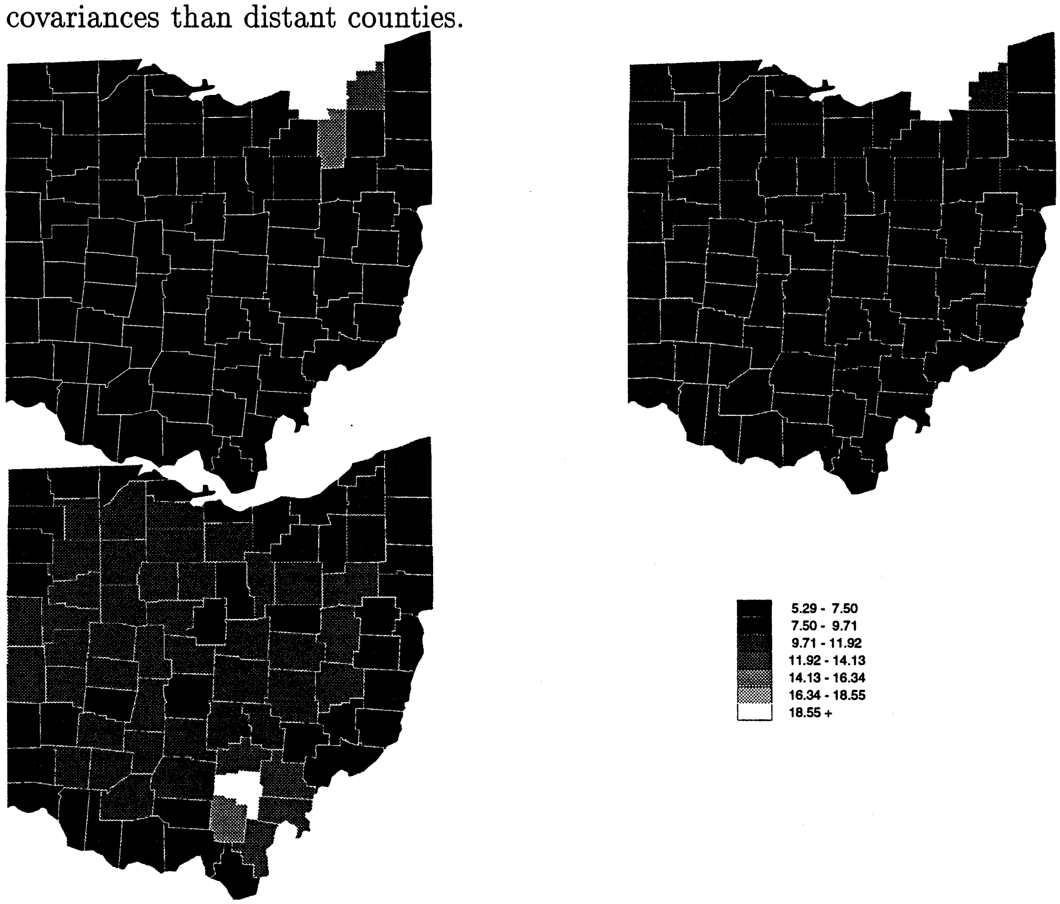

Figure 5: Posterior covariance maps for Cuyahoga, Lake, and Vinton counties.

As noted previously, estimates of the yield per acre harvested for each county, group pair are based on Model 2.1. Stated explicitly,

$$\hat{Y}_{i,j} = \hat{\mu} + \hat{C}_i + \hat{G}_j \tag{30}$$

for $i = 1, \cdots, N_C$ and $j = 1, \cdots, N_G$. For notational convenience, we represent all of these estimates by $\hat{Y}^{CG} = (\hat{Y}_{1,1}, \cdots, \hat{Y}_{1,N_G}, \cdots, \hat{Y}_{N_C,1}, \cdots, \hat{Y}_{N_C,N_G})$. Defining X_Y to be the

$(N_C N_G) \times (1 + N_C + N_G)$ matrix as dictated by the above relation this allows us to write

$$\hat{Y}^{CG} = X_Y \hat{\beta}. \tag{31}$$

We estimate the covariance matrix of $\hat{Y}^{CG}$ by $X_Y \hat{M} X_Y^T$. Obtaining the final estimates of the county yield requires some further work.

4 Results

Computation of estimates of total corn yield per county requires calculation of the intermediary quantities mentioned earlier, namely $\hat{P}_{i,j}$, an estimate of the proportion of acres harvested to total acres of farmland, and $A_{i,j}$, the total acres of farmland for each county, size group as reported by the census data. Estimation of $\hat{P}_{i,j}$ is performed using the voluntary sample data. Let $P_{i,j,k}$ equal the ratio of acres harvested to total acres of farmland for the kth responding farm in the ith county, jth size group. We define $\hat{P}_{i,j}$ to be

$$\hat{P}_{i,j} = \frac{1}{n_{i,j}} \sum_{k=1}^{n_{i,j}} P_{i,j,k} \tag{32}$$

for all $i = 1, \cdots, N_C$ and $j = 1, 2, 3$.

There were four county, size groups in which no data was observed. For the purposes of obtaining our initial estimates of the total yield, we used mean imputation across the size group to impute the missing values. To do this we computed the average, $\bar{P}_j$ for each $j = 1, 2, 3$ without the missing counties, then let missing $\hat{P}_{i,j}$ equal the average from the appropriate size group. Figure 6 displays the resulting estimates of $\hat{P}_{i,j}$.

As noted, the census data total acres in farmland was reported in size by county groups. There are twelve size groups in the census data which were condensed to three for are analysis. The total acres in farmland for some county size groups was not reported due to confidentiality concerns. Some missing values in the census data had to be imputed. We imputed values at the twelve size group level. When k of the size groups were not reported, we imputed the smallest $k-1$ values by taking the median of the size range and multiplying by the reported number of farms in that size group. We then imputed the remaining size group by subtracting the total acreage already allocated from the county total. The resulting exact totals and imputed values are displayed in Figure 7.

As stated in Section 1, we estimate expected total yield per county, $\hat{Y}_i$, by

$$\hat{Y}_i = \sum_{j=1}^{N_G} A_{i,j} P_{i,j} \hat{Y}_{i,j}. \tag{33}$$

Once again vectorizing our estimates, let $\hat{Y}^C = (\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_{N_C})$. Defining P as the appropriate $N_C \times (N_C N_G)$ matrix to produce the above relations, we write

$$\hat{Y}^C = P \hat{Y}^C = P X_Y \beta. \tag{34}$$

Clearly, P depends on the proportions estimated from our data. However, for the purpose of estimating the covariance matrix of $\hat{Y}^C$ we will ignore this dependence. It is our hope

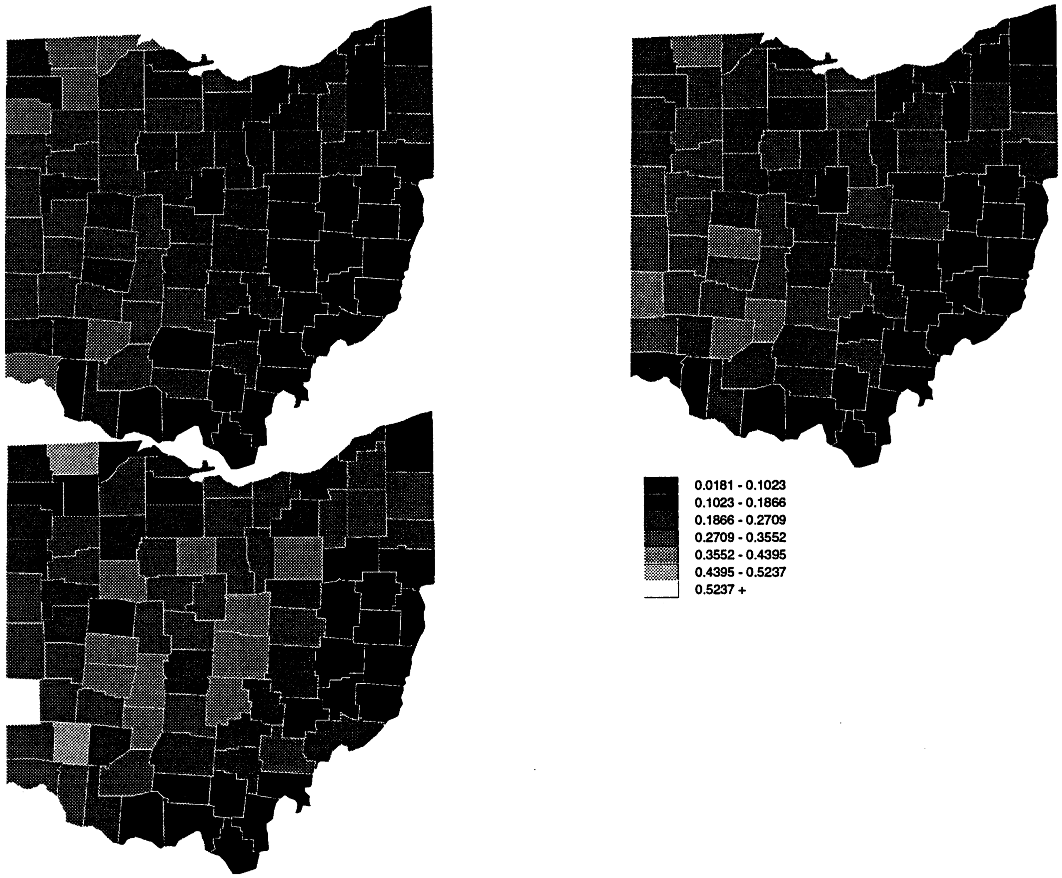

Figure 6: County maps of Ohio are shaded to reflect the proportions of acres harvested in total acres of farmland for each county group pair. The legend is also displayed. The upper left is Group 1. The upper right is Group 2. The lower left is Group 3.

that in the future, the proportions, like the total acreages in corn, can be obtained from the census data. We estimate the covariance of $\hat{Y}^C$ by $PX_Y\hat{M}X_Y^TP^T$. Expected total yield and the associated standard deviation estimates are pictured in Figure 8.

To summarize issues of special interest to statisticians, we first note that the general dependence structure generated by the model does indeed roughly correspond to the geographical features of Ohio. (See Section 3.2.) This lends some credence to the suggestion that the model used does capture relevant behavior. Furthermore, the Bayesian approach taken here, while allowing the imputation of prior information, permits analysis in the presence of variable sample sizes from county to county. Such unbalanced data sets typically cause serious problems in non-Bayesian approaches. On the other hand, we do not claim the model used here is best possible; some concerns and oversimplifications have been mentioned earlier. However, the overall results should encourage additional study of such models, and hybreds, in spatial statistics problems.

APPENDIX

We now provide arguments involving the existence of proper posteriors for a normal, hierarchical Bayesian model. The discussion borrows heavily from that given in Berger (1985, Section 4.6.3).

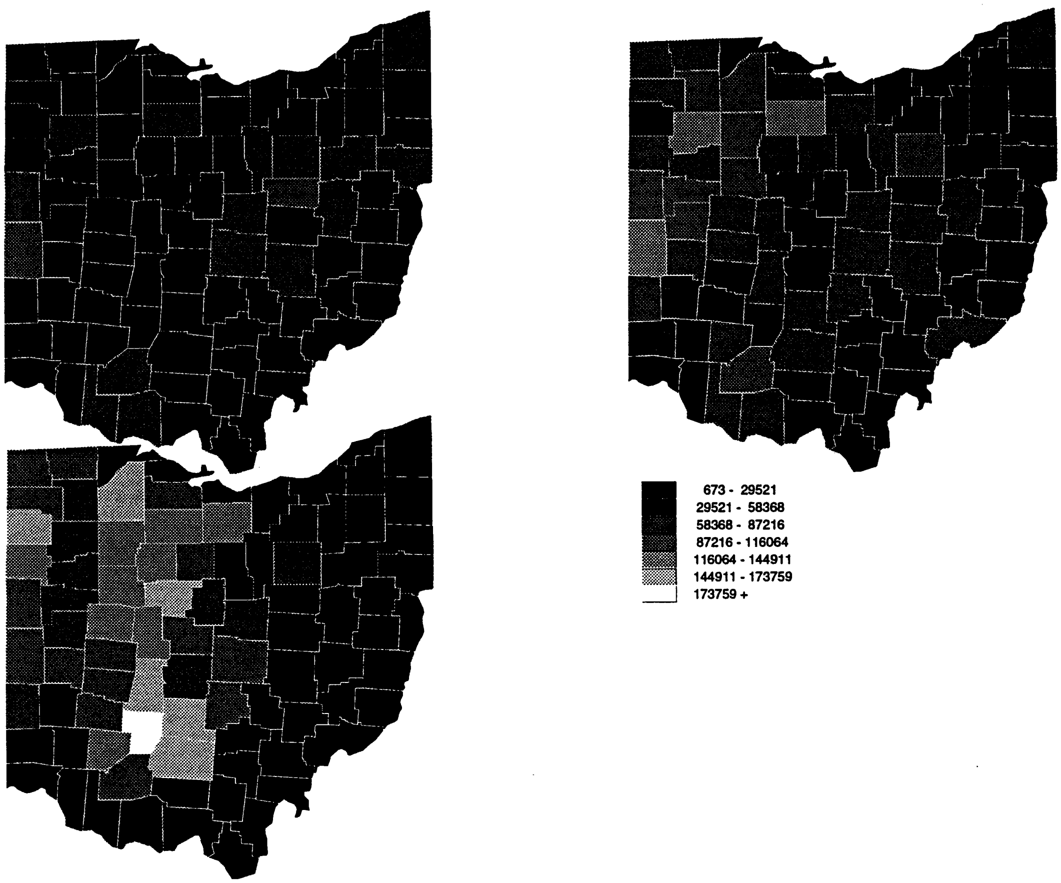

Figure 7: County maps of Ohio are shaded to reflect the total acres of farmland for each county group pair. The legend is also displayed. The upper left is Group 1. The upper right is Group 2. The lower left is Group 3.

Consider the following hierarchical model: For an $(n \times 1)$ data vector $\mathbf{Y}$, assume that

$$\mathbf{Y} \sim N(X\beta, \sigma^2 I)$$

where X is an $n \times p$, full rank design matrix, β is a $p \times 1$ vector of unknown regression coefficients, and σ^2 is the unknown model variance.

Suppose next that the first stage of the prior assigns a p-variate normal distribution to β with mean μ and covariance matrix A:

$$\beta \sim N(\mu, A).$$

Finally, the last stage of the prior assigns independent distributions, denoted respectively by $\pi, \pi_{2,1}$, and $\pi_{2,2}$, to the variables σ^2, μ, and A. For the moment, we only assume that $\pi_{2,2}$ insures that A is positive definite, with probability 1.

To investigate existence of a proper posterior, we consider the implied full joint distribution, denoted by j, and see whether or not it is integrable. Specifically, consider

$$j \propto \frac{\pi(\sigma^2)\ \pi_{2,1}(\mu)\ \pi_{2,2}(A)}{(\sigma^{2n}|A|)^{.5}} exp\{-.5\{(\mathbf{Y} - X\beta)^T(\mathbf{Y} - X\beta)/\sigma^2 + (\beta - \mu)^T A^{-1}(\beta - \mu)\}\}.$$

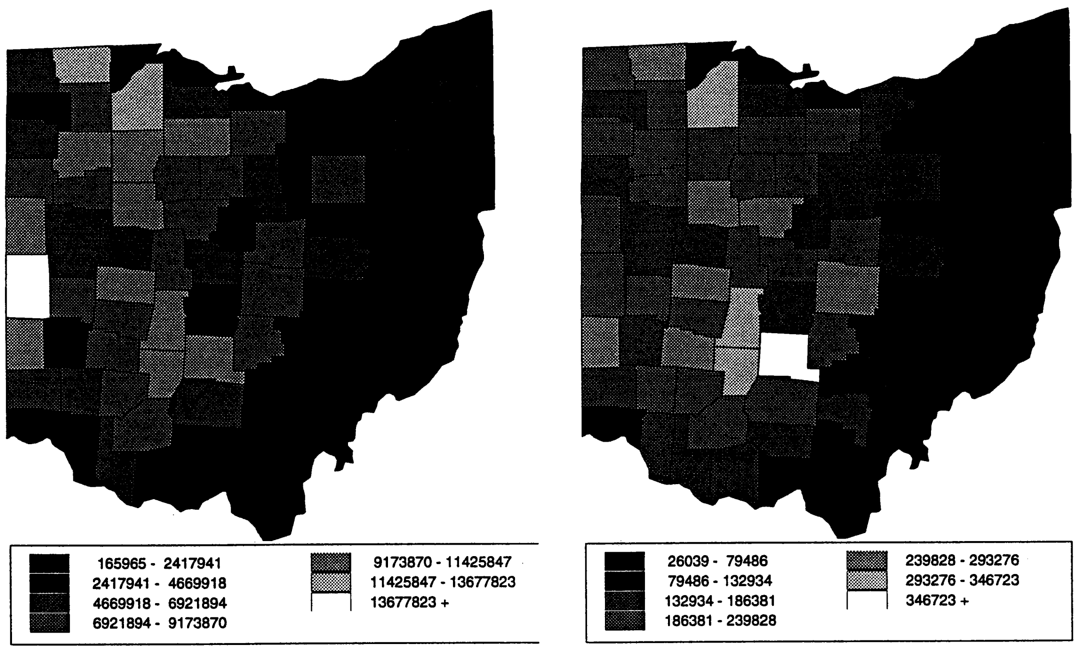

Figure 8: Estimate of total county yield are displayed on the left. Associated standard deviation estimates are pictured on the right.

Following Berger's analysis, consider the familiar representaion

$$(\mathbf{Y} - X\beta)^T(\mathbf{Y} - X\beta) = (\mathbf{Y} - X\hat{\beta})^T(\mathbf{Y} - X\hat{\beta}) + (\hat{\beta} - \beta)^T(X^TX)(\hat{\beta} - \beta),$$

where $\hat{\beta}$ is the usual least squares estimate $(X^TX)^{-1}X^T\mathbf{Y}$. Substituting this expression into j and integrating with respect to β, the resulting integrated joint distribution, denoted ij, is

$$ij \propto \pi(\sigma^2)\ \pi_{2,1}(\mu)\ \pi_{2,2}(A)\ (\sigma^2)^{-.5n} \frac{|\sigma^2(X^TX)^{-1}|^{.5}}{|\sigma^2(X^TX)^{-1} + A|^{.5}} exp\{-.5k\},$$

where

$$k = (\mathbf{Y} - X\hat{\beta})^T(\mathbf{Y} - X\hat{\beta})/\sigma^2 + (\hat{\beta} - \mu)^T[\sigma^2(X^TX)^{-1} + A]^{-1}(\hat{\beta} - \mu).$$

Note that a standard result from linear algebra implies that

$$\frac{|\sigma^2(X^TX)^{-1}|^{.5}}{|\sigma^2(X^TX)^{-1} + A|^{.5}} \leq 1.$$

Therefore, we can guarantee integrability if both $\pi_{2,1}$ and $\pi_{2,2}$ are bounded, even if the natural invariant distribution, $\pi(\sigma^2) = 1/\sigma^2$ is used, as long as n is large. Also, we can interpret the above bound to suggest that the boundedness of $\pi_{2,2}$ is typically necessary.

To apply this reasoning to our model, note that our prior for A is actually parameterized through three hyperparameters: a, σ_C^2, and σ_G^2. Integrability of our model follows directly from the previous discussion. Note that this argument suggests that using the natural noninformative priors, rather than bounded priors, for σ_C^2 and σ_G^2 would have lead to an improper posterior.

References

[1] J. E. Besag, *Spatial interaction and the statistical analysis of lattice systems*, J. R. Statist. Soc. B **36** (1974), 192–236.

[2] Alan E. Gelfand, Susan E. Hills, Amy Racine-Poon, and Adrian F. M. Smith, *Illustration of bayesian inference in normal data models using gibbs sampling*, J. Amer. Statist. Assoc. **85** (1990).

[3] Alan E. Gelfand and Adrian F. M. Smith, *Sampling-based approaches to calculating marginal densities*, J. Amer. Statist. Assoc. **85** (1990), 398–409.

[4] Stuart Geman and Donald Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*, IEEE Trans. Patt. Anal. Mach. Intell. **PAMI-6** (1984), 721–741.

[5] Ross Kinderman and J. Lauri Snell, *Markov random fields and their application*, American Mathematical Society, 1980.

[6] Nancy J. McMillan, *Computational methods for spatial statistics and image data*, Ph.D. thesis, Ohio State University, 1993.

[7] Christian Ritter and Martin A. Tanner, *Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler*, J. Amer. Statist. Assoc. **87** (1992), 861–868.

[8] Adrian F. M. Smith and G. O. Roberts, *Bayesian computation via the gibbs sampler and related markov chain monte carlo methods*, J. R. Statist. Soc. B **55** (1993), 3–23.