Spatial Correlation or Spatial Variation? A Comparison of Gamma/Poisson Hierarchical Models

Katja Ickstadt*and Robert L. Wolpert†
Institute of Statistics and Decision Sciences
Duke University, Durham NC 27708-0251
96-01‡

Revised March 27, 1996

Abstract

Apparent spatial dependence might arise in either of two different ways: from spatial correlation, or from dependence on underlying collateral information (observable or not) that varies spatially. In this paper a class of Bayesian hierarchical models is introduced to account for both spatial correlation and collateral dependence (including spatial trends) for count data. Markov chain Monte Carlo methods using data augmentation are used to evaluate Bayesian posterior distributions for several candidate models from this class; a new computational method is used to evaluate Bayes factors to aid in selecting among the models and so to aid in resolving correlation and spatial variation arising from collateral dependence. The methods are applied to a problem in forest ecology.

Key words: Bayes factors, Bayesian hierarchical models, covariates, data augmentation, Markov chain Monte Carlo.

*Supported by the Deutsche Forschungsgemeinschaft
†Supported by U.S. E.P.A. grant CR822047-01-0
‡Available at url: `ftp://ftp.isds.duke.edu/pub/WorkingPapers/96-01.ps`

1 Introduction

Current methods for analyzing spatial count data force the investigator to compromise. Auto-Poisson models (Besag 1974) cannot reflect positive correlation among counts, even for "nearby" observations; conjugate gamma/ Poisson models (Clayton and Kaldor 1987) cannot reflect correlation of either sign; Gaussian models (Cressie 1993, Cressie and Chan 1989) do not respect the discrete character of the data.

In this paper we introduce a flexible and widely applicable class of hierarchical Bayesian gamma/Poisson models for (possibly) spatially-dependent count data that (may) depend on covariates or other collateral information. We also develop a computational method, extending one of (Newton and Raftery 1994), for evaluating Bayes factors and posterior model probabilities used to solve the variable-selection and model-selection problems that arise in choosing a model from this flexible class. Related random-field models were introduced in (Wolpert and Ickstadt 1995), which did not include dependence on covariates and collateral information or methods for model comparison.

In some applications (communicable disease mapping, population modeling, etc.) apparent spatial dependence arises from some sort of local interaction; in others it has an underlying non-spatial cause: statistical dependence on some collateral information (possibly latent) that features spatial variation, such as regional concentrations of carcinogens or pollutants in disease mapping examples, or of nutrients, sunlight, or moisture in bioabundance examples. Non-spatial collateral variables (treatment, gender, or prior therapy for clinical trials, for example) may also have an explanatory role. Both sources of apparent spatial dependence typically lead to similar intensities (and hence positive apparent correlations) for nearby sites.

The models presented here feature both sources of apparent spatial dependence, and can be used to help distinguish spatial correlation from dependence on reported collateral information. When important covariates are unreported (or when their use is considered inappropriate), the missing spatial dependence can be modeled instead through spatial correlation.

The underlying data will be taken to be a finite or countable collection of counts $\{n_i\}_{i\in I}$ and vectors $\{z_i\}_{i\in I}$ of covariates (collateral variables of all kinds, random or not, often including vectors of spatial coordinates $\{x_i\}_{i\in I}$). At the first stage of the hierarchy the counts are regarded as the observed values of Poisson random variables N_i, conditionally independent given the

values λ_i of their (also random) means Λ_i; at the second stage the Λ_i are taken to be linear combinations of a set $\{\Gamma_j\}_{j\in J}$ of "impulses" with gamma distributions; at the third stage the impulses' distributions and the coefficients in the linear combinations are regarded as uncertain, with an arbitrary joint distribution.

In our analysis spatial and collateral dependence is modeled through the uncertain dependence on covariates z_i of the kernel $k = \{k_{ij}\}$ of linear coefficients relating the Poisson means Λ_i to the impulses Γ_j. We will concentrate on exploring the posterior distributions and expectations of the unobserved Poisson means and of the parameters governing the spatial and collateral dependence, although in other examples interest might center on the impulses Γ_j or on derived quantities such as the covariance. The computational methods used are based on a Markov chain Monte Carlo (MCMC) simulation of samples from the joint posterior distribution of all uncertain quantities, a hybrid Gibbs/Metropolis scheme featuring data augmentation.

The class of Bayesian hierarchical gamma/Poisson models introduced in Section 2 and applied to an example in forest ecology in Section 3 is rich and flexible, offering a wide range of possibilities for modeling covariate and spatial dependence. With this richness and flexibility comes the need to compare and select models; in a Bayesian framework we would like to compute posterior probabilities of models or, to avoid specifying prior probabilities, to compute Bayes factors (Kass and Raftery 1995). The necessary ingredient for either Bayes factors or posterior probabilities is the marginal density for each model, the integral of the conditional density of the observed data with respect to the prior. In Section 4 we develop a method for calculating these marginals from the MCMC stream, related to those of (Chib 1995, Gelfand and Dey 1994, Meng and Wong 1995, Newton and Raftery 1994, Raftery 1996). The results are then discussed in Section 5.

2 Methodology

Consider count data collected in a geographical region of interest denoted by $\mathcal{R}$ (usually a bounded region in the plane), divided into subregions ("plots") $\mathcal{R}_i$ indexed by a finite or countable set I. Our analysis is based only on the observed counts n_i of events in these subregions, and not on the location of events within plots. The locations of the plots are known, however, and perhaps other collateral variables; we collect these into a vector z_i associ-

ated with $\mathcal{R}_i$ and model the counts in a way that will reveal both spatial correlation and spatial variation.

In the examples below we partition a Duke Forest research plot $\mathcal{R}$ into a grid of square plots $\mathcal{R}_i$ and consider the counts N_i of hickory trees in each plot; we also record the spatial coordinates of the center of the plot in a vector z_i. Although we are able to observe the exact location of each tree, we only use the counts N_i for our analysis; see (Wolpert and Ickstadt 1995) for a method of analysis that does not require discretization or partition into subregions.

2.1 The model class

At the lowest stage of the hierarchy the counts are represented by independent Poisson random variables $\{N_i\}_{i\in I}$. At the second stage of the hierarchy the Poisson means are taken to be linear combinations $\Lambda_i \equiv \sum_{j\in J} k_{ij}\Gamma_j$ of unobserved independent gamma-distributed impulses $\{\Gamma_j\}_{j\in J}$, with uncertain nonnegative coefficients $\{k_{ij}\}_{i,j\in I,J}$ that may depend (implicitly) on covariates $\{z_i\}_{i\in I}$. In some applications $I = J$ and each Γ_j will be associated with a subregion $\mathcal{R}_j$ (cf. examples in Section 3).

Uncertainty about the parameters for the gamma distributions of the impulses $\Gamma_j \sim \mathsf{Ga}(\alpha_j, (\beta_j)^{-1})$ and about the values or collateral dependence of the coefficients k_{ij} is modeled at the third stage through dependence of k_{ij}^θ, α_j^θ, and β_j^θ on an uncertain parameter $\theta \in \Theta$, regarded (following the Bayesian paradigm) as random and accorded a prior distribution $\pi(d\theta)$. For simplicity we take $\Theta \subset \mathbb{R}^p$ for some p and take $\pi(d\theta)$ to have a density function $\pi(\theta)$ with respect to Lebesgue measure $d\theta$ on $\mathbb{R}^p$. Dependence on θ is indicated by superscripts; dependence on z_i remains implicit.

Thus, the class of hierarchical gamma/Poisson models introduced here, can be summarized:

$$\begin{array}{lrcl}
\text{Parameter:} & \theta & \sim & \pi(\theta) \\
\text{Impulses:} & \{\Gamma_j\}_{j\in J}|_\theta & \overset{\text{ind}}{\sim} & \mathsf{Ga}(\alpha_j^\theta, (\beta_j^\theta)^{-1}) \\
\text{Intensities:} & \Lambda_i & \equiv & \sum_{j\in J} k_{ij}^\theta \Gamma_j \\
\text{Point counts:} & \{N_i\}_{i\in I}|_{\theta,\{\Gamma_j\}} & \overset{\text{ind}}{\sim} & \mathsf{Poi}(\Lambda_i)\,.
\end{array} \tag{1}$$

Usually the effect of covariates z_i is to modulate the Poisson intensity locally, only in plot $\mathcal{R}_i$; in the spirit of proportional hazard (Cox) survival models (Cox 1972), we write the parameter in two parts as $\theta = (\theta_0, \theta_c)$ and

the kernel k_{ij}^{θ} as a product $k_{ij}^{\theta} = \exp\Big(h(z_i, \theta_c)\Big) k_{ij}^{\theta_0}$ of a "baseline" kernel $k_{ij}^{\theta_0}$ and a covariate effect $\exp\Big(h(z_i, \theta_c)\Big)$. We take $h(\cdot)$ to be linear, leading to the form

$$k_{ij}^{\theta} = \exp(z_i \cdot \theta_c) k_{ij}^{\theta_0} \tag{2}$$

While other possibilities (e.g., an additive decomposition $k_{ij}^{\theta} = h(z_i, \theta_c) + k_{ij}^{\theta_0}$ or $k_{ij}^{\theta} = z_i \cdot \theta_c + k_{ij}^{\theta_0}$) may be more appropriate for particular applications, the multiplicative model has the merits of simplicity and positivity for all values of θ. Multiplicative incorporation of collateral information into the analysis of intensities for a different class of spatial Poisson models is also studied by (Diggle 1990).

Features (including mean vectors and covariance matrices) of the joint distributions of the random variables $\{\Gamma_j\}_{j \in J}$ and $\{N_i\}_{i \in I}$, conditional on $\theta \in \Theta$, can easily be studied from the negative logarithms of their Laplace transforms, the so-called Laplace exponents:

$$\begin{aligned}
-\log \mathsf{E}^{\theta}\Big[e^{-\sum_{j \in J} t_j \Gamma_j}\Big] &= \sum_{j \in J} \log(1 + t_j (\beta_j^{\theta})^{-1}) \alpha_j^{\theta} \\
-\log \mathsf{E}^{\theta}\Big[e^{-\sum_{i \in I} \phi_i N_i} \Big| \{\Gamma_j\}\Big] &= \sum_{i,j \in I,J} (1 - e^{-\phi_i}) k_{ij}^{\theta} \Gamma_j \qquad (3) \\
-\log \mathsf{E}^{\theta}\Big[e^{-\sum_{i \in I} \phi_i N_i}\Big] &= \sum_{j \in J} \log\Big(1 + \sum_{i \in I} (1 - e^{-\phi_i}) k_{ij}^{\theta} (\beta_j^{\theta})^{-1}\Big) \alpha_j^{\theta} \qquad (4) \\
\mathsf{E}^{\theta}\Big[N_i\Big] &= \sum_{j \in J} k_{ij}^{\theta} (\beta_j^{\theta})^{-1} \alpha_j^{\theta} \\
\mathsf{Cov}^{\theta}\Big[N_i, N_{i'}\Big] &= \sum_{j \in J} \Big(\delta_{i'}^{i} + k_{i'j}^{\theta} (\beta_j^{\theta})^{-1}\Big) k_{ij}^{\theta} (\beta_j^{\theta})^{-1} \alpha_j^{\theta} \qquad (5)
\end{aligned}$$

where $\delta_{i'}^{i}$ is the Kronecker symbol, one if $i = i'$ and otherwise zero.

While the $\{N_i\}_{i \in I}$ are conditionally independent Poisson random variates with means $\{\lambda_i\}_{i \in I}$ (cf. (1), (3)), it follows from (4) that *un*conditionally the $\{N_i\}_{i \in I}$ are dependent, distributed as sums of independent negative-binomial random variates. From (5) we see that N_i and $N_{i'}$ are uncorrelated (and in fact independent) only if $k_{i\cdot}^{\theta}$ and $k_{i'\cdot}^{\theta}$ are orthogonal in the sequence space $\ell^2(J, (\beta_{\cdot}^{\theta})^{-2} \alpha_{\cdot}^{\theta})$.

When covariates are absent, the hierarchical Bayesian models given by (1) are a special case of the gamma/Poisson random field models defined in

(Wolpert and Ickstadt 1995). Here the spaces $\mathcal{X}$ of possible event locations and $\mathcal{S}$ of impulse locations reduce to finite or countably infinite sets $\mathcal{X} = \{x_i\}_{i\in I}$ and $\mathcal{S} = \{s_j\}_{j\in J}$, respectively. The present case is simpler technically (the models are formulated for random variables instead of for random fields), but includes the new feature of collateral dependence.

From (4) we can, in principle, compute the likelihood function for the Poisson counts, $\mathsf{P}[\{N_i = n_i\}_{i\in I} \mid \theta]$; it's just the coefficient of $\prod_{i\in I} {z_i}^{n_i}$ in the Taylor expansion of $\mathsf{E}^{\theta}\Big[\exp(-\sum_{i\in I}\phi_i N_i)\Big]$, for $\phi_i \equiv -\log z_i$. Because this computation is tedious and numerically unstable, however, we instead employ an MCMC approach to generate dependent samples from the joint posterior distribution of θ and $\{\Gamma_j\}_{j\in J}$.

2.2 Data augmentation and the MCMC scheme

Our numerical methods below require sampling from the complete conditional distribution of the impulses $\{\Gamma_j\}_{j\in J}$, given θ and the observed counts $\{N_i = n_i\}_{i\in I}$, a mixture distribution from which direct sampling is cumbersome. Therefore, following (Tanner and Wong 1987) we introduce latent "augmentation" variables $\{N_{ij}\}_{i,j\in I,J}$ which can be sampled easily and which lead to a mixture-model representation of model (1) and to conjugate (gamma) complete conditional distributions for the $\{\Gamma_j\}_{j\in J}$.

Conditional on the values $\{\gamma_j\}_{j\in J}$ of $\{\Gamma_j\}_{j\in J}$ and $\{n_i\}_{i\in I}$ of $\{N_i\}_{i\in I}$ set $\lambda_{ij} \equiv k_{ij}^{\theta}\gamma_j$, $\lambda_{i+} \equiv \sum_{j\in J}\lambda_{ij}$, $p_{ij} \equiv \lambda_{ij}/\lambda_{i+}$, and let $\{N_{ij}\}_{i\in I}$ have independent multinomial $\mathsf{MN}(n_i, p_{i\cdot})$ distributions. We might regard the random variables N_{ij} as the portions of count n_i attributable to the j^{th} impulse, $j \in J$; in some applications these impulses might be thought of as causes for the events (sources of nutrition for bioabundance studies, or of carcinogens for cancer disease mapping examples). We now turn to the joint distribution of all uncertain quantities.

Theorem 1 *The joint distribution of $\{N_i\}_{i\in I}$, $\{N_{ij}\}_{i,j\in I,J}$, $\{\Gamma_j\}_{j\in J}$, and θ has a density function with respect to the product of counting measure for discrete and Lebesgue measure for continuous variables, given by*

$$f\Big(\{n_i\},\{n_{ij}\},\{\gamma_j\},\theta\Big) \;=\; \pi(\theta)\left(\prod_{j\in J}\frac{{\gamma_j}^{\alpha_j^\theta+n_{+j}-1}(\beta_j^\theta)^{\alpha_j^\theta}}{\Gamma(\alpha_j^\theta)}\right)$$
$$\times\;\left(\prod_{i,j\in I,J}\frac{(k_{ij}^\theta)^{n_{ij}}}{n_{ij}!}\right)\exp\Big(-\sum_{j\in J}\Big(\beta_j^\theta+k_{+j}^\theta\Big)\gamma_j\Big)$$

where $n_{+j}\equiv\sum_{i\in I}n_{ij}$ *and* $k_{+j}^\theta\equiv\sum_{i\in I}k_{ij}^\theta$.

Proof. Direct calculation. □

This leads directly to the necessary full conditional distributions:

Corollary 1

1. *The conditional distribution of the augmented data vector* $\{N_{ij}\}_{j\in J}$ *given the observed count* n_i*, the impulses* $\{\gamma_j\}_{j\in J}$*, and the parameter* θ *is multinomial with parameters* n_i *and* $p_{i\cdot}$*, independent for* $i\in I$.

2. *The conditional joint distribution of* $\{\Gamma_j\}_{j\in J}$ *given* $\{n_i\}_{i\in I}$*,* $\{n_{ij}\}_{i,j\in I,J}$*, and* θ*, is again gamma, independent for* $j\in J$*, with parameters* $\alpha_j^\theta+n_{+j}$ *and* $\beta_j^\theta+k_{+j}^\theta$.

3. *The conditional density of* θ *given* $\{n_i\}_{i\in I}$*,* $\{n_{ij}\}_{i,j\in I,J}$*, and* $\{\gamma_j\}_{j\in J}$ *is given by*

$$f\Big(\theta\mid\{n_i\},\{n_{ij}\},\{\gamma_j\}\Big) \;\propto\; \pi(\theta)\left(\prod_{j\in J}\frac{{\gamma_j}^{\alpha_j^\theta+n_{+j}-1}(\beta_j^\theta)^{\alpha_j^\theta}}{\Gamma(\alpha_j^\theta)}\right)$$
$$\times\;\left(\prod_{i,j\in I,J}\frac{(k_{ij}^\theta)^{n_{ij}}}{n_{ij}!}\right)\exp\Big(-\sum_{j\in J}\Big(\beta_j^\theta+k_{+j}^\theta\Big)\gamma_j\Big)\,.$$

Proof: Immediate from Theorem 1 and the definition of $\{n_{ij}\}_{i,j\in I,J}$. □

We now study the (analytically intractable) posterior distribution of the uncertain θ and the $\{\Gamma_j\}_{j\in J}$ by simulating steps from an ergodic Markov chain with the posterior for its stationary distribution (Gelfand and Smith

1990, Tierney 1994). Corollary 1 leads directly to the following hybrid Gibbs/ Metropolis MCMC scheme:

Given a prior density $\pi(\theta)$ on Θ, a parametric kernel $\{k_{ij}^{\theta}\}_{i,j\in I,J}$, a transition probability kernel $Q(\theta, d\theta^*) = q(\theta,\theta^*)d\theta^*$, parameters $\{\alpha_j^{\theta}\}$ and $\{\beta_j^{\theta}\}$ for the gamma distributions of the impulses Γ_j, and initial points $\theta^0 \in \Theta$ and nonnegative integers $\{n_{ij}^0\}_{i,j\in I,J}$ satisfying $n_{i+}^0 = n_i$, we can generate successive points starting at $t = 1$ as follows:

1. Gibbs step to update the impulses variables:

   Given $\{n_i\}_{i\in I}$, $\{n_{ij}^{t-1}\}_{i,j\in I,J}$, and θ^{t-1},

   - set $\alpha_j^t \equiv \alpha_j^{\theta^{t-1}} + n_{+j}^{t-1}$ and $\beta_j^t \equiv \beta_j^{\theta^{t-1}} + k_{+j}^{\theta^{t-1}}$;
   - generate $\gamma_j^t \overset{\text{ind}}{\sim} \mathsf{Ga}\left(\alpha_j^t, (\beta_j^t)^{-1}\right)$;
   - set $\lambda_{ij}^t \equiv k_{ij}^{\theta^{t-1}}\gamma_j^t$, $\lambda_{i+}^t \equiv \sum_{j\in J}\lambda_{ij}^t$, and $p_{ij}^t \equiv \lambda_{ij}^t/\lambda_{i+}^t$.

2. Gibbs step to update the augmentation points:

   Given $\{n_i\}_{i\in I}$, $\{\gamma_j^t\}_{j\in J}$, and θ^{t-1},

   - generate $n_{ij}^t \sim \mathsf{MN}(n_i, p_{i\cdot}^t)$, independent for $i \in I$.

3. Metropolis step to update the parameter θ:

   Given $\{n_i\}_{i\in I}$, $\{n_{ij}^t\}_{i,j\in I,J}$, $\{\gamma_j^t\}_{j\in J}$, and θ^{t-1},

   - generate a new candidate $\theta^* \sim Q(\theta^{t-1}, d\theta^*)$;
   - calculate the acceptance probability

$$\begin{aligned} P^* \;\equiv\; & \frac{\pi(\theta^*)q(\theta^*,\theta^{t-1})}{\pi(\theta^{t-1})q(\theta^{t-1},\theta^*)} \prod_{i,j\in I,J}\left[\frac{k_{ij}^{\theta^*}}{k_{ij}^{\theta^{t-1}}}\right]^{n_{ij}^t} \\ \times\; & \prod_{j\in J}\left[\frac{(\gamma_j^t)^{\alpha_j^{\theta^*}+n_{+j}^t-1}(\beta_j^{\theta^*})^{\alpha_j^{\theta^*}}}{(\gamma_j^t)^{\alpha_j^{\theta^{t-1}}+n_{+j}^t-1}(\beta_j^{\theta^{t-1}})^{\alpha_j^{\theta^{t-1}}}}\frac{\Gamma(\alpha_j^{\theta^{t-1}})}{\Gamma(\alpha_j^{\theta^*})}\right] \\ \times\; & \exp\Big(-\sum_{j\in J}\left(\beta_j^{\theta^*} - \beta_j^{\theta^{t-1}} + k_{+j}^{\theta^*} - k_{+j}^{\theta^{t-1}}\right)\gamma_j^t\Big) \end{aligned}$$

   - generate $\theta^t \sim \begin{cases}\theta^* & \text{with probability } \min(1, P^*) \\ \theta^{t-1} & \text{otherwise.}\end{cases}$

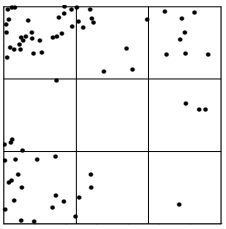
4. Set $t \leftarrow t+1$ and return to step 1.

3 An Example

3.1 The data

As an illustrative example we analyze the spatial density of hickory trees in a $140m \times 140m$ research area in Duke Forest, Durham, North Carolina, for evidence of spatial correlation and systematic variation. The underlying data set consists of a complete census count of every tree with diameter of at least $2.5cm$ at height $1.5m$, including the spatial coordinates (within $0.1m$), the diameter at height $1.5m$ (within $0.1cm$), and an indicator of the tree's condition; it contains 4501 trees of 24 different species. The data were collected in 1951–1952 and updated in 1974, 1982, and 1989 by members of the Duke University Botany Department in an ongoing study of forest maturation (Bormann 1953, Christensen 1977).

Our analysis is based on the full subset of all 75 hickory trees with diameter at least $10cm$ in the 1974 census count. Figure 1a shows the exact tree locations in the research plot, Figure 1b indicates the counts of trees in a 3×3 lattice of equally sized square plots ("quadrats") numerically and through shading. As suggested by (Christensen 1977) we treat the four hickory species present (*carya glabra* (pignut hickory), *carya ovalis* (sweet pignut hickory), *carya ovata* (shagbark hickory), and *carya tomentosa* (mockernut hickory)) as a single group. Our analysis below makes no use of the trees' location within quadrats, nor of their condition or diameter.



29	10	8
5	0	3
16	3	1

(a) Exact locations (b) Plot counts

Figure 1. Hickory tree empirical density.

3.2 The models

Figure 1 reveals some spatial trends: intensities are systematically higher in the west than in the east, and in the north than in the south within the research area. Questions arise about how much of this phenomenon can be explained as spatial correlation, or clustering, and how much as systematic spatial trends.

Model (1) can help answering these questions. Set $I = J = \{1, \ldots, 9\}$ and with each quadrat $j \in J$ associate an impulse $\Gamma_j \sim \mathsf{Ga}(\alpha_j, \beta_j)$. Then model the Poisson intensity for the i^{th} quadrat as $\Lambda_i \equiv \sum_{j \in J} k_{ij}^{\theta} \Gamma_j$, $i \in I$, where the kernel $\{k_{ij}^{\theta}\}_{i,j \in I,J}$ depends on $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) \in \Theta = \mathbb{R}^4$ and $z_i = (y_i, x_i) \in \mathcal{R}$ (spatial coordinates of the center of $\mathcal{R}_i$) through the relation (cf. (2)):

$$k_{ij}^{\theta} = \begin{cases} \exp(\theta_1 + \theta_3 y_i + \theta_4 x_i) & \text{for } i = j \\ \exp(\theta_2 + \theta_3 y_i + \theta_4 x_i) & \text{for } i \sim j \\ 0 & \text{otherwise,} \end{cases}$$

with $i \sim j$ indicating that the quadrats $\mathcal{R}_i$ and $\mathcal{R}_j$ are neighbors. The parameter θ_1 measures the overall log-intensity, the parameter θ_2 measures possible interaction between neighboring quadrats, and the parameters θ_3 and θ_4 measure dependence on the two included covariates, measures y_i of "northness" and x_i of "eastness" (Ripley 1981).

To assess how much of the variability of our data should be attributed to spatial correlation (i.e., to non-zero $\exp(\theta_2)$) and how much to spatial trends (i.e., to non-zero θ_3 and θ_4) we would like to compare this model to some nested models with fewer parameters:

Model	Parameters	Fixed Values
M_1	$\theta = (\theta_1)$	$\exp(\theta_2) = \theta_3 = \theta_4 = 0$
M_2	$\theta = (\theta_1, \theta_2)$	$\theta_3 = \theta_4 = 0$
M_3	$\theta = (\theta_1, \theta_3, \theta_4)$	$\exp(\theta_2) = 0$
M_4	$\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$	

For each model we have chosen independent standard normal prior distributions for each parameter θ_k, and independent $\mathsf{Ga}(\alpha_j, (\beta_j)^{-1})$ distributions for the Γ_j with $\alpha_j = 75/9$ and $\beta_j = 1$ for each $j \in J$; in our example the parameters α_j and β_j do not depend on θ.

3.3 Posterior analysis

Estimated posterior mean intensities $\mathsf{E}[\Lambda_i]$ based on 5,000 MCMC steps are presented in Figure 2 for each of the four models, both numerically and through shading. We observe that the models M_1 and M_2, although they also preserve the spatial trends, smooth the data much more than the models M_3 and M_4. The posterior mean intensities of models M_1 and M_2 differ very little, as do those of models M_3 and M_4, indicating that nearest-neighbor correlation (the influence of parameter θ_2, included only in models M_2 and M_4) may have a negligible effect.

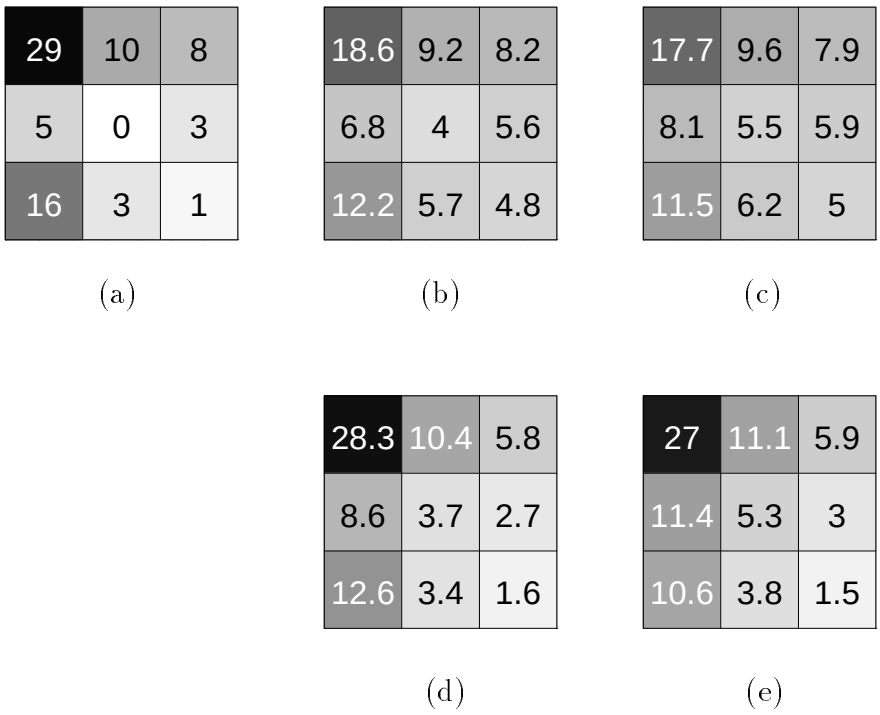


Figure 2. Hickory tree empirical (a) and posterior densities for model M_1 (b), model M_2 (c), model M_3 (d), and model M_4 (e).

Figure 3 shows the prior (curve) and the estimated posterior (histogram) density function for all four parameters of model M_4 with the posterior means $(-0.52, -2.71, 0.60, -0.83)$ indicated by vertical lines. This suggests that the systematic trends should be included in the analysis, but the spatial correlation is very weak (i.e., $\exp(\theta_2)$ is small).

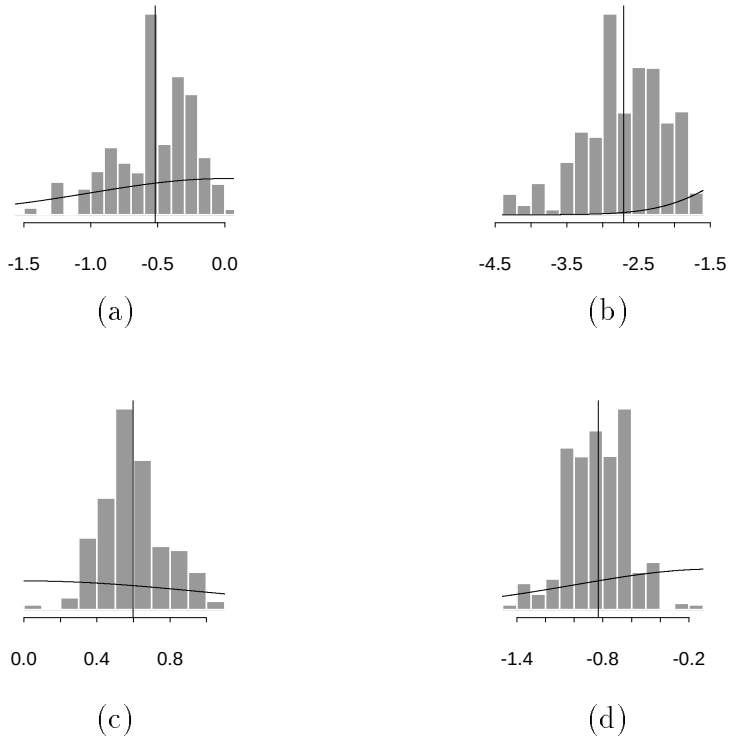


Figure 3. Prior (solid line) and posterior (histogram) density functions and posterior means (vertical lines) for θ_1 (a), θ_2 (b), θ_3 (c), and θ_4 (d).

4 Model Comparison

While the graphical outputs (cf. Fig. 2) are very suggestive of which models fit well and which do not, a more objective and more quantitative approach would be to construct a mixture model assigning prior probabilities $\mathsf{P}[M_m]$ (and perhaps utilities U_m) to model M_m and to prefer the model with the greatest posterior probability (or expected utility); these probabilities (and their ratios, the Bayes Factors) depend on the data and model only through the *marginal probability densities* (sometimes called marginal likelihoods), the integral over all parameters and latent variables of the joint density function for the data, parameters, and latent variables. For simplicity we consider only zero-one loss (or equal utilities), and prefer the model with the greatest

posterior probability.

4.1 Bayes factor calculations

Here we present a method to calculate these marginal probabilities. Although an analytical evaluation via the likelihood function for models of type (1) is intractable, they can easily be computed numerically within the MCMC sampling scheme using data augmentation. Our method generalizes that of (Newton and Raftery 1994), which does not apply directly to problems that include latent variables or augmented data, confirming the suggestion in (Raftery 1993) that sampling methods can be extended to cover this case; for another method suitable for Gibbs schemes with latent variables (but not for our hybrid Gibbs/Metropolis scheme), see (Chib 1995). The computations will be given first in general, for an arbitrary model including latent variables, then specific formulas will be given for models of type (1).

In this section let x denote the observed value of the data vector X, z a possible value of a (random) data augmentation vector Z, ξ a vector of latent variables, and θ the vector of uncertain parameters. In practice any unobserved quantities might be viewed either as variables (hence part of Z) or parameters (hence part of ξ); usually one choice will be more convenient than the other, as shown below.

Here we consider only the case in which the joint probability distribution of X, Z, ξ, and θ under each model M_m is absolutely continuous with respect to a product of Lebesgue measure (for continuous quantities) and counting measure (for discrete quantities), with a density function $f_m(x, z, \xi, \theta)$; in a slight abuse of notation we will denote marginal and conditional densities for all quantities x, z, ξ, θ all by $f_m(\cdot)$ or $f_m(\cdot \mid \cdot)$, distinguishing them only by their arguments, and denote both Lebesgue and counting measure by dz, $d\xi$, and $d\theta$; thus the prior densities are $\pi_m(\theta) = f_m(\theta)$, the likelihoods $f_m(x \mid \theta)$, etc. Recall that the Bayes factor of model M_m against model M_n is simply the ratio of marginal probability densities,

$$B_{mn} = \frac{f_m(x)}{f_n(x)}. \tag{6}$$

Now let Z^t, ξ^t, and θ^t be steps of an ergodic Markov chain with stationary distribution $f_m(z, \xi, \theta \mid x)$; the superscript t again indicates the MCMC iteration.

Theorem 2 *If (z^t, ξ^t, θ^t) are steps of an ergodic Markov chain with stationary distribution $f_m(z, \xi, \theta \mid x)$, and if $g_m(x, z, \xi, \theta)$ is measurable and integrable with*

$$\bar{g}_m(x) \equiv \iiint g_m(x, z, \xi, \theta)\, f_m(x, z, \xi, \theta)\, dz\, d\xi\, d\theta,$$

then the marginal probabilities can be computed by the almost-sure limits

$$\lim_{T\to\infty} \bar{g}_m(x) \left[\frac{1}{T}\sum_{t=1}^{T} g_m(x, z^t, \xi^t, \theta^t)\right]^{-1} = f_m(x).$$

Proof: Direct application of the ergodic theorem in the form of (Meyn and Tweedie 1993, Theorem 17.1.7, p. 416) or (Tierney 1994, Theorem 3). □

Corollary 2

$$f_m(x) = \lim_{T\to\infty} \left[\frac{1}{T}\sum_{t=1}^{T} \frac{f_m(z^t \mid x, \xi^t, \theta^t)}{f_m(x, z^t \mid \xi^t, \theta^t)}\right]^{-1} = \lim_{T\to\infty} \left[\frac{1}{T}\sum_{t=1}^{T} \frac{1}{f_m(x \mid \xi^t, \theta^t)}\right]^{-1}. \tag{7}$$

Proof: Apply Theorem 2 with the $f_m(x, z, \xi, \theta)\, dz\, d\xi\, d\theta$-integrable function $g_m(x, z, \xi, \theta) \equiv f_m(z \mid x, \xi, \theta)/f_m(x, z \mid \xi, \theta) = 1/f_m(x \mid \xi, \theta)$, and $\bar{g}_m(x) = 1$. □

Corollary 3 *For models of type (1) simulation-consistent estimates of the marginal probabilities of the observed data $\{n_i\}_{i\in I}$ are given by*

$$f_m(\{n_i\}) = \lim_{T\to\infty} \left[\frac{1}{T}\sum_{t=1}^{T} \prod_{i\in I} \frac{n_i!\, \exp(\lambda_{i+}^t)}{(\lambda_{i+}^t)^{n_i}}\right]^{-1} \tag{8}$$

$$= \lim_{T\to\infty} \left[\frac{1}{T}\sum_{t=1}^{T} \prod_{j\in J} \left(\frac{\Gamma(\alpha_j^{\theta^t})\left(\beta_j^{\theta^t} + k_{+j}^{\theta^t}\right)^{\alpha_j^{\theta^t}+n_{+j}^t}}{\Gamma(\alpha_j^{\theta^t} + n_{+j}^t)\left(\beta_j^{\theta^t}\right)^{\alpha_j^{\theta^t}}}\right) \prod_{i\in I} \frac{n_i!}{\left(k_{i+}^{\theta^t}\right)^{n_i}}\right]^{-1} \tag{9}$$

where $\lambda_{ij}^t \equiv k_{ij}^{\theta^t}\gamma_j^t$ and $\lambda_{i+}^t \equiv \sum_{j\in J} \lambda_{ij}^t$.

Proof: For models of type (1) the observed data are $x = \{n_i\}_{i \in I}$, the augmented data are $z = \{n_{ij}\}_{i,j \in I,J}$, the latent variables are $\xi = \{\gamma_j\}_{j \in J}$, and the parameter is θ.

The estimator of Equation (8) is simply the application of Equation (7) to models of type (1); Equation (9) follows from Theorems 1 and 2 and direct calculation (using the gamma integral and multinomial theorem) that $\bar{g}_m(x) \equiv \iiint g_m(x, z, \xi, \theta)\, f_m(x, z, \xi, \theta)\, dz\, d\xi\, d\theta = 1$. □

Theorem 2 and its corollaries allow one to compute marginal probabilities $f_m(\{n_i\}_{i \in I})$ and hence Bayes factors $B_{mn} = f_m(\{n_i\}_{i \in I}) / f_n(\{n_i\}_{i \in I})$ to compare models of type (1) that differ in their priors $\pi_m(\theta)$, kernels k_{ij}^{θ}, or hyperparameters α_j^{θ}, β_j^{θ}. For example, models with and without interactions for neighboring sites x_i can be compared, or those with and without explicit dependence on collateral information.

4.2 Results

Table 1 shows Bayes factors for comparing the models $\{M_m\}_{m=1\ldots4}$ of Section 3.2. Recall $\theta \in \Theta \subset \mathbb{R}^m$ in model M_m, and that models M_2 and M_4 feature spatial correlation, while models M_3 and M_4 feature explicit dependence on collateral (spatial) information. Evidently model M_3 would be preferred to each competitor, all other things being equal, but the one-parameter model M_1 might be preferred by those with an extreme preference for parsimony.

	M_1	M_2	M_3	M_4
M_1	1	$1/B_{21}$	$1/B_{31}$	$1/B_{41}$
M_2	0.0086	1	$1/B_{32}$	$1/B_{42}$
M_3	9948.5	$1.15 \cdot 10^6$	1	$1/B_{43}$
M_4	300.96	$3.49 \cdot 10^4$	0.0303	1

Table 1. Bayes factors for models M_1–M_4.

A formal way to include preference for parsimonious models would be to assign unequal prior probabilities (or, equivalently, utilities) to the models reflecting this predisposition, and to compute their posterior probabilities. Table 2 shows the prior and posterior probabilities of each model M_m for three different prior distributions: an equiprobable prior assigning $\mathsf{P}[M_m] \propto 1$, a prior somewhat predisposed to parsimony assigning $\mathsf{P}[M_m] \propto e^{-m/2}$, and a more aggressively pro-parsimony prior with $\mathsf{P}[M_m] \propto 4^{-m}$. Each of these is

of the form $\mathsf{P}[M_m] \propto e^{-\Pi m/2}$, with (respectively) $\Pi = 0, 1, 2\log 4$, complexity penalties suggested by the discussion in (Foster and George 1994, Table 2) of the Least-Square, Adjusted R^2, and Risk Inflation criteria, respectively.

	M_1	M_2	M_3	M_4
Neutral	0.2500	0.2500	0.2500	0.2500
	0.0001	0.0000	0.9705	0.0294
Mild	0.4551	0.2760	0.1674	0.1015
	0.0003	0.0000	0.9817	0.0180
Aggressive	0.7529	0.1882	0.0471	0.0118
	0.0016	0.0000	0.9909	0.0075

Table 2. Prior and posterior probabilities of models M_1–M_4.

These results are consistent with visual impressions from Figure 2, which suggests that models M_1 and M_2 (which lack explicit dependence on the collateral (spatial) information, and so cannot reflect an overall spatial trend) over-smooth the data, while the inclusion of spatial interaction in models M_2 and M_4 does not appear to improve the fit; even a strong preference for parsimony is not enough to lead to a preference for the simplest model, M_1, over the most-favored model M_3. See the Discussion Section for further comments on spatial correlation and spatial trends in this example.

It is worth remarking that ergodic convergence of $g_m(x, z^t, \xi^t, \theta^t)$ for the Bayes factor calculations is much slower than that for the model parameters $\{\gamma_j^t\}_{j\in J}$ and θ^t. The histograms of Figure 3 suggest that each θ_k^t has a nearly-normal distribution with moderate tails, while a glance at Figure 4 shows that the quantities $g_m(x, z^t, \xi^t, \theta^t)$ averaged in the Bayes factor calculations have much heavier tails and hence slower convergence (the figure shows $g_4(x, z^t, \xi^t, \theta^t)$ for the sequence from Equation (9), which converges more quickly than that of Equation (8) and Corollary 2; note in Figure 4a that the vast majority of the points are below the mean of about $1 \cdot 10^{10}$, while the few values that extend as high as $6 \cdot 10^{13}$ dominate the ergodic average). Figure 4b suggests that the logarithms $Y^t \equiv \log g_m(x, z^t, \xi^t, \theta^t)$ have nearly a normal distribution; if the sampling distribution of the Y^t were *exactly* normal with some mean μ and variance σ^2, then we could exploit the rapid convergence of the sample mean $\bar{Y}_T$ and variance $S_T^2(Y)$ to estimate the marginal probability densities $f_m(x)$ using the expectation of the log-normal quantities $g_m(x, z^t, \xi^t, \theta^t)$,

$$f_m(x) = \lim_{T\to\infty} \bar{g}_m(x) \exp(\bar{Y}_T + S_T^2(Y)/2). \tag{10}$$

The slow convergence of ergodic averages used in estimating Bayes factors was earlier noted in (Newton and Raftery 1994) and also in (Kass and Raftery 1995), where the possibility of using logarithms is suggested in a somewhat different context.

(a) (b)

Figure 4. Sample histograms for (a) $g_4(x, z^t, \xi^t, \theta^t)$ and (b) $\log g_4(x, z^t, \xi^t, \theta^t)$.

5 Discussion

In the example we illustrated model comparison by comparing models with and without spatial correlation, with and without a linear spatial trend, all with a fixed 3×3 partition of the $140m \times 140m$ research plot into nine elements. The linearity of the spatial trend is not essential—any smooth large-scale trend will appear linear locally, and we do not suggest using the present model estimates to make predictions outside the small research plot.

One might also explore different discretizations; Figure 5 shows empirical density plots for several different choices of partitions.

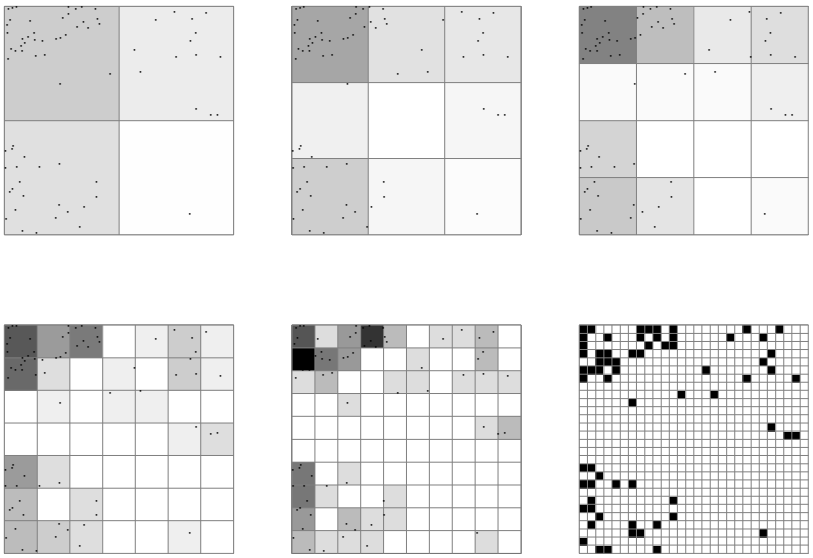


Figure 5. Empirical hickory tree density plots for various discretizations.

Different spatial patterns emerge at different resolutions; the apparent presence of a linear trend and absence of significant interaction at the 3×3 level may not reflect the models' behavior in more refined partitions, where intensities are more likely to be similar in neighboring quadrats. It is disconcerting that inference and conclusions may depend on an inadvertent and arbitrary choice of the level of refinement.

One alternative, pursued in (Wolpert and Ickstadt 1995), is to dispense with discretization altogether and model the precise locations of each event; another would be to use the model comparison methods of Section 4 to compare models at different levels of refinement. In doing so one may wish to penalize the additional complexity of more refined models; a $k \times k$ model will include k^4 linearly independent latent variables $\{\gamma_j\}_{j \in J}$ and $\{n_{ij}\}_{i,j \in I,J}$ (81 for our 3×3 example, but 625 for a 5×5 case and $10,000$ for a 10×10 model). Of course the methods are applicable to other lattice structures (hexagonal, triangular, etc.), and to discrete settings without a lattice structure at all—disease mapping or population modeling, for example, where in place of quadrats we might have counties or regions and where the concern about arbitrary level of discretization may not arise.

Figure 5 suggests that hickory density is higher near the edges of the research plot (which also holds some twenty other tree species, including hundreds of rather evenly-spaced oaks; see (Christensen 1977)). Perhaps this is evidence that hickories are urging into this predominantly oak stand, or receding from it; an interesting object for further study would be the evolution of spatial patterns within the plot for each species, and the spatial study of biodiversity.

Acknowledgments

The authors would like to thank Norm Christensen for sharing data and insight and Michael Lavine, Jim Berger, and Noel Cressie for provocative questions and helpful suggestions. This work was supported by grants from the United States Environmental Protection Agency, the United States National Science Foundation, and the Deutsche Forschungsgemeinschaft.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society (Series B)*, 35(2):192–236.

Bormann, F. H. (1953). The statistical efficiency of sample plot size and shape in forest ecology. *Ecology*, 34:474–487.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313–1322.

Christensen, N. L. (1977). Changes in structure, pattern and diversity associated with climax forest maturation in Piedmont, North Carolina. *American Midland Naturalist*, 97:176–188.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age–standardized relative risks for use in disease mapping. *Biometrics*, 43:671–681.

Cox, D. R. (1972). Regression models and life–tables (with discussion). *Journal of the Royal Statistical Society (Series B)*, 34:187–220.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York, NY, USA.

Cressie, N. A. C. and Chan, N. H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, 84(406):393–401.

Diggle, P. J. (1990). A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society (Series A)*, 153:349–362.

Foster, D. P. and George, E. I. (1994). The risk inflation criterion for multiple regression. *Annals of Statistics*, 22(4):1947–1975.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society (Series B)*, 56(3):501–514.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90:773–796.

Meng, X.-L. and Wong, W. H. (1995). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. Technical Report 365, University of Chicago. To appear in *Statistica Sinica*.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, New York, NY, USA.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society (Series B)*, 56:3–48.

Raftery, A. E. (1993). Discussion of three papers on the Gibbs sampler and other Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society (Series B)*, 55:85.

Raftery, A. E. (1996). Hypothesis testing and model selection. In Gilks, W. R., Spiegelhalter, D. J., and Richardson, S., editors, *Markov Chain Monte Carlo in Practice*, pages 163–187, New York, NY, USA. Chapman & Hall.

Ripley, B. D. (1981). *Spatial Statistics*. John Wiley & Sons, New York, NY, USA.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–550.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, 22(4):1701–1762.

Wolpert, R. L. and Ickstadt, K. (1995). Gamma/Poisson random field models for spatial statistics. Discussion Paper 95-43, Duke University ISDS, USA.