

**MARKOV CHAIN MONTE CARLO FOR
AUTOLOGISTIC REGRESSION MODELS
WITH APPLICATION TO THE
DISTRIBUTION OF PLANT SPECIES**

Fred W. Huffer¹
Department of Statistics
The Florida State University
Tallahassee, FL 32306-4330
Tel. (904) 644-6696
Email: huffer@stat.fsu.edu

and

Hulin Wu
Frontier Science & Technology
Research Foundation
1244 Boylston Street, Suite 303
Brookline, MA 02167-2104
wu@sdac.harvard.edu

Acknowledgements: We thank Dr. D.W. Crumpacker of the University of Colorado at Boulder for providing us with species and climate data. This research was partially supported by EPA Grant #CR818052-01-0 to Dr. Crumpacker. We would also like to thank the referees and an associate editor for their helpful comments.

¹corresponding author

Markov Chain Monte Carlo for Autologistic Regression Models with Application to the Distribution of Plant Species

Fred W. Huffer¹ and Hulin Wu²

¹Department of Statistics, Florida State University,
Tallahassee, FL 32306-4330, U.S.A.

²Frontier Science & Technology Research Foundation,
1244 Boylston Street, Suite 303, Brookline, MA 02167-2104, U.S.A.

SUMMARY

In this paper, we explore using autologistic regression models for spatial binary data with covariates. Autologistic regression models can handle binary responses exhibiting both spatial correlation and dependence on covariates. We use Markov chain Monte Carlo (MCMC) to estimate the parameters in these models. The distributional behavior of the MCMC maximum likelihood estimates (MCMC MLEs) is studied via simulation. We find that the MCMC MLEs are approximately normally distributed and that the MCMC estimates of Fisher information may be used to estimate the variance of the MCMC MLEs and to construct confidence intervals. Finally, we illustrate by example how our studies may be applied to model the distribution of plant species.

1 Introduction

This project began as an attempt to model the distribution of plant species in terms of climate variables like temperature and rainfall. In particular, we were given data on the

Key words: Ecological data; Environmental statistics; Maximum likelihood estimate; Sampling distribution; Spatial binary data; Spatial statistics.

distribution of about 180 plant species in the state of Florida, and data on the values of 9 climate variables which were expected to be important factors in determining the distribution of the plant species. This data was assembled by Dr. D.W. Crumpacker of the University of Colorado at Boulder. The plant distribution data was obtained by digitizing maps from Little (1978). The climate information was compiled from a number of sources.

Figure 1 presents four of the distribution maps, species No. 1 (*Chamaecyparis Thyoides*), No. 4 (*Pinus Clausa*), No. 38 (*Castanea Pumila*) and No. 158 (*Zanthoxylum Clava-herculis*).

Place of Figure 1

The climate variables we are using are the following:

TMM = mean minimum temperature in degrees centigrade of coldest month
(usually January).

TM = mean temperature in degrees centigrade of coldest month (usually January).

TAV = mean annual temperature in degrees centigrade.

LT = lowest temperature recorded from 1931 to 1990 in degrees centigrade.

FZF = median freeze-free period in days.

PRCP = mean total annual precipitation in millimeters.

MI = moisture index = $(PRCP)/(TAV \times 58.93)$.

where $TAV \times 58.93$ = estimate of mean annual potential evapotranspiration
by the Holdridge method.

PMIN = mean total precipitation of driest month in millimeters.

ELV = elevation in feet.

Inspection of maps like those in Figure 1 suggests that there are strong climate

effects on species distribution. In particular, the species boundaries are often observed to roughly follow contour lines for one of the temperature variables TMM, TM, TAV, LT or FZF. Inspection also reveals a strong degree of spatial correlation in the data, that is, whether a species is present or absent at a given site is strongly related to its presence or absence at neighboring sites. Some correlation is to be expected simply because the binary responses (presence or absence) depend on the climate covariates which are themselves spatially correlated and display definite spatial patterns. However, for many species the degree of spatial correlation is much larger than can be explained in terms of the covariates alone.

Ecologists have analyzed similar data (sometimes using pollen abundance as a proxy for the actual species distribution) by a variety of means. Box et al. (1993) built a climatic-envelope model for many woody Florida plant species. This paper is particularly relevant to the current work as they use very similar climate variables and many of the same plant species that we do. Bartlein et al. (1986) and Huntley et al. (1989) constructed response surfaces for some plant species data by a locally-weighted averaging technique. This gave them graphical displays of the relationship between species distribution and climate. Austin et al. (1990) used logistic regression to model the dependence on the climate variables, but they did not consider spatial interaction/correlation in their models. In this paper, we try to model the binary data (species present or absent) in ways which properly account for both the spatial correlation and the dependence on the climate covariates.

This kind of spatial binary data with covariates occurs frequently in many fields such as geology, meteorology, medical science, and environmental science, so it is of great interest to develop and study models for dealing with this data. The model we shall use in this paper is the autologistic regression model which is explained in the next section (Section 2). The autologistic regression model is a special case of the general autologistic model introduced by Besag (1974, 1975) and studied by many authors (see Zhao and Prentice (1990), Geyer and Thompson (1992), Wu (1994), Wu and Huffer (1996)). See Gumpertz, Graham and Ristaino (1997) for an interesting application of the autologistic regression model to investigate the effects of soil variables on the

incidence of disease in agricultural fields.

The parameters in this model will be estimated by a Markov Chain Monte Carlo (MCMC) approximation to the maximum likelihood estimates as in Geyer (1991a). This method of parameter estimation is reviewed in Section 3. Simulation studies on the distributional properties of MCMC estimates and confidence intervals are presented in Sections 4 and 5. We find that the MCMC estimates are roughly normally distributed and that confidence intervals based on the estimated Fisher information perform very well so long as the amount of spatial correlation is not too large. In Section 6, we apply the methods proposed in this paper to one of the plant species (No. 38).

2 The Autologistic Regression Model

We shall suppose that our data is recorded at m locations (sites) forming a subset \mathcal{S} of a rectangular lattice. Each site in \mathcal{S} is described by giving coordinates (k, ℓ) specifying the row and column of the lattice at which it is located. The sites in \mathcal{S} are numbered from 1 to m in some arbitrary fashion. At each site i we observe a binary response y_i and a $p \times 1$ vector of covariates \mathbf{x}_i . Taken altogether, the m binary responses constitute a map $Y = (y_i)$. The autologistic regression model specifies the conditional probability p_i that $y_i = 1$ given all the other values y_j ($j \neq i$) as follows:

$$\begin{aligned} p_i &= P(y_i = 1 | \text{all other values}) = P(y_i = 1 | \text{nearest neighbors}) \\ &= \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad \text{where} \quad \eta_i = \beta_0 + \mathbf{x}_i' \boldsymbol{\beta}_1 + \gamma y_i^*. \end{aligned} \tag{1}$$

Here y_i^* denotes the neighborhood sum for site i , that is,

$$y_i^* = \sum_{j=1}^m y_j I(i \sim j) = \sum_{j: i \sim j} y_j \tag{2}$$

where $i \sim j$ indicates that sites i and j are “neighbors”. For a site i with coordinates (k, ℓ) , the neighbors j are any members of \mathcal{S} occupying the four nearest locations $(k - 1, \ell)$, $(k + 1, \ell)$, $(k, \ell - 1)$, $(k, \ell + 1)$. A site in the interior of \mathcal{S} will have 4 neighbors. Sites on the boundary of \mathcal{S} will have fewer neighbors.

The parameters in this model are the intercept β_0 , a $p \times 1$ vector β_1 which specifies the covariate effects, and a parameter γ which determines the degree of spatial interaction or correlation in the data. When $\gamma = 0$, the model (1) reduces to the ordinary logistic regression model which is appropriate for independent binary responses given the covariates. When $\beta_1 = 0$, the model becomes the special case of the autologistic model used in Besag (1974, 1975).

In this paper we will concentrate on the model described above. There are many other models that could be used for spatial binary data. The model specified in (1) and (2) uses only the four nearest neighbors. One might consider higher-order autologistic models which use a larger number of neighbors. One could also replace the logistic model in (1) by the corresponding probit or complementary log-log model (McCullagh and Nelder, 1989). An interesting possibility (suggested by a referee) is a model in which the responses y_i at each site are assumed to be conditionally independent given the values of the probabilities p_i whose probits are assumed to follow a spatial Gaussian process. Choosing a model based on the autologistic model, as we have done, has some clear advantages. Our model is a natural generalization of the commonly used and well understood logistic regression model. Also, our model leads to an exponential family with a low-dimensional sufficient statistic (see Section 3.1). This is crucial for the implementation of MCMC estimation schemes. One final remark: we know of no model for correlated spatial binary data with covariates which does not involve very extensive computation.

3 Markov Chain Monte Carlo Method

The spatial correlation term γy_i^* in (1) causes difficulty for parameter estimation. Carrying out exact maximum likelihood estimation is difficult because the likelihood function is not tractable except when the number of sites m is quite small. Two estimation methods, the coding method and maximum pseudolikelihood (MPL) method have been proposed by Besag (1974, 1975). But these two methods are not efficient. Wu and Huffer (1996) and other authors concluded that the MCMC method is superior to the

other two methods. In particular, Wu and Huffer (1996) have carried out simulation studies comparing the coding method, MPL method, and the MCMC method on data generated from the model (1). They found that when the amount of spatial interaction (as determined by γ) is sufficiently large, the MCMC MLE's are a substantial improvement on the other two methods. For the species data we are modeling, the spatial interaction is quite large. For this reason we shall restrict our attention to the MCMC MLE's in this paper.

3.1 Description of the MCMC Method

Markov chain Monte Carlo methods can be used to approximate the MLE for any family of distributions having probability densities known up to a constant of proportionality (Geyer, 1991a,b, 1992, 1994; Geyer and Thompson, 1992) We shall first describe the MCMC method in general, following the notation of Geyer (1991b), and then specialize our discussion to the model (1).

Suppose the probability measure P_θ of our data y has a density or mass function f_θ (with respect to a measure μ) which can be written as:

$$f_\theta(y) = \frac{1}{z(\theta)} h_\theta(y)$$

where the function h_θ is known, but the normalizing constant

$$z(\theta) = \int h_\theta(y) d\mu(y)$$

is intractable. If we can generate observations Y_1, Y_2, \dots, Y_n from P_ψ , then we can estimate the ratio $z(\theta)/z(\psi)$ using the relation

$$\frac{z(\theta)}{z(\psi)} = \int \frac{h_\theta(y)}{h_\psi(y)} f_\psi(y) d\mu(y) = \mathbb{E}_\psi \frac{h_\theta(Y)}{h_\psi(Y)} \approx \frac{1}{n} \sum_{i=1}^n \frac{h_\theta(Y_i)}{h_\psi(Y_i)}.$$

Here we are taking θ as a variable and ψ as fixed. This gives an estimate $l_n(\theta)$ of the log-likelihood:

$$l_n(\theta) = \log \frac{h_\theta(Y_{\text{obs}})}{h_\psi(Y_{\text{obs}})} - \log \left(\frac{1}{n} \sum_{i=1}^n \frac{h_\theta(Y_i)}{h_\psi(Y_i)} \right). \quad (1)$$

Here we are using Y_{obs} to denote the observed data, and Y_1, Y_2, \dots, Y_n to denote the Monte Carlo sample generated from P_ψ . The maximizer of $l_n(\theta)$ is the Monte Carlo approximant of the MLE.

When the probability measures P_θ form an exponential family with $\theta \in R^p$, we can write

$$h_\theta(y) = e^{\theta' t}$$

where $t = t(y)$ is the vector of sufficient statistics. In this case, for given observed data Y_{obs} , the maximum likelihood estimate is obtained by finding θ which satisfies $E_\theta t(Y) = t(Y_{\text{obs}})$. The Monte Carlo analog of this result can be obtained by differentiating the log-likelihood function (1) and setting it equal to 0. i.e.

$$\frac{\sum_{j=1}^n T_j e^{(\theta-\psi)' T_j}}{\sum_{j=1}^n e^{(\theta-\psi)' T_j}} = T_{\text{obs}} \quad (2)$$

where we define $T_{\text{obs}} = t(Y_{\text{obs}})$ and $T_j = t(Y_j)$ for $j = 1, 2, \dots, n$. The Newton-Raphson method can be used to solve this equation. The solution of Equation (2) for θ is the Monte Carlo MLE denoted $\hat{\theta}_n$. Note that this solution depends on the Monte Carlo sample Y_1, Y_2, \dots, Y_n only through the sufficient statistics T_1, T_2, \dots, T_n which form a set of points in R^p .

In our situation the data y consists of a binary map $y = (y_i)_{i=1}^m$ and the joint probability function of model (1) forms an exponential family with

$$h_\theta(y) = \exp \left(\beta_0 \sum_i y_i + \beta_1' \left(\sum_i \mathbf{x}_i y_i \right) + (\gamma/2) \sum_i y_i y_i^* \right)$$

so that

$$t(y) = \left(\sum_i y_i, \sum_i \mathbf{x}_i y_i, \frac{1}{2} \sum_i y_i y_i^* \right) \text{ and} \\ \theta = (\beta_0, \boldsymbol{\beta}_1, \gamma).$$

Note that

$$\frac{1}{2} \sum_i y_i y_i^* = \sum_{i < j} y_i y_j I(i \sim j).$$

See Cressie (1991) and Besag (1974) for more detail. Thus the MCMC method outlined in the previous paragraphs applies in our situation.

We shall generate our Monte Carlo sample Y_1, \dots, Y_n using the Gibbs sampler (Geman and Geman (1984)). A single iteration of the Gibbs sampler consists of visiting the m sites in some order and updating the value at each site using the conditional probability given in (1). Let G_0 be the map (a binary m -vector) giving the initial or starting values for the m sites, and let G_i be the map after i iterations of the Gibbs sampler. Our Monte Carlo sample consists of $Y_j = G_{b+j\Delta}$ for $j = 1, \dots, n$ where b is a suitably chosen ‘burn-in’ or ‘warm-up’ period, and Δ is a suitably chosen ‘spacing’.

The closer ψ is to the exact MLE $\hat{\theta}$, the smaller the Monte Carlo error is for a fixed Monte Carlo sample size n . The maximum pseudolikelihood estimate (MPLE) is often a good candidate for ψ ; it is easily computed using standard software for logistic regression.

3.2 MCMC Estimation of the Fisher Information

The first derivative of the log likelihood approximant is:

$$\frac{\partial l_n(\theta)}{\partial \theta} = T_{\text{obs}} - \frac{\sum_{j=1}^n T_j e^{(\theta-\psi)'T_j}}{\sum_{j=1}^n e^{(\theta-\psi)'T_j}} = -\frac{\sum_{j=1}^n u_j e^{(\theta-\psi)'u_j}}{\sum_{j=1}^n e^{(\theta-\psi)'u_j}}$$

where $u_j = T_j - T_{\text{obs}}$. The second derivative evaluated at the MCMC maximum likelihood estimate $\hat{\theta}_n$ is

$$\left. \frac{\partial^2 l_n(\theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}_n} = -\frac{\sum_{j=1}^n u_j u_j' e^{(\hat{\theta}_n-\psi)'u_j}}{\sum_{j=1}^n e^{(\hat{\theta}_n-\psi)'u_j}}.$$

In obtaining the above expression, we have used the fact that

$$\left. \frac{\partial l_n(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = 0.$$

Thus we can estimate the Fisher information evaluated at $\hat{\theta}_n$ by

$$\hat{I}_n = -\left. \frac{\partial^2 l_n(\theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}_n} = \frac{\sum_{j=1}^n u_j u_j' e^{(\hat{\theta}_n-\psi)'u_j}}{\sum_{j=1}^n e^{(\hat{\theta}_n-\psi)'u_j}}. \quad (3)$$

We obtain the Fisher information estimate (3) as a by-product of our parameter estimation program. It does not require any extra computational effort.

In our situation, the Markov chain obtained from the Gibbs sampler has a finite (although very large) state space consisting of the 2^m possible binary m -vectors. Moreover, there is a positive probability of going from any state to any other state in a single

iteration of the Gibbs sampler. This guarantees that our Markov chain is ergodic and also that the sample average of any quantity obeys a central limit theorem (see Billingsley (1968)). This allows us to apply the theory in Section 1.1 of Geyer and Thompson (1992) and Section 3 of Geyer (1994) to conclude that $\hat{\theta}_n \rightarrow \hat{\theta}$ and $\hat{I}_n \rightarrow I(\hat{\theta})$ in probability as $n \rightarrow \infty$. Here we use

$$I(\hat{\theta}) = - \left. \frac{\partial^2 l(\theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}}$$

(where $l(\theta)$ is the exact log-likelihood function) to denote the Fisher information evaluated at the exact MLE $\hat{\theta}$. (Note: our models belong to an exponential family, so the ‘observed’ and ‘expected’ Fisher information are the same.)

This discussion suggests using the diagonal elements of \hat{I}_n^{-1} to estimate the variances of the MCMC parameter estimates. We expect these variance estimates to be asymptotically valid, that is, they should be accurate if the Monte Carlo sample size n and the number of sites m are both sufficiently large. We study this issue via simulation in the next section.

Unfortunately, the general theory mentioned above does not give us any useful information on the rate of convergence. We must decide on empirical grounds if our sample size n is sufficiently large. The sample size should be chosen so that the Monte Carlo variability $\text{Cov}(\hat{\theta}_n - \hat{\theta})$ is small relative to the ‘true’ variability $\text{Cov}(\hat{\theta} - \theta)$. The ‘true’ variability is estimated by \hat{I}_n^{-1} . Using the results in Section 3 of Geyer (1994) we can estimate the Monte Carlo variability by

$$\text{Cov}(\hat{\theta}_n - \hat{\theta}) \approx C_n = \hat{I}_n^{-1} \hat{A} \hat{I}_n^{-1}. \quad (4)$$

Here \hat{A} is an estimate of the covariance matrix of $\bar{z} = n^{-1} \sum_i z_i$ where the series z_i is defined for $i = 1, 2, \dots, n$ by

$$z_i = \frac{u_i e^{(\hat{\theta}_n - \psi)' u_i}}{n^{-1} \sum_{j=1}^n e^{(\hat{\theta}_n - \psi)' u_j}}.$$

Such an estimate can be constructed by standard time series methods. Let $\hat{R}(s)$ be the sample autocovariance at lag s defined by

$$\hat{R}(s) = n^{-1} \sum_{i=1}^{n-s} (z_{i+s} - \bar{z})(z_i - \bar{z})'$$

for $0 \leq s \leq n - 1$ and by $\hat{R}(s) = \hat{R}(-s)'$ for $-(n - 1) \leq s < 0$. Now $\text{Cov}(\bar{z})$ can be estimated by the ‘window’ estimator

$$\hat{A} = n^{-1} \sum_{s=-(n-1)}^{n-1} \lambda(s) \hat{R}(s)$$

where $\lambda(s)$ is an appropriately chosen ‘lag window’. (See Priestley (1981), Ch. 6.) One simple choice is the ‘truncated periodogram’ window which gives an estimate of the form

$$\hat{A} = n^{-1} \sum_{s=-k}^k \hat{R}(s). \quad (5)$$

The value of k is chosen after examination of the autocovariances $\hat{R}(s)$ and is typically very small relative to n .

We will use this approach to estimate the Monte Carlo variability in Section 6.

4 Distributional Properties of MCMC Estimates

Theoretical study of the behavior of MCMC estimates is difficult, so we shall approach this problem by conducting simulation studies. In our simulations we shall use the autologistic regression model (1) on a complete 40×40 lattice; the number m of sites is 1600. When computing the neighborhood sums y_i^* , the corner sites have two neighbors, and the sites along the edges have three neighbors. We shall restrict consideration to a single covariate ($p = 1$) so that η_i in (1) can be written simply as $\eta_i = \beta_0 + \beta_1 x_i + \gamma y_i^*$. This covariate has the form of a diagonal sine wave; for a site i with coordinates (k, ℓ) , the covariate takes on the value

$$x_i = 2.5 \times \sin(0.1 \times (k + \ell)).$$

For each set of true parameter values $(\beta_0, \beta_1, \gamma)$, 500 sets of the pseudo-observed data were generated by a Gibbs sampler; the 500 data sets are obtained from 500 independent Markov chains. For each pseudo-observed data set, MCMC estimates of the parameters are calculated. So we have 500 independent MCMC estimates for each set of true parameter values, which may be examined for their bias, variance, and

other distributional properties. We shall use this same simulated data throughout this section.

Our MCMC estimates were computed using a sample size of $n = 1000$ generated from a Gibbs sampler started from the “zero-state”, i.e., the initial value is $y_i = 0$ for all sites i . We used a burn-in or warm-up period of $b = 100$ and a spacing of $\Delta = 2$. These values of n , b and Δ are chosen on the basis of our experience and the suggestions of Geyer and Thompson (1992) and Geyer (1991a, 1992). Our computational work was done using an S-PLUS interface to FORTRAN on SUN SPARC Stations.

For generating our Monte Carlo sample, we take ψ to be the maximum pseudolikelihood estimate (MPLE) obtained by fitting a logistic regression model for the response y_i on the covariates x_i and y_i^* . Occasionally, the MPLE turned out to be a bad value for ψ and led to a Monte Carlo sample $T_1, T_2, \dots, T_{1000}$ which did not contain the pseudo-observed value T_{obs} within its convex hull. For these samples the MCMC MLE does not exist, and attempting to solve (2) by Newton-Raphson iteration leads to a sequence of estimates $\theta_1, \theta_2, \theta_3, \dots$ which drift off to infinity. When this happened, we took the first step θ_1 of the Newton-Raphson iteration as a new value for ψ and generated another Monte Carlo sample using this new value. This somewhat ad hoc adjustment worked (at least for the range of parameter values reported in our tables) and produced samples which led to well-defined MCMC MLE’s.

Our choice of parameter values was motivated partly by the species data we wish to model. In most of the species we examined, the covariate effects and the spatial interaction were both quite large. So in these simulations we have taken $\beta_1 = 2$ which produces strong covariate effects. (We have also run simulations using other values of β_1 and the results were qualitatively similar.) We use a wide range of γ values. In applications γ is generally positive, but out of curiosity we have also examined negative values of γ and have given results for one of these ($\gamma = -1.5$) in our tables.

We shall examine only univariate aspects of the sampling distribution of the MCMC estimates, that is, we shall examine each parameter estimate separately. Studying the multivariate behavior of the MCMC estimates is a more complicated problem we shall not undertake here. In this subsection we shall mainly be concerned with showing

that the estimates are approximately normally distributed with the correct mean. The next subsection examines the behavior of the estimated Fisher information and the approximate confidence intervals which may be formed using it.

First we shall examine the following quantities for each sample of 500 estimates. *Bias*, the difference between the estimator sample mean and the known true value of the parameter; *skewness*, the skewness coefficient $g_1 = m_3/m_2^{3/2}$; *kurtosis*, the excess kurtosis coefficient $g_2 = m_4/m_2^2 - 3$. Here m_2, m_3 , and m_4 are the estimator sample moments computed about the mean. To test whether the estimator is normally distributed, the coefficients of skewness and excess kurtosis may be compared with their null distributions under sampling from a normal distribution: When the simulation sample size, N , is big enough, g_1 and g_2 are approximately normally distributed with mean zero and standard deviation $(6/N)^{1/2}$ and $(24/N)^{1/2}$ respectively (Ratkowsky, 1983, p. 25–26). In our simulation study, $N = 500$, and thus the standard deviations of g_1 and g_2 are 0.11 and 0.22, respectively.

Place of Figure 2

Place of Tables 1, 2 and 3

The sampling distributions of $\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}$ are summarized in Tables 1, 2, and 3. These tables show that the estimation error (both bias and variance) and the skewness increase as the spatial interaction increases. There is no such apparent trend for the kurtosis coefficients. We display Q-Q normal plots of the three parameter estimates in two cases,

one with $\gamma = 0.2$ and the other with $\gamma = 0.8$ representing a small and moderately large value of γ respectively. The parameters (β_0, β_1) were equal to $(1.0, 2.0)$ in both cases. These plots show that our MCMC estimators are approximately normally distributed, although some skewness and kurtosis exist. In summary, in our simulations the normal approximation to the distribution of the MCMC estimates appears reasonable so long as the spatial interaction γ is not too large.

In our simulation study for fixed finite sample size, a number of things start to go wrong when the value of γ becomes large. First, as γ increases, the correlation between successive iterations of the Gibbs sampler increases. Eventually, this correlation becomes large enough so that a sample size of 1000 with the “warm-up” and “spacing” values chosen in our simulations is no longer appropriate. (We need to either enlarge our sample size or increase the warm-up and spacing values.) Second, when γ exceeds some critical value a phase transition occurs and the realizations of the Gibbs sampler tend toward states in which nearly all of the values y_i are the same. Third, when γ is large, the maximum pseudolikelihood estimate performs poorly and is no longer suitable for us to use as the value ψ for generating our Monte Carlo samples. Because of these reasons, in our simulations the performance of the MCMC MLE deteriorates as γ increases. We have tried values of γ larger than those we report in our tables and the deterioration in performance continues and accelerates. The first and third problems mentioned above can be dealt with by fine-tuning our MCMC procedure and improving our method for choosing ψ . Thus, it may be possible to improve the performance for large γ to some degree.

The range of γ values for which the MCMC MLE performs well increases with the value of β_1 . One reason for this may be that the critical value of γ at which a phase transition occurs increases with the value of β_1 . As one indication of this trend we give Table 4. We ran a number of simulations (similar to those described above) holding β_0 fixed at 1.0 and varying the values of β_1 and γ . When γ was sufficiently large, the MCMC MLE would fail to exist for many of the Monte Carlo samples, that is to say the procedure described earlier, using the MPLE or the first step θ_1 of the Newton-Raphson iteration as ψ , did not produce a Monte Carlo sample for which the MCMC

MLE was well-defined. For each value of β_1 that we used, Table 4 records the largest value γ for which the MCMC MLE existed for all samples. This largest value increases steadily with β_1 .

Place of Table 4

5 Asymptotic Variance and Confidence Intervals

For a maximum likelihood estimate, the inverse of the Fisher information matrix may be used as a rough estimate of the covariance matrix of the parameter estimates. In many situations this estimate can be shown to be asymptotically valid, that is, it becomes progressively more accurate as the “sample size” or “amount of data” increases. For the autologistic regression model, the MCMC estimate of the Fisher information matrix is given in equation (3). We shall use the diagonal elements of the inverse of this matrix to estimate the variances of our parameter estimates. We shall call this type of variance estimate the *asymptotic variance*. For each set of parameter values used in our simulation study, we generated 500 data sets and computed MCMC estimates for each of them. For each parameter, the sample variance of the 500 estimates supplies us with a value we shall refer to as the *empirical variance*. The empirical variance will be a fairly accurate estimate of the true variance of our parameter estimates. In this subsection we shall compare the asymptotic and empirical variances.

Table 5 summarizes the asymptotic variance and empirical variance for the different parameter cases. It shows that the mean of the asymptotic variance estimates is usually very close to the empirical variance. The standard deviation of the asymptotic variance estimates is relatively small when γ is small, but increases steadily with the value of γ . Thus the asymptotic variance seems to be a roughly unbiased estimator of the actual variance, but one that is highly variable for larger values of γ . In spite of this difficulty at large values of γ , the confidence intervals (described below) based on the asymptotic

variance estimates work very well even for larger values of γ .

Place of Table 5

Let $\hat{\theta}_i$ be the estimate of some parameter θ , and $\hat{\sigma}_i^2$ be the asymptotic variance estimate for that parameter obtained from the i^{th} data set in one of our simulation cases. Let $r_i = (\hat{\theta}_i - \theta)/\hat{\sigma}_i$. We call r_i the standardized estimate. We have examined the normality of r_i and found that the distribution of r_i is closer to normal (in fact, $N(0, 1)$) than that of the original estimate $\hat{\theta}_i$. The sampling distribution of the standardized estimates in the two cases $(\beta_0, \beta_1, \gamma) = (1, 2, 0.2)$ and $(\beta_0, \beta_1, \gamma) = (1, 2, 0.8)$ is summarized in Table 6. The corresponding Q-Q normal plots are shown in Figure 3.

Place of Figure 3

Place of Table 6

We may form confidence intervals for each parameter based on the approximate $N(0, 1)$ distribution of the values $r_i = (\hat{\theta}_i - \theta)/\hat{\sigma}_i$, i.e., we take the $(1 - \alpha)$ confidence interval of θ to be $\hat{\theta}_i \pm z_{\alpha/2}\hat{\sigma}_i$, where $\hat{\theta}_i$ is the MCMC estimate, $z_{\alpha/2}$ is the normal quantile corresponding to the $(1 - \alpha)$ confidence level, and $\hat{\sigma}_i$ is the asymptotic standard deviation estimate.

For each parameter case, we have generated 500 data sets and obtained the 500 corresponding sets of parameter estimates and confidence intervals. Table 7 provides the actual coverage probabilities of the 99%, 95%, 90%, 70%, 50%, 30% and 10% confidence intervals. The smaller confidence levels are included only to test the normal approximation. In order to test that the coverage probabilities are equal to the nominal confidence level, we have used the exact Binomial two-sided test. For the level of significance $\alpha = 0.05$, the critical values for the various confidence levels are: 99%, (0.980, 0.998); 95%, (0.930, 0.969); 90%, (0.872, 0.926); 70%, (0.659, 0.740); 50%, (0.456, 0.544); 30%, (0.260, 0.341); 10%, (0.074, 0.128).

Place of Table 7

Table 7 shows that the actual coverage probabilities of our confidence intervals are close to the nominal confidence levels. In all cases, the difference between the actual coverage probability and the nominal confidence level is insignificant at the level $\alpha = 0.05$. Since these excellent results hold for a broad range of values for γ , we are optimistic that these confidence intervals will be reliable in most applications of the autologistic regression model.

Some interesting research which in many ways parallels our own was carried out by Graham (1994). His work deals with hypothesis testing for a variety of autologistic models: he considers both bi-directional (asymmetric) and second-order autologistic models in addition to the simple first-order autologistic models used in the present paper. However, Graham (1994) does *not* include covariates in his models. Also, Graham studies hypothesis testing whereas our paper concentrates on the properties of point estimates and confidence intervals. Much of the emphasis in both our paper and Graham (1994) is placed on showing that standard likelihood-based procedures have their usual properties when applied to the autologistic model with a Monte Carlo

approximation of the likelihood function. In particular, Graham carries out simulations to verify that both the Monte Carlo likelihood ratio test statistic and the Monte Carlo Wald test statistic have their usual asymptotic chi-squared distributions. Other similarities include the following: both papers use Monte Carlo estimates of the Fisher information matrix, and both papers use the MPLE as the parameter ψ for generating the Monte Carlo samples used to approximate the likelihood function.

6 Application to the Distribution of Plant Species

To illustrate how the MCMC method can be used in the Florida plant species data, we shall use the distribution data for plant species No. 38 (*Castanea pumila*) introduced in Section 1. The other species can be modeled in a similar way. For this analysis, we have “digitized” the map of Florida so that it is represented as a collection of 1845 sites forming a subset of a rectangular 68×80 lattice. These 1845 sites form the set \mathcal{S} of Section 2. At each site i we record $y_i = 1$ or 0 according to whether the species was present or absent at this site.

For purposes of illustration, we shall restrict ourselves here to autologistic regression models which involve only a single climate variable. (The bootstrap study in Section 6.2 and some of the other simulation work in this section becomes much more difficult with more variables.) After fitting a variety of models, we found that the distribution of species No. 38 could be explained fairly well in terms of the single climate variable FZF. For now, we shall simply assume that a model using FZF is appropriate.

We selected FZF in an informal manner, but more formal methods are available to us. The estimated log-likelihood in (1) can be used to compute Monte Carlo approximations to any of the standard test statistics based on the log-likelihood function or its derivatives (such as the likelihood ratio statistic, Rao’s score statistic, or Wald’s statistic). Using these Monte Carlo approximations, we could emulate the variable selection methods which are used for generalized linear models. One drawback of this approach is that it can be very time consuming; each test may require simulating one or more Monte Carlo samples. Selecting the ‘best’ of our 9 climate variables in this way

would take a lot of time; an ‘all possible subsets’ search is probably out of the question. We are working on faster methods of variable selection based on the pseudo-likelihood function. A careful consideration of model selection and other related issues will be given in another paper (Huffer and Wu, 1996).

6.1 Markov Chain Monte Carlo Estimates

In obtaining our maximum likelihood estimates, we implemented the MCMC as follows: We start the Gibbs sampler from the “observed state” (taking the observed map for species No. 38 as the starting state) and take the warm-up period to be 100 and the spacing to be 2. We used a sample size of $n = 2000$. For this data, the MPLE is not a good choice for ψ . We obtained our final choice for ψ by a process of successive modification beginning with the MPLE, that is, we generated a sequence of Monte Carlo samples and used each sample to suggest a new value of ψ to be used for the next sample. The new value for ψ was typically the first step θ_1 in the Newton-Raphson iterations for solving (2), the same procedure we used in our simulation studies. This approach to choosing ψ is fairly ad hoc; a more systematic approach is the iterative approach with constrained step sizes suggested by Geyer and Thompson (1992). Our results are reported in Table 8. The first line of the table gives the MCMC estimates. The second line gives asymptotic standard errors which are the square roots of the diagonal elements of \hat{I}_n^{-1} (see equation (3)). The last line contains estimates of the Monte Carlo (MC) standard errors; these are the square roots of the diagonal elements of C_n in (4). The matrix C_n was computed using $k = 20$ in (5). We chose $k = 20$ because the sample autocovariances $\hat{R}(s)$ had decayed to negligible values by lag 20.

Place of Table 8

The estimates given in Table 8 are probably good enough for most practical pur-

poses; the Monte Carlo errors are small relative to the standard errors. If more accurate estimates are desired, a larger sample size n can be used; the standard deviation of the Monte Carlo errors will be proportional to $n^{-1/2}$. Taking the estimates at face value for the moment, we see, by comparing the estimates in line 1 with the asymptotic standard errors in line 2, that both the climate and spatial correlation effects are highly significant.

After obtaining the estimates in Table 8, we performed various checks to gain confidence in the validity of the MCMC results. First, we simulated a very large sample ($n = 20,000$) using a Gibbs sampler started from the “observed state” with ψ set equal to the estimates in Table 8. Time series plots of the sufficient statistics T_1, T_2, \dots, T_n appeared to be stationary, and the corresponding autocorrelations decayed rapidly to zero, reaching negligible values before lag 20. (Since the spacing is 2, “lag 20” means 40 Gibbs sampler iterations.) We also computed MCMC MLEs using this very large sample; they were close to the values given in Table 8. Then we simulated a number of samples starting the Gibbs sampler from radically different initial states, again with ψ set equal to the estimates in Table 8. Three different starting states we used were (a) $y_i = 0$ for all i , (b) $y_i = 1$ for all i , and (c) the reverse of the “observed state” obtained by interchanging 1’s and 0’s. In all these cases, the time series plots of the sufficient statistics were indistinguishable from plots started from the “observed state” well before 100 iterations of the Gibbs sampler. We concluded that our Markov chain “mixes” fairly rapidly. The warm-up of $b = 100$ is more than adequate, and samples of moderate size like $n = 2000$ should lead to reasonably accurate estimates.

Next, we wished to verify the accuracy of the Monte Carlo standard errors displayed in Table 8. We carried out the following simulation study. We generated 100 samples of size $n = 2000$. Each sample was produced by a Gibbs sampler started from the “observed state” with a spacing of 2, warm-up of 100, and with ψ set equal to the estimates in Table 8. We used $k = 20$ in (5) to compute the Monte Carlo standard errors. For each of these 100 samples we computed the same quantities we display in Table 8. We summarize our results by reporting the means and standard deviations in Table 9. Since the true θ is unknown, we do not know what the ‘correct’ values are for

line 1. However, we see by comparing lines 2 and 5 (noting the approximate equality), and then noting that the values in line 6 are small relative to those in line 5, that our estimated Monte Carlo standard errors are fairly good indicators of the true variability in the MCMC estimates. Also, since the values in line 4 are very small relative to those in line 3, we see that our MCMC estimates of the asymptotic standard errors are quite accurate.

Place of Table 9

The remarks in the previous paragraphs say nothing about how well our model (with the parameter estimates in Table 8) actually fits the data. We can gain some idea about the adequacy of the estimated model by using the Gibbs sampler to generate random samples of maps from this model and then comparing these generated maps with the actual species map. An overall comparison can be obtained by using the model to compute fitted probabilities for each site and comparing these fitted values with the actual data values (0 or 1) at each site. The fitted probabilities are estimated using one long run of our Markov chain, that is, we use the Gibbs sampler to generate N samples (maps) and take the proportion of “1’s” among the N values for a given site to be the fitted probability for that site. We have used our model to compute fitted probabilities for species No. 38. A gray scale plot of these values is presented in Figure 4. We can see that this closely resembles the observed distribution map in Section 1.

Place of Figure 4

Wu (1994) and Wu and Huffer (1996) compared the MCMC MLEs with estimates

obtained using two other methods: the coding method and maximum pseudo-likelihood method. The fitted probabilities produced by the MCMC estimates were closer to the actual data values than those produced by the other two methods. More precisely, the MCMC estimates were best in the sense of minimizing the sum of the absolute fitting errors.

6.2 Bootstrap Study Using Gibbs Sampler

We would like to use the asymptotic standard errors reported in Table 8 to form confidence intervals for β_0 , β_1 , and γ as we did Section 5. Would such intervals be (at least approximately) valid in our current situation? To address this question we conducted a parametric bootstrap study. We used a Gibbs sampler with $\psi = (10.467, -0.0516, 2.5620)$ (these values are very close to the estimates in Table 8) to generate 500 independent maps. Using the procedures described earlier in Section 4, we succeeded in obtaining MCMC estimates for 497 of these samples. These 497 estimates were normalized as in Section 5. The distribution of these normalized estimates is summarized in Table 10.

Place of Table 10

Table 10 shows that the normalized MCMC estimates have approximately a standard normal distribution. This suggests that interval estimates based on the estimated Fisher information will be fairly reliable in this situation.

6.3 Evaluation of the Impact of Climate Change

The proposed model with the estimated parameters can be used to predict the impact of climate change (such as the global warming effect). In this subsection, we illustrate this idea by evaluating the impact of changes in the climate variable FZF on the distribution

of species No. 38. We consider two scenarios: (a) the values of FZF increase by 10% at each site, and (b) the values of FZF decrease by 10%. (In scenario (a), some values of FZF must be truncated so as not to exceed 365.) We forecast the probabilities of the presence of the species at each site using a Gibbs sampler and the parameter estimates in Table 8. The procedure is similar to the one which is used to create Figure 4 in Subsection 6.1 (see Wu and Huffer, 1996; and Wu, 1994 for more details.) The gray scale plots of the forecasted probability map are presented in Figure 5. In scenario (a), the distribution of species No. 38 will be reduced about 20% (in the sense of total area); while in scenario (b), the distribution of the species will increase about 30%.

Place of Figure 5

7 Conclusion and Discussion

We proposed the autologistic regression model for modeling the distribution of plant species in terms of climate covariates. Since the spatial correlation is explicitly included in the autologistic model, this model is more adequate and reasonable than models such as logistic regression which ignore the spatial correlation. (See the discussion in Section 1 of Wu and Huffer (1996).) However the parameters in the model cannot easily be estimated due to the intractable likelihood function. Fortunately the MCMC method, which has been well developed recently and is now applied in many fields, may be used to tackle this problem. The MCMC procedure is derived under our model setup. Some properties of MCMC MLE are studied via simulation. The application of the method is illustrated by constructing a simple model for a particular plant species. We used this model to predict changes in the species distribution with changes in the climate. In this paper we focused on the investigation of the MCMC method in this application. Some other important problems, such as model selection, spatial correlation testing and model checking, will be addressed elsewhere (Wu, 1994; Huffer

and Wu, 1996).

References

- Austin, M. P., Nicholls, A. O., and Margules, C. R. (1990), Measurement of the realized qualitative niche: environmental niches of five eucalyptus species. *Ecological Monographs*, 60(2), 161-177.
- Bartlein, P. J., Prentice, I. C., and Webb, T. (1986). Climatic response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography*, 13, 35-57.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society (with Discussion)*, Series B, 36, 192–236.
- (1975). Statistical analysis of non-lattice data. *The Statistician*, 24, 179–195.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Box, E. O., Crumpacker, D. W., and Hardin E. D. (1993). A Climatic model for location of plant species in Florida, U.S.A.. *Journal of Biogeography*, 20, 629-644.
- Cressie, N. (1991). *Statistics for Spatial Data*, New York: John Wiley.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Geyer, C. J. (1991a). Markov chain Monte Carlo maximum likelihood. *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (E.M. Keramides, ed.), 156–163.
- (1991b). Reweighting Monte Carlo mixtures. *Technical Report No.568*, School of Statistics, University of Minnesota.

- (1992). Practical Markov chain Monte Carlo (with discussion). *Statistical Science*, Vol.7, No.4, 473–511.
- (1994). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society, Ser.B* **56** 261–274.
- Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society, Ser.B* **54** 657–699.
- Graham, J. (1994). Monte Carlo Markov chain likelihood ratio test and Wald test for binary spatial lattice data. *Technical Report, Dept. of Statistics, North Carolina State University, Raleigh, NC.*
- Gumpertz, M.L., Graham, J.M., and Ristaino, J.B. (1997). Autologistic model of spatial pattern of phytophthora epidemic in bell pepper: effects of soil variables on disease presence. *Journal of Agricultural, Biological, and Environmental Statistics*, Vol.2, No.2, 131–156.
- Huffer, F.W. and Wu, H. (1996). Model selection in autologistic regression models with applications. Paper in preparation.
- Huntley, B., Bartlein, P. J., and Prentice, I. C. (1989). Climatic control of the distribution and abundance of beech (*Fagus L.*) in Europe and north America. *Journal of Biogeography*, 16, 551-560.
- Little, Jr., E. L. (1978). *Atlas of United States Trees*, Volume 5. Florida. Misc. Publ. No. 1361, USDA Forest Service. Washington, D. C.: U.S. Government Printing Office. 256 maps, with indices of common and scientific names.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, London: Chapman and Hall.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.
- Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling*, Marcel Dekker, Inc.

- Wu, H. (1994). *Regression Models for Spatial Binary Data with Application to the Distribution of Plant Species*, Ph.D. Dissertation, Department of Statistics, Florida State University.
- Wu, H. and Huffer, F.W. (1997). Modeling the distribution of plant species using the autologistic regression model. *Environmental and Ecological Statistics*, 4, 49–64.
- Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika*, 77, 642–648.

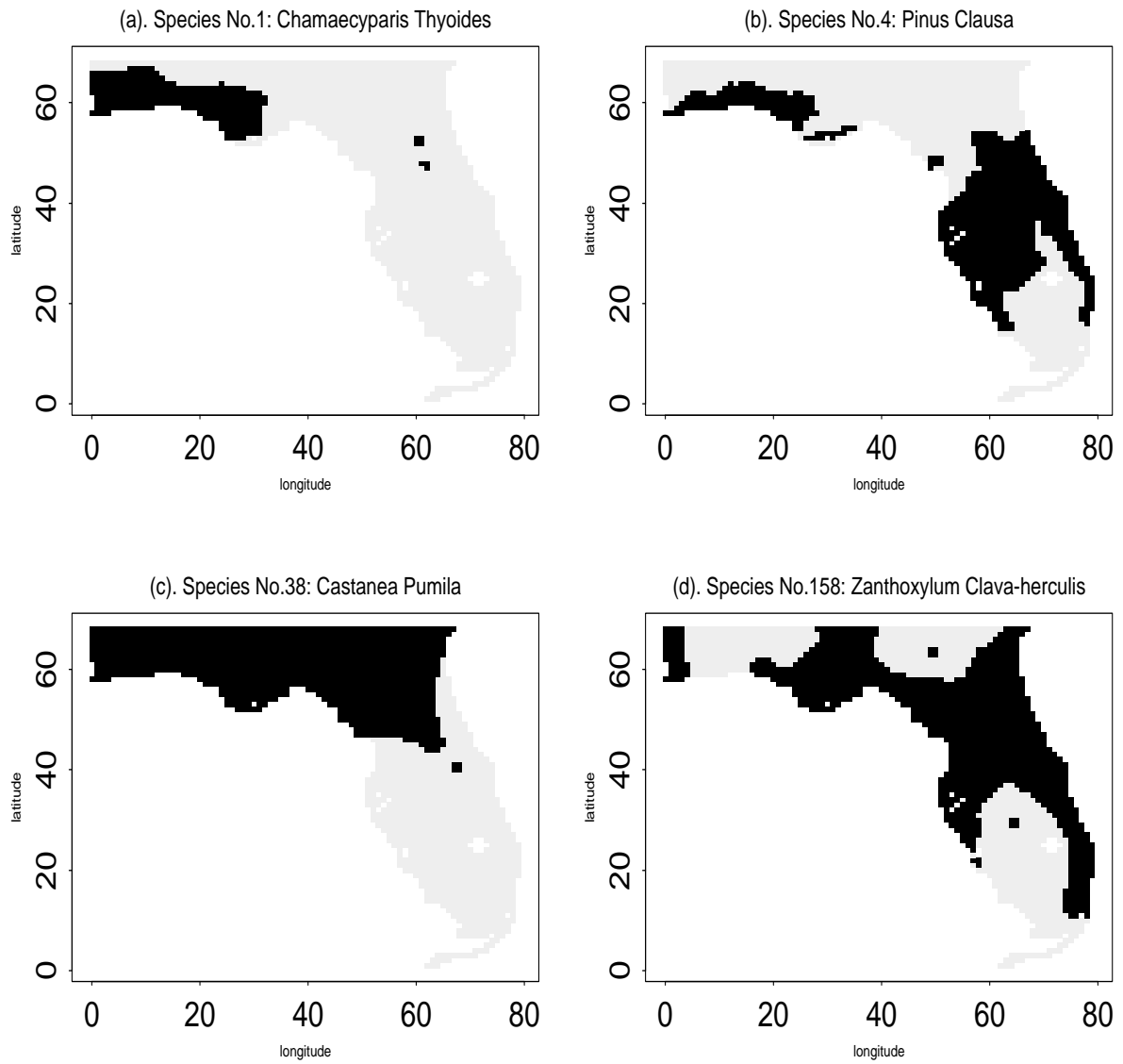


Figure 1: Distribution maps of plant species

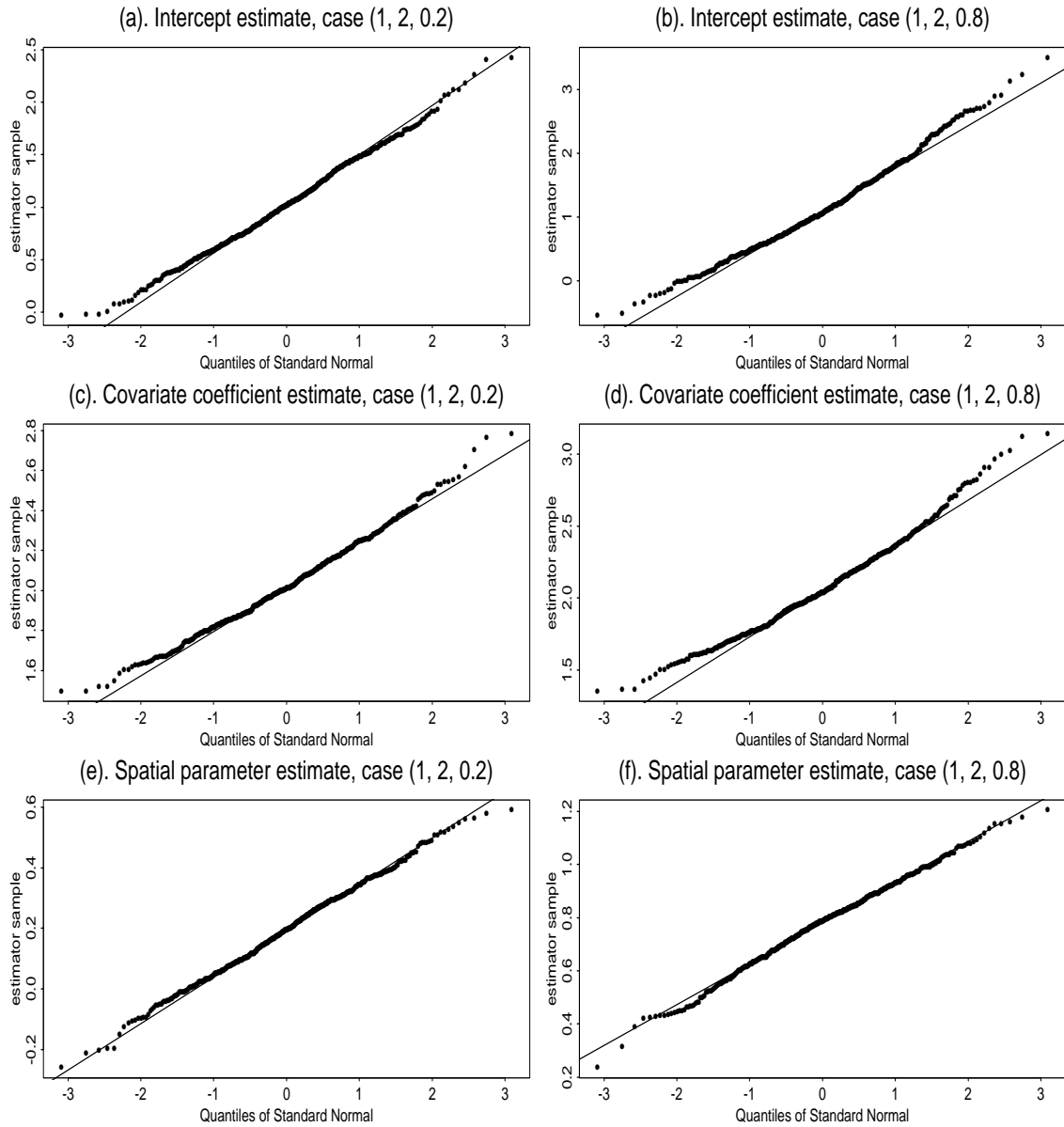


Figure 2: Q-Q normal plots of the MCMC estimators

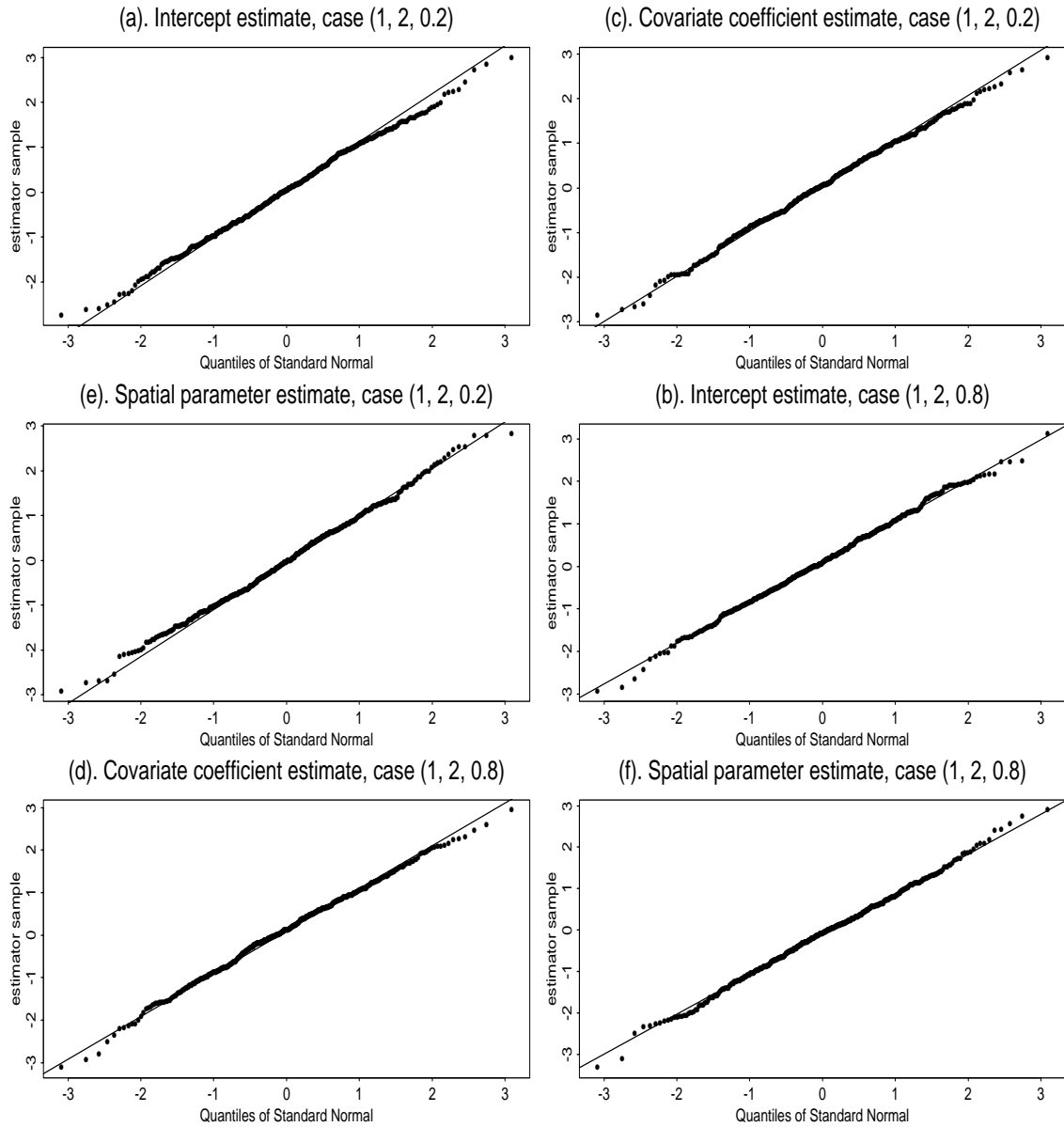


Figure 3: Q-Q normal plots of the standardized estimates

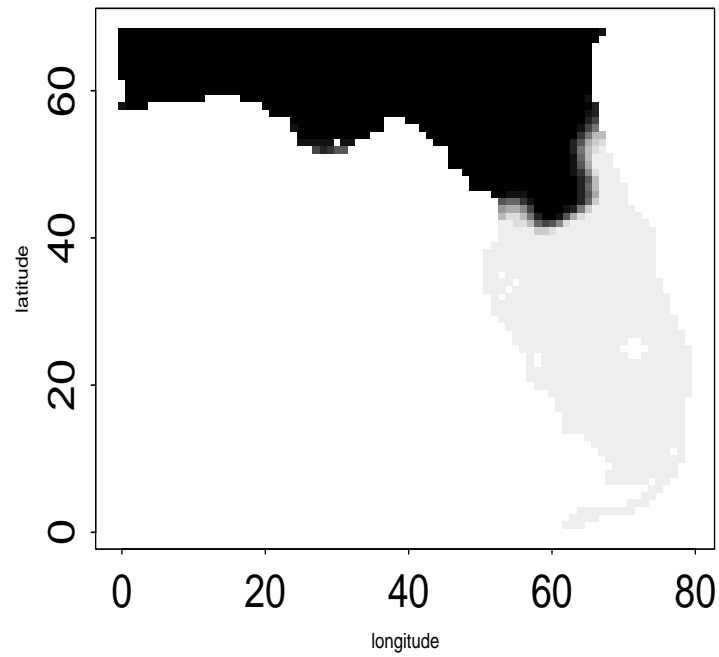
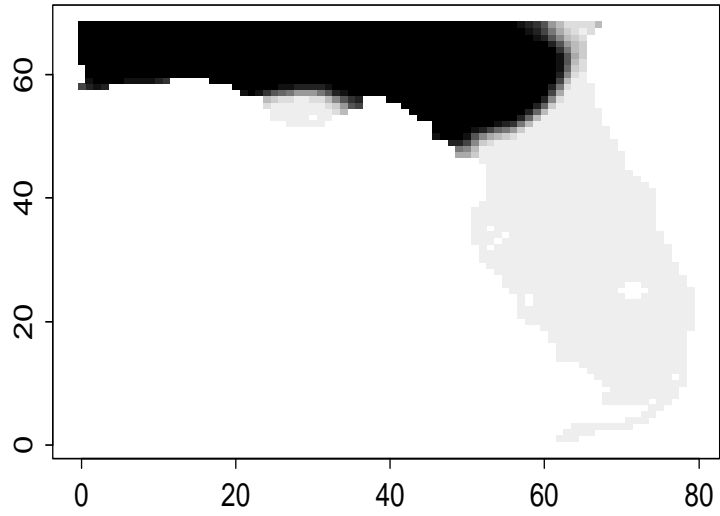


Figure 4: Gray-scale plot of the fitted probabilities for Species No. 38

(a) Impact of 10% increase of FZF



(b) Impact of 10% decrease of FZF

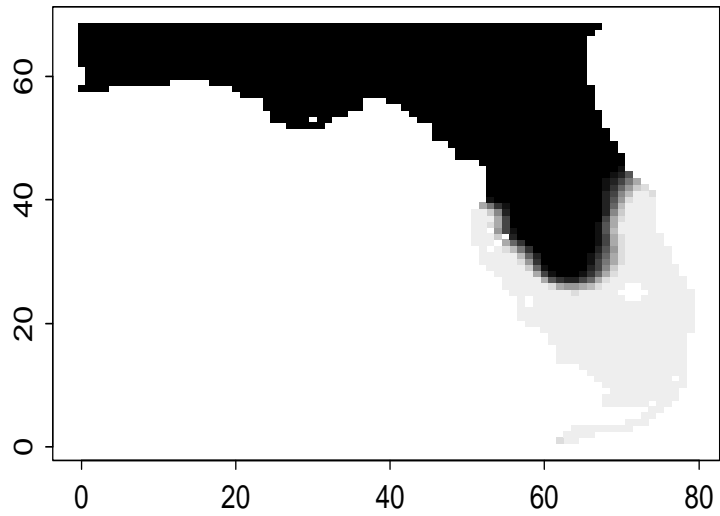


Figure 5: Gray-scale plots of the impact of FZF on the distribution of Species No. 38

Table 1: Sampling distribution of $\hat{\beta}_0$

True $(\beta_0, \beta_1, \gamma)$	Mean	S.D.	Bias	Skewness (g_1)	Kurtosis (g_2)
(1, 2, -1.5)	1.0063	0.1551	0.0063	0.1318	-0.1748
(1, 2, 0.0)	1.0157	0.3679	0.0157	0.2370*	0.0400
(1, 2, 0.2)	1.0282	0.4346	0.0282	0.1827	-0.1132
(1, 2, 0.4)	1.0488	0.5064	0.0488	0.2824*	-0.0715
(1, 2, 0.6)	1.0463	0.5818	0.0463	0.3036*	-0.2154
(1, 2, 0.8)	1.1282	0.6757	0.1282	0.4348**	0.1161
(1, 2, 1.0) ¹	1.1821	0.8474	0.1821	0.6176**	1.3421**
(1, 2, 1.5)	1.2743	1.1532	0.2743	0.7468**	0.7793**

Table 2: Sampling distribution of $\hat{\beta}_1$

true $(\beta_0, \beta_1, \gamma)$	Mean	S.D.	Bias	Skewness (g_1)	Kurtosis (g_2)
(1, 2, -1.5)	2.0116	0.1079	0.0116	-0.0285	0.5295*
(1, 2, 0.0)	2.0183	0.1920	0.0183	0.3239*	0.5432*
(1, 2, 0.2)	2.0242	0.2170	0.0242	0.3307**	0.1626
(1, 2, 0.4)	2.0337	0.2452	0.0337	0.5288**	0.3732
(1, 2, 0.6)	2.0326	0.2736	0.0326	0.4070**	-0.0256
(1, 2, 0.8)	2.0679	0.3087	0.0679	0.5189**	0.3532
(1, 2, 1.0) ¹	2.0892	0.3746	0.0892	0.7865**	1.9230**
(1, 2, 1.5)	2.1225	0.4478	0.1225	0.7754**	0.7375**

Table 3: Sampling distribution of $\hat{\gamma}$

true $(\beta_0, \beta_1, \gamma)$	Mean	S.D.	Bias	Skewness (g_1)	Kurtosis (g_2)
(1, 2, -1.5)	-1.5059	0.0997	-0.0059	0.2682*	0.7506**
(1, 2, 0.0)	-0.0040	0.1386	-0.0040	-0.1024	-0.1545
(1, 2, 0.2)	0.1939	0.1478	-0.0061	-0.0243	-0.1448
(1, 2, 0.4)	0.3894	0.1530	-0.0106	-0.1066	-0.1122
(1, 2, 0.6)	0.5919	0.1553	-0.0081	-0.1682	-0.3673
(1, 2, 0.8)	0.7773	0.1556	-0.0227	-0.1892	0.0393
(1, 2, 1.0) ¹	0.9722	0.1681	-0.0278	-0.1893	0.0553
(1, 2, 1.5)	1.4845	0.1723	-0.0155	-0.2612*	0.6410*

¹ In this case, there is an outlier. These are the results after deleting the outlier.

* the value exceeds two standard deviations

** the value exceeds three standard deviations

Table 4: Largest value of γ for which MCMC MLE always existed

Coefficient β_1	0.1	0.5	1.0	1.5	2.0
Critical Value for γ	1.1	1.2	1.8	1.9	2.5

Table 5: Asymptotic variance and empirical variance

γ	-1.5	0.0	0.2	0.4	0.6	0.8	1.0	1.5
E.V.	0.024	0.135	0.189	0.256	0.338	0.457	0.718	1.330
$\hat{\beta}_0$ A.V.	0.027	0.145	0.191	0.256	0.345	0.479	0.658	1.310
S.D.	0.009	0.015	0.024	0.039	0.064	0.108	0.232	1.532
E.V.	0.012	0.037	0.047	0.060	0.075	0.095	0.140	0.200
$\hat{\beta}_1$ A.V.	0.012	0.039	0.049	0.061	0.077	0.099	0.126	0.204
S.D.	0.023	0.006	0.008	0.013	0.018	0.027	0.043	0.169
E.V.	0.010	0.019	0.022	0.023	0.024	0.024	0.028	0.030
$\hat{\gamma}$ A.V.	0.011	0.021	0.022	0.023	0.024	0.026	0.027	0.029
S.D.	0.027	0.001	0.002	0.002	0.003	0.004	0.014	0.061

Notes:

1. In all cases, the true values are $\beta_0 = 1, \beta_1 = 2$.
2. E.V. is the empirical variance.
3. A.V. is the average of the asymptotic variance estimates.
4. S.D. is the standard deviation of the asymptotic variance estimates.

Table 6: Sampling distribution of standardized estimates

Parameter Case	Estimate	Mean	S.D.	Skewness	Kurtosis
(1, 2, 0.2)	$\hat{\beta}_0$	0.0267	0.9941	-0.0393	-0.1893
	$\hat{\beta}_1$	0.0413	0.9769	-0.0991	-0.0593
	$\hat{\gamma}$	-0.0263	1.0073	0.0707	-0.0995
(1, 2, 0.8)	$\hat{\beta}_0$	0.1035	0.9706	-0.0409	-0.0152
	$\hat{\beta}_1$	0.1008	0.9749	-0.1816	0.0647
	$\hat{\gamma}$	-0.1147	0.9794	-0.0331	0.1495

Table 7: Actual coverage probability of confidence intervals

True	γ	-1.5	0.0	0.2	0.4	0.6	0.8	1.0	1.5
	99%	0.992	0.992	0.988	0.994	0.994	0.992	0.992	0.994
	95%	0.960	0.952	0.960	0.954	0.948	0.956	0.932	0.954
C.I.	90%	0.912	0.902	0.918	0.900	0.904	0.894	0.884	0.914
of	70%	0.696	0.732	0.694	0.680	0.694	0.724	0.677	0.700
β_0	50%	0.532	0.514	0.476	0.480	0.488	0.502	0.481	0.450
	30%	0.354	0.278	0.298	0.298	0.278	0.304	0.293	0.270
	10%	0.126	0.078	0.106	0.096	0.084	0.108	0.084	0.088
	99%	0.986	0.984	0.988	0.992	0.996	0.990	0.992	0.996
	95%	0.946	0.960	0.960	0.958	0.952	0.952	0.944	0.964
C.I.	90%	0.898	0.920	0.900	0.918	0.910	0.916	0.888	0.908
of	70%	0.706	0.738	0.716	0.690	0.706	0.716	0.671	0.690
β_1	50%	0.498	0.514	0.496	0.494	0.490	0.498	0.491	0.496
	30%	0.302	0.322	0.290	0.292	0.296	0.304	0.313	0.290
	10%	0.096	0.118	0.122	0.100	0.094	0.122	0.108	0.086
	99%	0.992	0.992	0.986	0.990	0.996	0.992	0.988	0.984
	95%	0.944	0.956	0.946	0.950	0.950	0.942	0.944	0.950
C.I.	90%	0.898	0.904	0.906	0.892	0.898	0.904	0.876	0.888
of	70%	0.708	0.710	0.696	0.686	0.688	0.708	0.685	0.686
γ	50%	0.504	0.500	0.478	0.484	0.502	0.518	0.497	0.468
	30%	0.318	0.310	0.266	0.296	0.300	0.330	0.269	0.296
	10%	0.114	0.100	0.088	0.104	0.114	0.118	0.080	0.100

Notes:

1. In all cases, the true values are $\beta_0 = 1, \beta_1 = 2$.
2. For case $\gamma = 1.0$, the sample size is 499. One observation is an outlier.

Table 8: Estimation results for species No. 38

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$
Estimate	10.385	-0.051348	2.5609
Asymp. S.E.	3.133	0.009894	0.2794
MC S.E.	0.103	0.000318	0.0081

Table 9: Summary of MCMC estimation results for 100 samples

Line No.			$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$
1	Estimate	Mean	10.393	-0.051331	2.5548
2		S.D.	0.117	0.000367	0.0080
3	Asymp. S.E.	Mean	3.116	0.009803	0.2785
4		S.D.	0.055	0.000178	0.0047
5	MC S.E.	Mean	0.109	0.000343	0.0081
6		S.D.	0.0086	0.000027	0.00085

Table 10: Sampling distribution of normalized estimates for Species No. 38

Estimate	Mean	S.D.	Skewness	Kurtosis
$\hat{\beta}_0$	0.1204	1.0294	-0.2226	0.1942
$\hat{\beta}_1$	-0.1214	1.0327	0.2691	0.1559
$\hat{\gamma}$	-0.0724	0.9545	-0.3976	0.2225