

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1996 - 8th Annual Conference Proceedings

MARKOV CHAIN MONTE CARLO METHODS FOR MODELING THE SPATIAL PATTERN OF DISEASE SPREAD IN BELL PEPPER

Jonathan M. Graham

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Graham, Jonathan M. (1996). "MARKOV CHAIN MONTE CARLO METHODS FOR MODELING THE SPATIAL PATTERN OF DISEASE SPREAD IN BELL PEPPER," *Conference on Applied Statistics in Agriculture*.
<https://doi.org/10.4148/2475-7772.1321>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

MARKOV CHAIN MONTE CARLO METHODS FOR MODELING THE SPATIAL PATTERN OF DISEASE SPREAD IN BELL PEPPER

Jonathan M. Graham
Department of Mathematical Sciences
University of Montana
Missoula, MT 59812

With exponential family models for dependent data, such as the autologistic model for binary spatial lattice data, maximum likelihood estimates can be obtained using Markov chain sampling methods by simulating an ergodic Markov chain which converges weakly to the equilibrium distribution of the model. This Markov chain Monte Carlo maximum likelihood (MCMCML) procedure provides a competitor to the usual pseudolikelihood estimation method often used for modeling discrete lattice data. Within this MCMCML framework, it is also possible to conduct formal inference using MCMC analogues to the usual likelihood ratio, Wald, and Lagrange multiplier tests, for which the asymptotic distributions are known subject to some mild regularity conditions. Here, the MCMC methodology will be discussed as it pertains to the autologistic model for binary data and will be used to model the spatial pattern of disease spread in bell pepper caused by the pathogen *Phytophthora capsici*.

Keywords: Autologistic model, Markov chain Monte Carlo (MCMC)

1. Introduction

Markov chain Monte Carlo (MCMC) methods for estimating parameters from distributions which have an intractable form have been used increasingly in applications in recent years. Examples include the analysis of disease incidence data in agricultural studies (Besag 1974, Gumpertz et. al. 1995), the distribution of plant species (Huffer & Wu 1995), genealogical studies (Geyer & Thompson 1994), and image restoration (Geman & Geman 1984). However, little attention has been given to the parallel development of formal inference procedures within this MCMC framework. Here, an application is considered which necessitates the use of hypothesis testing to address important physical questions about the data. Three MCMC inference procedures corresponding to the usual likelihood ratio test, Wald test, and Lagrange multiplier test, are developed and compared to other related procedures based on the pseudolikelihood estimation method, through simulation and an application to *Phytophthora* root and crown rot disease incidence data in bell pepper plants.

Data collected over a geographical region are often not independent. Hence, modeling spatial data should incorporate this spatial dependence. In some cases, as with binary data on a lattice or grid, a Markov random field (MRF) assumption on the dependence structure may be appropriate. Under such an assumption, the dependence structure of the data can be defined through interactions between groups of neighboring sites on the lattice. Because the realization of a random variable at a given site is dependent on the values at neighboring sites, this MRF assumption enables lattice data to be modeled in a natural way

through conditional probabilities, where the probability distribution for each site-random variable conditioned on the values at neighboring sites is specified. Under a properly defined neighborhood system, this conditional probability model has an equivalent joint specification known as the Gibbs distribution (Geman & Geman 1984). For binary data under an MRF assumption, a special type of Gibbs model known as the autologistic model is often used.

The impetus for much of the work in this paper originated from an applied problem in plant pathology, where interest lay in understanding the mechanism of disease spread in bell pepper plants caused by the pathogen *Phytophthora capsici* (Ristaino, et. al. 1993). To study this, the spatial patterns of spread were analyzed for disease incidence data collected from three North Carolina bell pepper fields, using autologistic models of different orders and directions of dependence. Viewing each field as a 20x20 lattice of sites, each consisting of 2 or 3 plants, a “0” or “1” is recorded at each site indicating whether the plants at that site are healthy or diseased respectively. In this way, binary spatial lattice data were collected at different times and fields over the 1992 growing season. An example of the data from one of these fields appears in Figure 1, where “•” and “o” indicate diseased and healthy sites, respectively. An initial inspection of this field (labeled Field 1 1992) reveals a greater degree of disease spread within rows than across rows.

The primary objective of this paper is to investigate the applicability of formal MCMC-based inference procedures, specifically to autologistic models in the spatial lattice data setting. Such procedures will enable the practitioner to test for differences in the direction and magnitude of disease spread in the *Phytophthora* data. Section 2 reviews the form of autologistic model and demonstrates its flexibility in modeling data with the different levels and directions of spatial dependence. Sections 3 and 4 provide the necessary background for the estimation and inference procedures respectively. After developing these inference methods and the corresponding asymptotic theory, two simulation studies comparing the adequacy and power of tests using these methods are summarized in Section 5. Finally, Section 6 presents an example using *Phytophthora capsici* disease incidence lattice data to illustrate the use of the inference procedures in performing model selection and for answering specific questions pertaining to the mechanism of disease spread.

2. Autologistic Model

Binary data, often collected as the presence or absence of some characteristic under study, are frequently analyzed using a logistic regression model. However, binary spatial lattice data may exhibit some form of spatial dependence, rendering the usual logistic model inappropriate. Under such a scenario, an alternative model of “logistic” form, where the site random variables are now regressed on themselves through their dependence on random variables at neighboring sites is termed the autologistic (AL) model. Besag (1974) formalized the notion of an AL model, establishing various properties of the model and demonstrating its flexibility in modeling different levels and directions of spatial dependence.

2.1 Background

Let $D = \{s_1, \dots, s_m\} \subset \mathbb{R}^d$, d a positive integer, be a lattice of $m (< \infty)$ sites on which

a realization of the spatial process $\mathbf{Z} = \{Z_1, \dots, Z_m\}$ is observed, and let $S_N = \{(s_i, N_i) : i = 1, \dots, m\}$ be a neighborhood system defined on D , where N_i is the set of all neighbors of site s_i , $i = 1, \dots, m$. A Markov random field (MRF) assumption on \mathbf{Z} with respect to S_N requires the form of the resulting conditional probability distribution to satisfy the spatial Markov Property: $Pr(Z_i = z_i | Z_j = z_j : j \neq i) = Pr(Z_i = z_i | Z_j = z_j : s_j \in N_i), \forall i, j$, namely that the conditional distribution of Z_i given all other z_j in the lattice depends only on the z_j at neighboring sites of s_i . Under the additional assumptions of positivity and pairwise-only site dependence, application of the Hammersley-Clifford Theorem necessarily yields the general form of the AL model for binary data as:

$$Pr(Z_i = z_i | \{z_j : j \neq i\}) = \frac{\exp\{\alpha_i z_i + \sum_{j=1}^m \beta_{ij} z_i z_j\}}{1 + \exp\{\alpha_i + \sum_{j=1}^m \beta_{ij} z_j\}}, z_i = 0, 1,$$

where $\boldsymbol{\theta} = (\{\alpha_i\}, i = 1, \dots, m; \{\beta_{ij}\}, i, j = 1, \dots, m)$ is the vector of model parameters and $\beta_{ii} = 0$ for identifiability (Besag 1974, Cressie 1993).

In this general framework, suppose now for simplicity that the data $\mathbf{z} = \{z_{jk} : j = 1, \dots, m_1, k = 1, \dots, m_2\}$ are collected on a two-dimensional lattice $D = \{(j, k) : j = 1, \dots, m_1, k = 1, \dots, m_2\}$. The further restrictions, $\alpha = \alpha_i$ and $\beta = \beta_{ij} \forall s_j \in N_i$, give rise to the simplest form of the AL model with first-order dependence structure known as the Ising model. The resulting first-order AL(α, β) or Ising model has the form:

$$Pr(Z_{jk} = 1 | z_{j'k'} : (j', k') \neq (j, k); \alpha, \beta) = \text{expit}(\alpha + \beta n_{jk}) = \frac{\exp\{\alpha + \beta n_{jk}\}}{1 + \exp\{\alpha + \beta n_{jk}\}},$$

$$\text{where: } n_{jk} = \sum_{(j', k') \in N(j, k)} z_{j'k'} = [z_{j-1, k} + z_{j+1, k} + z_{j, k-1} + z_{j, k+1}],$$

the number of nearest neighbors which are diseased. This model resembles the classic logistic regression model; however, the n_{jk} are not independent of the Z_{jk} . The corresponding joint (Gibbsian) distribution is given by: $Pr(\mathbf{Z} = \mathbf{z} | \alpha, \beta) = [c(\alpha, \beta)]^{-1} \exp\{\alpha S + \beta N\}$ where $S = \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} z_{jk}$, and $N = \frac{1}{2} \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} z_{jk} n_{jk}$ are the sufficient statistics for (α, β) , and $c(\alpha, \beta)$ is the normalizing constant. To avoid edge effects on a finite lattice, these sums are computed only over the inner lattice sites (those with a complete set of neighbors).

The β -parameter measures the strength of the spatial dependence in the data, and the α -parameter contains information about the number of one-realizations in the lattice. A unit increase in n_{jk} corresponds to an increase by β in the log odds of observing a one to a zero at a particular site, given the values at neighboring sites. In terms of the *Phytophthora* data, this translates to an increase by β in the log odds of disease at a given site for each additional neighboring site with diseased plants. Larger values of β indicate stronger spatial dependence between neighboring sites, with $\beta = 0$ giving the independent logistic regression model. The correlation between sites generally decreases with distance; however, there is a critical point for β at $\beta_c = 2 \sinh^{-1}(1) \approx 1.76$ where for $\beta > \beta_c$ and $\alpha = -2\beta$, the model exhibits long-range correlation yielding two completely different types of realizations from the same model parameters (Pickard 1987). In models with high dependence, this

can cause severe problems with estimation, as the estimation methods are highly sensitive to small changes in the data. Pickard (1987) explores the effects of this critical value on estimation and inference for the Ising model, showing that the underlying MRF can exhibit phase transitions and long-range correlation, leading to identifiability problems. Gidas (1993) concludes that under an identifiability assumption, consistency of the parameter estimates for points beyond the critical value is achieved, whereas asymptotic normality is not.

2.2 Model Hierarchy

There are two natural extensions of this first-order Ising model to higher orders and different directions of dependence respectively. For the former, the obvious extension is to a second-order model which includes diagonally adjacent sites as neighbors in addition to the first-order neighbors. This second-order AL model, denoted $AL(\alpha, \beta, \delta)$, is given by:

$$Pr(Z_{jk} = z_{jk} \mid z_{j',k'} : (j', k') \neq (j, k); \alpha, \beta, \delta) = \frac{\exp\{z_{jk}(\alpha + \beta n_{jk} + \delta d_{jk})\}}{1 + \exp\{\alpha + \beta n_{jk} + \delta d_{jk}\}} \text{ where:}$$

n_{jk} and d_{jk} represent the number of adjacent (horizontal and vertical) and diagonal nearest neighbors with a value of one, respectively. Here, β describes the dependence between horizontally or vertically adjacent neighbors, and δ the dependence between diagonally adjacent neighbors. Extensions to higher-order dependence models might be useful with the *Phytophthora* data in examining the extent of spatial dependence between pepper plants at varying distances of separation. The corresponding joint specification is: $Pr(\mathbf{Z} = \mathbf{z} \mid \alpha, \beta, \delta) = [c(\alpha, \beta, \delta)]^{-1} \exp\{\alpha S + \beta N + \delta D\}$, where $S = \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} z_{jk}$, $N = \frac{1}{2} \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} z_{jk} n_{jk}$, and $D = \frac{1}{2} \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} z_{jk} d_{jk}$ are the sufficient statistics for (α, β, δ) and $c(\alpha, \beta, \delta)$ is the normalizing constant.

The second extension of the first-order Ising model is to split the dependence parameter β into two parameters to incorporate two possible directions of dependence. This bi-directional first-order AL model, denoted $AL(\alpha, \beta_1, \beta_2)$, has the form:

$$Pr(Z_{jk} = z_{jk} \mid z_{j',k'} : (j', k') \neq (j, k); \alpha, \beta_1, \beta_2) = \frac{\exp\{z_{jk}(\alpha + \beta_1 n_{1jk} + \beta_2 n_{2jk})\}}{1 + \exp\{\alpha + \beta_1 n_{1jk} + \beta_2 n_{2jk}\}}$$

where: $n_{1jk} = [z_{j,k-1} + z_{j,k+1}]$, and $n_{2jk} = [z_{j-1,k} + z_{j+1,k}]$, the number of within-row and between-row nearest neighbors with a value of one respectively. β_1 represents the between-rows spatial interaction, and β_2 the within-row interaction. Such a distinction may be important if one is interested in investigating differences in the spread of disease between rows and within rows in the field. The corresponding joint specification is: $Pr(\mathbf{Z} = \mathbf{z} \mid \alpha, \beta_1, \beta_2) = [c(\alpha, \beta_1, \beta_2)]^{-1} \exp\{\alpha S + \beta_1 N_1 + \beta_2 N_2\}$, where $S = \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} z_{jk}$, $N_1 = \frac{1}{2} \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} z_{jk} n_{1jk}$, and $N_2 = \frac{1}{2} \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} z_{jk} n_{2jk}$ are the sufficient statistics for $(\alpha, \beta_1, \beta_2)$ and $c(\alpha, \beta_1, \beta_2)$ is the normalizing constant. The hierarchy inherent with these AL models enables the study of different levels and directions of dependence in the neighborhood structure through parameter estimation and hypothesis testing, so that model selection can be performed.

3. Estimation Methods

Parameter estimation methods in models of intractable form such as the AL model have received significant attention in the literature. Early methods such as coding (Besag 1974) and pseudolikelihood (Besag 1975) maximized certain functions of the *conditional* probability distribution. More recently, efforts to estimate the unwieldy normalizing constant present in the *joint* probability specification have led to a variety of approximate maximum likelihood estimation procedures, as summarized in Geyer & Thompson (1992). This latter paper is the origin of the Markov chain Monte Carlo (MCMC) maximum likelihood estimation procedure upon which the inference procedures developed here are based. MCMC methods have become increasingly popular in recent years with efforts focused primarily on Markov chain convergence rates and the development of more efficient estimation algorithms (Besag & Green 1993, Geyer & Thompson 1994).

This section provides a brief review of the pseudolikelihood (PL) and Markov chain Monte Carlo maximum likelihood (MCML) estimation methods, both of which will be used to develop inference procedures in a later section. Some discussion on the asymptotic properties of the resulting estimators is also given. Although asymptotic theory is typically studied on the basis of increasing sample sizes, it is here viewed with respect to increasing *lattice* sizes. Under the framework of Gidas (1993), let $D(m)$ denote the d -dimensional lattice of m sites, and consider a sequence of samples $\{\mathbf{Z}_m = \{Z_i : s_i \in D(m)\}$ for increasing m , and an associated expanding sequence of MRFs $\{\pi_m\}$ with common parameter vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. Throughout this paper, the limit as the lattice size $m = |D(m)| \rightarrow \infty$ will be interpreted in the manner given by van Love (Gidas 1993), namely that as $m \rightarrow \infty$, $|D(m') - D(m)|/|D(m)| \rightarrow 0$ for every m and m' where $m' > m$ is the number of sites in the next step of the sequence of increasing lattices.

3.1 Pseudolikelihood Method

Likelihood functions resulting from MRF models generally contain an intractable normalizing constant, preventing maximum likelihood estimation in even simple AL models. The pseudolikelihood (PL) estimation method attempts to circumvent this problem by maximizing the product of the conditional densities. The normalizing constant is absent from the conditional densities allowing this PL function to be maximized by standard numerical methods. This method was developed by Besag (1975) and a good general introduction to PL functions can be found in Strauss (1992).

For binary lattice data $\mathbf{Z} = \mathbf{z}$ on a lattice $D = \{(j, k) : j = 1, \dots, m_1, k = 1, \dots, m_2\}$ where \mathbf{Z} has a Gibbs distribution with parameter vector $\boldsymbol{\theta} \in \Theta$, the PL function is defined as the product over *all* sites in D of the conditional probability densities of the Z_{jk} given the values at the remaining sites:

$$PL(\boldsymbol{\theta} | \mathbf{z}) = \prod_{(j,k) \in D} Pr(Z_{jk} = z_{jk} | z_{j'k'}, (j', k') \neq (j, k)) = \prod_{(j,k) \in D} \frac{\exp\{z_{jk} \cdot \mathbf{t}'_{jk} \boldsymbol{\theta}\}}{1 + \exp\{\mathbf{t}'_{jk} \boldsymbol{\theta}\}},$$

in the context of the autologistic model, where \mathbf{t}_{jk} is the vector of sufficient statistics for $\boldsymbol{\theta}$ associated with the site (j, k) . For example, in the Ising model, $\mathbf{t}_{jk} = (1, n_{jk})$ and $\boldsymbol{\theta} = (\alpha, \beta)$

following earlier notation. The popularity of the PL method in practice stems from its ease of implementation and well-behaved concave PL functions, which can be numerically maximized to obtain the maximum pseudolikelihood estimator (MPLE) $\hat{\boldsymbol{\theta}}_{PL}$. Although maximization of this PL function yields consistent estimators (see Geman & Graffigne 1987 for a proof of consistency), it should be emphasized that the PL function is *not* a true likelihood, so that estimators obtained are not MLEs and hence not necessarily asymptotically efficient (Strauss 1992). Asymptotic normality of the MPLE, to this author's knowledge, has not been proven for increasing lattice sizes; and in fact, the asymptotic standard errors from the "pseudolikelihood information matrix" are valid only in the case of spatial independence.

3.3 Markov Chain Monte Carlo Maximum Likelihood Method

The inability to compute MLEs directly with Gibbs-MRF models led to the use of pseudolikelihood as a method of estimation for the model parameters. Geyer & Thompson (1992) developed a *Monte Carlo Maximum Likelihood* (MCML) method which relies on Markov chain sampling methods such as the Gibbs sampler to simulate an ergodic Markov chain which converges in distribution to the desired MRF.

Suppose we have a single realization $\mathbf{z} = (z_1, \dots, z_m)$ on a lattice $D(m)$ of m sites from a spatial lattice process $\mathbf{Z} = (Z_1, \dots, Z_m)$ with Gibbs density given by:

$$f(\mathbf{z} | \boldsymbol{\theta}) = c^{-1}(\boldsymbol{\theta}) \exp \left\{ \sum_{i=1}^p U_i(\mathbf{z})\theta_i \right\}, \quad (1)$$

where $c(\boldsymbol{\theta}) = \int \exp \{ \sum_{i=1}^p U_i(\mathbf{z})\theta_i \} d\mu(\mathbf{z})$ is the normalizing constant, and μ is counting measure over $\boldsymbol{\Omega} = \Omega_1 \times \dots \times \Omega_m$, where Ω_i is the single-site state space ($\Omega_i = \{0, 1\}$ for binary data). The vectors $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ and $\mathbf{U}(\cdot) = (U_1(\cdot), \dots, U_p(\cdot))$ are the natural parameter vector and statistic respectively for this exponential family distribution. When $c(\boldsymbol{\theta})$ is not of closed form, straightforward maximum likelihood estimation is not feasible; however, Geyer & Thompson (1992) show that the generation of a Markov chain of n lattice samples $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ using MCMC sampling methods yields a consistent MCMC estimator of $c(\boldsymbol{\theta})/c(\boldsymbol{\psi})$ for any $\boldsymbol{\theta}, \boldsymbol{\psi} \in \Theta$, given by:

$$d_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{k=1}^n \exp \{ \mathbf{T}'_k(\boldsymbol{\theta} - \boldsymbol{\psi}) \} \xrightarrow{a.s.} \frac{c(\boldsymbol{\theta})}{c(\boldsymbol{\psi})} \text{ as } n \rightarrow \infty, \quad (2)$$

for any fixed $\boldsymbol{\theta}$. Using (1), the log-likelihood of $\boldsymbol{\theta}$ given \mathbf{z} can be rewritten in terms of the ratio $c(\boldsymbol{\theta})/c(\boldsymbol{\psi})$, and approximated via (2) to give a Monte Carlo approximate log-likelihood: $l_n(\boldsymbol{\theta} | \mathbf{z}) = -\log[d_n(\boldsymbol{\theta})] + \sum_{i=1}^p U_i(\mathbf{z})\theta_i$. Assuming $\hat{\boldsymbol{\theta}}_{MC}$ and $\hat{\boldsymbol{\theta}}_{ML}$, the maxima of the approximate and true likelihoods, respectively, exist and are unique, strong convergence of the maximizers of the form $\hat{\boldsymbol{\theta}}_{MC} \xrightarrow{a.s.} \hat{\boldsymbol{\theta}}_{ML}$ as $n \rightarrow \infty$ follows from the concavity of the likelihoods (Geyer & Thompson 1992). A unique maximum $\hat{\boldsymbol{\theta}}_{MC}$ of $l_n(\boldsymbol{\theta} | \mathbf{z})$ exists if the vector of sufficient statistics $\mathbf{U}(\mathbf{z})$ is contained in the convex hull of the vectors of sufficient statistics $\mathbf{T}_1, \dots, \mathbf{T}_n$ from the Markov chain $\boldsymbol{\psi}$ -samples (Penttinen 1984).

The method of Markov chain Monte Carlo maximum likelihood (MCML) has received much attention in recent years, and has been useful in solving a variety of complex problems (Geyer & Thompson 1992). The main drawback of this method is the tremendous

amount of computing effort required to obtain the Monte Carlo maximum likelihood estimator (MCMLE) in some large-scale problems. The major advantages of MCML lie with the corresponding asymptotic theory and its reliance on maximum likelihood theory.

Note that the convergence of the MCMLE $\hat{\theta}_{MC}$, given above, is to the MLE $\hat{\theta}_{ML}$, not the true θ_0 . Under an identifiability assumption on the true θ_0 , Gidas (1993) proved the strong convergence of the MLE to θ_0 for any $\theta_0 \in \Theta$ as the *lattice size* m increases subject to the van Love conditions stated earlier. This result and the convergence of $\hat{\theta}_{MC}$ to $\hat{\theta}_{ML}$ as $n \rightarrow \infty$ can be combined to give the convergence of $\hat{\theta}_{MC} \xrightarrow{a.s.} \theta_0$ as first n and then $m \rightarrow \infty$.

A related result on the asymptotic normality of the MCMLE $\hat{\theta}_{MC}$ also involves a simultaneous treatment of the two types of asymptotics present, again stemming from initially separate results. Under simple regularity conditions, Geyer (1994) demonstrates the asymptotic normality of $\hat{\theta}_{MC} - \hat{\theta}_{ML}$ at a rate of \sqrt{n} , where n is the Monte Carlo sample size. Under stricter conditions, Gidas (1993) establishes the asymptotic normality of the true MLE $\hat{\theta}_{ML}$ at a rate of \sqrt{m} , where m is the lattice size. These two asymptotic results may be combined to give: $\sqrt{m}(\hat{\theta}_{MC} - \theta_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}^{-1}(\theta_0))$ as n , then $m \rightarrow \infty$, under the assumptions discussed. The primary motivation for reviewing results on the consistency and asymptotic normality of the MCML estimator is to use these results to find the asymptotic distributions of test statistics resulting from this method. In the subsequent sections, test statistics based on these estimation procedures will be developed, their asymptotic distributions found, and the adequacy of the asymptotics with simulated and real data examined.

4. Inference Methods

In this section, several methods for testing hypotheses on the model parameters in autologistic models are presented. In general, for a vector of model parameters $\theta = (\theta_1, \dots, \theta_p)$, tests of general form $H: \theta \in \Theta_H$ vs. $A: \theta \in \Theta_H^C$ will be considered, where Θ_H is the parameter space for θ under the null hypothesis H . The first two presented methods treat the PL function as a likelihood, yielding “pseudolikelihood ratio test (PLRT)” statistics which are transformed to approximate a chi-square distribution. The final three testing methods extend the Markov chain Monte Carlo maximum likelihood methodology to a Monte Carlo likelihood ratio test (MCLRT), Wald test (MCWT), and Lagrange multiplier test (MCLMT).

4.1 Pseudolikelihood Ratio Tests

Since the pseudolikelihood (PL) is not a true likelihood function, it does not make sense to develop a pseudolikelihood ratio test (PLRT) as a ratio of likelihoods in the usual way. Letting λ_P denote the PLRT statistic, simulation demonstrates clearly that $-2 \log \lambda_P$ is not approximately chi-squared distributed. However, given the ease of implementation and popularity of the PL method in practice, a formal testing procedure based on some transformation of λ_P would be extremely useful. Here, two tests which match the first two moments of a transformed PLRT statistic to those of a χ^2 -random variable are discussed. No formal justification for these transformations is provided; however, a heuristic argument is given.

The coding method of estimation (Besag 1974) considers a partition of the lattice D into particular sublattices $D_i, i = 1, \dots, s$, such that the random variables Z_{jk} at sites in a given sublattice are conditionally independent, given the values at neighboring sites. Taking

the product of the site probabilities over the sites in each sublattice yields a sequence of “coding likelihood” functions $CL_1(\boldsymbol{\theta}|\mathbf{z}), \dots, CL_s(\boldsymbol{\theta}|\mathbf{z})$ corresponding to D_1, \dots, D_s . Because $\{D_1, \dots, D_s\}$ partition D , the PL function can be expressed as: $PL(\boldsymbol{\theta}|\mathbf{z}) = \prod_{i=1}^s CL_i(\boldsymbol{\theta}|\mathbf{z})$. Using this equality, and letting $\hat{\boldsymbol{\theta}}_{PL}$ and $\hat{\boldsymbol{\theta}}_{PL}^H$ be the unrestricted and restricted MPLEs respectively, the PLRT statistic can be approximated as:

$$\lambda_P = \frac{PL(\hat{\boldsymbol{\theta}}_{PL}^H | \mathbf{z})}{PL(\hat{\boldsymbol{\theta}}_{PL} | \mathbf{z})} = \prod_{i=1}^s \frac{CL_i(\hat{\boldsymbol{\theta}}_{PL}^H | \mathbf{z})}{CL_i(\hat{\boldsymbol{\theta}}_{PL} | \mathbf{z})} \approx \prod_{i=1}^s \lambda_{C_i},$$

where the $\lambda_{C_i}, i = 1, \dots, s$ are LRT statistics for the coding likelihoods. Under H_0 , the $-2 \log \lambda_{C_i}$ statistics, $i = 1, \dots, s$, are *dependent* identically-distributed random variables with an asymptotic χ^2 -distribution, so $-2 \log \lambda_P$ is approximately the sum of dependent identically-distributed χ^2 -random variables. This provides the basis for the first PLRT, denoted by $PLRT_C$, which is an approximation for a sum of dependent χ^2 -random variables.

PLRT_C Method: Viewing $-2 \log \lambda_P$ as $\sum_{i=1}^s X_i$ where X_1, \dots, X_s are dependent identically distributed $X^2(r)$ random variables with an assumed common correlation ρ between each pair (of CLRT statistics), the first two moments of a chi-squared random variable can be matched with those of $a_c \sum X_i$ where $a_c > 0$ is a constant. Simple computation verifies that if $a_c = 1/[1 + (s - 1)\rho]$ where s is the number of coding sets, then the means and variances of $a_c \sum X_i$ and a $\chi^2(sra_c)$ -variable are equal. Hence, if an estimate of $\rho = Corr(X_i, X_j) = Corr(-2 \log \lambda_{C_i}, -2 \log \lambda_{C_j})$ can be obtained, the PLRT may be useful as an asymptotic chi-squared test. This can be done with the Gibbs sampler by generating sample data from the distribution with the PL estimates as parameter values, and calculating the sample correlation between pairs of $-2 \log \lambda_{C_i}$ test statistics as estimates of ρ .

PLRT_P Method The second potential PLRT, denoted $PLRT_P$, attempts to match the mean and variance of the statistic $-2a_p \log \lambda_P$ ($a_p > 0$, a constant) to that of a χ^2 -random variable. Suppose μ_P and σ_P^2 are the mean and variance of $-2 \log \lambda_P$. Then matching the first two moments of $-2a_p \log \lambda_P$ with those of a χ^2 -random variable yields $a_p = 2\mu_P/\sigma_P^2$ as the choice of a_p such that $-2a_p \log \lambda_P$ has the same mean and variance as a $\chi^2[2\mu_P^2/\sigma_P^2]$ -distribution. Again, μ_P and σ_P^2 are in general unknown but can be estimated using the Gibbs sampler.

Both of these PLRT methods will be investigated through simulation studies in later sections. The chi-square approximations can be justified by examining plots of the empirical density of $-2 \log \lambda_P$ for simulated data. Other efforts to develop a test procedure based on PL estimation include a Cholesky decomposition of the PL-based estimated information matrix which also exploits the coding-PL relationship (Miller 1981), and a subset selection method based on the deviance statistic in the presence of covariate data (Huffer & Wu 1995).

4.2. Monte Carlo Likelihood Ratio Test

A natural question stemming from the MCML method is whether similar ideas (estimation of the ratio of normalizing constants) can be used for developing a hypothesis testing procedure. Suppose that a realization \mathbf{z} on a lattice $D(m)$ of size m is observed. As with MCML estimation, we choose some $\boldsymbol{\psi} \in \Theta \subset \mathfrak{R}^p$, where Θ is the unrestricted parameter

space for θ , and let P_ψ denote the probability measure having density f_ψ with respect to a measure μ . With the likelihood $L(\theta|z)$ given by (1), the usual likelihood ratio test (LRT) statistic for testing $H: \theta \in \Theta_H$ vs. $A: \theta \notin \Theta_H$, is:

$$\begin{aligned} \lambda_m &= \frac{\sup_{\theta \in \Theta_H} L(\theta|z)}{\sup_{\theta \in \Theta} L(\theta|z)} = \frac{\sup_{\theta \in \Theta_H} [(c(\theta) c(\psi))^{-1} \exp\{\sum_{i=1}^p U_i(z)\theta_i\}]}{\sup_{\theta \in \Theta} [(c(\theta) c(\psi))^{-1} \exp\{\sum_{i=1}^p U_i(z)\theta_i\}]} \\ &= \frac{[c(\hat{\theta}_{ML}^H)/c(\psi)]^{-1} \exp\{\sum_{i=1}^p U_i(z)\hat{\theta}_{MLi}^H\}}{[c(\hat{\theta}_{ML}) c(\psi)]^{-1} \exp\{\sum_{i=1}^p U_i(z)\hat{\theta}_{MLi}\}} = \frac{\tilde{L}(\hat{\theta}_{ML}^H|z)}{\tilde{L}(\hat{\theta}_{ML}|z)}, \end{aligned} \tag{3}$$

where $\hat{\theta}_{ML}^H$ is the MLE of θ_0 under H , and $\hat{\theta}_{ML}$ is the unrestricted MLE of θ_0 . Application of the MCMC methodology is now clear from expression (3). Generation of a sufficiently large number of ψ -samples will enable estimation of the numerator and denominator ratios $c(\hat{\theta}_{ML}^H)/c(\psi)$ and $c(\hat{\theta}_{ML})/c(\psi)$ in (3) using the Monte Carlo approximant $d_n(\theta)$ in (2). The critical point here is that the same ψ needs to be used for both approximations, making the choice of ψ an important consideration in terms of the number of Monte Carlo ψ -samples required for convergence of $d_n(\theta)$ in the two maximizations. Intuitively, choices of ψ between θ_0 and θ_0^H should allow for more rapid convergence and more accurate approximations.

For all $n = 1, 2, \dots$, define λ_{mn} as:

$$\lambda_{mn} = \frac{[d_n(\hat{\theta}_{MC}^H)]^{-1} \exp\{\sum_i U_i(z)\hat{\theta}_{MCi}^H\}}{[d_n(\hat{\theta}_{MC})]^{-1} \exp\{\sum_i U_i(z)\hat{\theta}_{MCi}\}} = \frac{\sum_{k=1}^n \exp\{(\mathbf{T}_k - \mathbf{U})'(\hat{\theta}_{MC}^H - \psi)\}}{\sum_{k=1}^n \exp\{(\mathbf{T}_k - \mathbf{U})'(\hat{\theta}_{MC}^H - \psi)\}},$$

where $\tilde{L}_n(\cdot|z)$ is the Monte Carlo approximant to the likelihood. The terms \mathbf{T}_k and \mathbf{U} are the vector of sufficient statistics for ψ from the k^{th} Monte Carlo sample, and for θ respectively. It then follows from the MCML development that as $n \rightarrow \infty$, $\lambda_{mn} \xrightarrow{a.s.} \lambda_m$.

In well-behaved distributions, such as the exponential family models studied here, the statistic $-2 \log \lambda_m$ converges in distribution to a chi-squared random variable as the “sample size”, or number of sites m increases. Appealing to the general theory for exponential families of distributions, the convergence of $-2 \log \lambda_m \xrightarrow{D} \chi^2(r)$ as $m \rightarrow \infty$ follows from the consistency and asymptotic normality of the MLE shown by Gidas (1993). Letting $n, m \rightarrow \infty$ in sequence, it is then straightforward to show that $-2 \log \lambda_{mn} \xrightarrow{D} \chi^2(r)$.

4.3. Monte Carlo Wald Test

In maximum likelihood estimation, the inverse of the observed information matrix is an asymptotic estimator for the variance-covariance matrix of the parameter estimates. Here it is shown that the MCML methodology leads naturally to a MCMC-based estimator of the inverse information matrix, providing the basis for a Monte Carlo Wald Test (MCWT).

For the class of Gibbs distributions within the exponential family, the information matrix is given by: $I(\theta) = \mathbf{E}[(\partial \log L / \partial \theta)(\partial \log L / \partial \theta)'] = [\mathbf{Var}(\mathbf{U}(z))]^{-1}$, where $\mathbf{Var}(\mathbf{U}(z))$ is

simply the variance-covariance matrix of the sufficient statistics. Computation of the first- and second-order partials of $\log c(\boldsymbol{\theta})$ yields an alternative expression for $I(\boldsymbol{\theta})$ as:

$$I(\boldsymbol{\theta}) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} [\log c(\boldsymbol{\theta})] = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} [\log(c(\boldsymbol{\theta})/c(\boldsymbol{\psi}))], \quad (4)$$

where division by $c(\boldsymbol{\psi})$ in the final equality in (4) does not change the expression. Written this way, the application of the MCMC methodology is apparent. Using (2) to estimate $c(\boldsymbol{\theta})/c(\boldsymbol{\psi})$ provides a Monte Carlo approximation to the information matrix as:

$$I_{MC}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} [\log \{d_n(\boldsymbol{\theta})\}] = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} [\log \{ \frac{1}{n} \sum_{k=1}^n \exp\{\mathbf{T}'_k(\boldsymbol{\theta} - \boldsymbol{\psi})\} \}]. \quad (5)$$

where \mathbf{T}_k is the vector of sufficient statistics in the k^{th} $\boldsymbol{\psi}$ -generated lattice sample, $k = 1, \dots, n$. Finally, substitution of the MCMLE $\hat{\boldsymbol{\theta}}_{MC}$ for $\boldsymbol{\theta}$ in (5) gives the estimated Monte Carlo information matrix.

Consider a test of the form $\mathbf{H}: \mathbf{R}(\boldsymbol{\theta}) = \mathbf{0}$ vs. $\mathbf{A}: \mathbf{R}(\boldsymbol{\theta}) \neq \mathbf{0}$, where $\mathbf{R}(\boldsymbol{\theta}) = (R_1(\boldsymbol{\theta}), \dots, R_r(\boldsymbol{\theta}))$ is a vector of $r \leq p$ independent hypotheses on $\boldsymbol{\theta}$. Assume that each $R_i(\cdot)$ possesses continuous first-order partial derivatives, and let $D(\boldsymbol{\theta}) = \partial \mathbf{R}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}' = \{\partial R_i / \partial \theta_j\}_{i,j}$, $i = 1, \dots, r$, $j = 1, \dots, p$, be the $r \times p$ matrix of all first order partials. The Monte Carlo Wald test statistic is then defined as:

$$W_{mn} = \mathbf{R}'(\hat{\boldsymbol{\theta}}_{MC}) [D(\hat{\boldsymbol{\theta}}_{MC}) \hat{\mathbf{I}}_{MC}^{-1}(\hat{\boldsymbol{\theta}}_{MC}) D'(\hat{\boldsymbol{\theta}}_{MC})]^{-1} \mathbf{R}(\hat{\boldsymbol{\theta}}_{MC}),$$

As with the MCMC likelihood ratio test statistic, this MCMC Wald test statistic can be shown to converge in distribution to a chi-square distribution with degrees of freedom equal to the number of independent hypotheses r , as the Monte Carlo size n and lattice size m increase sequentially to ∞ . A proof of this result follows from the strong convergence and asymptotic normality of the MCMLE shown by Gidas (1993) under fairly strict regularity conditions, and can be found in Graham (1995).

4.4. Monte Carlo Lagrange Multiplier Test

In addition to the likelihood ratio test and Wald test, a third commonly used testing procedure which requires only one maximization under the reduced model is the Lagrange multiplier test or Rao's score test. Suppose again that we wish to test the set of independent hypotheses on the $p \times 1$ vector of model parameters $\boldsymbol{\theta}$ with form: $\mathbf{H}: \mathbf{R}(\boldsymbol{\theta}) = \mathbf{0}$ vs. $\mathbf{A}: \mathbf{R}(\boldsymbol{\theta}) \neq \mathbf{0}$. Let $\mathbf{s}(\boldsymbol{\theta}) = [(1/\sqrt{m})(\partial l_m(\boldsymbol{\theta}|\mathbf{z})/\partial \boldsymbol{\theta})]_{p \times 1}$, and $\mathbf{C}(\boldsymbol{\theta}) = [(1/m)(\partial^2 l_m(\boldsymbol{\theta}|\mathbf{z})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}')]_{p \times p}$, with $\hat{\mathbf{s}}$ and $\hat{\mathbf{C}}$ representing the values of \mathbf{s} and \mathbf{C} evaluated at the restricted MLE $\hat{\boldsymbol{\theta}}^H$, and $l_m(\cdot|\mathbf{z})$ denoting the true *unrestricted* log-likelihood. Under suitable regularity conditions, the Lagrange multiplier test statistic given by: $LM = \hat{\mathbf{s}}' \hat{\mathbf{C}}^{-1} \hat{\mathbf{s}}$ has an asymptotic χ^2 -distribution with degrees of freedom equal to r , the number of independent hypotheses being tested.

For Gibbs models, however, closed form expressions for the first and second partials of the log-likelihood function $l_m(\boldsymbol{\theta}) = -\log c(\boldsymbol{\theta}) + \mathbf{U}'(\mathbf{z})\boldsymbol{\theta}$ do not exist. Using again the fact

that $\partial/\partial\theta \log c(\theta) = \partial/\partial\theta \log[c(\theta)/c(\psi)]$, it can be shown that the LM test statistic takes the form:

$$LM_m = \left[U(z) - \frac{\partial}{\partial\theta} \log \frac{c(\theta)}{c(\psi)} \right]' \left[-\frac{\partial^2}{\partial\theta\partial\theta'} \log \frac{c(\theta)}{c(\psi)} \right]^{-1} \left[U(z) - \frac{\partial}{\partial\theta} \log \frac{c(\theta)}{c(\psi)} \right] \Big|_{\hat{\theta}^H} \quad (6)$$

where as before, the Monte Carlo approximant $d_n(\theta)$ can be used to estimate $c(\theta)/c(\psi)$, and $\hat{\theta}_{MC}^H$ to estimate $\hat{\theta}^H$, yielding a Markov chain Monte Carlo Lagrange multiplier (MCLM) test statistic LM_{mn} (see Graham (1995) for details).

Under the assumption that the first and second order partials of $c(\theta)$ are continuous in a neighborhood of θ_0 , it can be shown that $LM_{mn} \xrightarrow{a.s.} LM_m$ as $n \rightarrow \infty$. And as with the previous two MCMC test statistics, convergence in distribution under H of LM_m to a chi-square random variable with r degrees of freedom follows from the convergence and asymptotic normality of the MCMLE (Graham 1995). Combining these results, it follows that under H , $LM_{mn} \xrightarrow{D} \chi^2(r)$ random variable as n and then $m \rightarrow \infty$.

5. Simulation Studies

In an effort to compare the PL and MCMC inference procedures, two simulation studies are performed. The first of these investigates the accuracy of the asymptotic distributions of the test statistics for a number of different sets of hypotheses and lattice sizes. The second study examines the power of different tests and how it changes under different levels of spatial dependence. In all cases, only first- and second-order AL models are considered over nested lattices of size 20x20, 40x40, and 60x60. Edge effects were handled by performing inference only on an inner portion of each lattice with an outer “guard” region of width 1 used only in the conditioning. Values at the edge sites of the lattice samples were generated under a torus assumption.

5.1 Simulation Study: Empirical Distributions of Test Statistics

The purpose of this first series of simulations is to compare the empirical distributions of the test statistics under the null hypothesis, computed from generated Markov chain lattice samples, to their hypothesized asymptotic χ^2 -distributions. Five separate simulations are run for five different models and corresponding tests on the model parameters. The first three of these analyzed the test $H: \beta = \beta_0$ vs. $A: \beta \neq \beta_0$ for $\beta_0 = 0.25, 0.50, 1.00$ respectively, for lattice samples from a first-order $AL(\alpha, \beta)$ distribution ($\alpha = -2\beta$ in all cases). The fourth examines the test $H: \beta = \gamma$ vs. $A: \beta \neq \gamma$ for $\beta = \gamma = 0.5$ in the bi-directional first-order $AL(\alpha, \beta, \gamma)$ model, and the fifth the test $H: \delta = 0$ vs. $A: \delta \neq 0$ for $\delta = 0$ in the second-order $AL(\alpha, \beta, \delta)$ model. The following general procedure is used for each of these five cases.

Let θ represent the vector of parameters for a given model, and θ_0 the true value of θ under H . Using a Gibbs sampler, 500 lattice samples of size 64x64 are generated from the autologistic model under the null hypothesis $\theta = \theta_0$. From each of the 500 generated samples, the unrestricted and restricted MPLEs, $\hat{\theta}_{PL}$, $\hat{\theta}_{PL}^H$, and MCMLEs $\hat{\theta}_{MC}$, $\hat{\theta}_{MC}^H$, are calculated. To compute the MCMLEs, an additional single stream of 2000 MCMC samples from the autologistic model with parameter $\psi = \hat{\theta}_{PL}$ is used. Using these maxima, the various test statistics for the $PLRT_C$, $PLRT_P$, MCLRT, MCWT, and MCLMT methods,

given by $-2a_C \log \lambda_P$, $-2a_P \log \lambda_P$, $-2 \log \lambda_{mn}$, W_{mn} , and LM_{mn} respectively, are found. The resulting 500 values for each test statistic comprise the empirical distribution of the test statistics. To compare this to the proposed χ^2 -distributions, the 500 values were categorized into the 20 groups with endpoints at every fifth percentile of the proposed distribution. With each group having an expected count of 25 under H , the accuracy of the asymptotics is assessed using a simple chi-squared GOF test for each of the five testing methods over all five cases and all three lattice sizes, with the final results reported in Table 1.

The details of this simulation study are not reported here (see Graham 1995 for details); only the salient features are indicated. In viewing the table, it is clear that there are severe departures from a chi-squared distribution for the test statistics under the $PLRT_C$ testing method in two cases. The two problematic cases represent the presence of moderately high dependence ($\beta = 1.00$), and second-order dependence ($\beta = \delta = 0.50$). This poor behavior for the $PLRT_C$ was consistently observed in AL models with high first-order dependence or second-order dependence. For each of the remaining methods except for the MCLRT method, there were minor departures at the 5% significance level. Based on this study, the test statistics for the MCMC methods as well as for the $PLRT_P$ method seem to approximate a chi-squared distribution reasonably well for simple AL models. It is also encouraging that neither the lattice size nor the level of dependence ($\beta_0 = 0.25, 0.50, 1.00$ indicating low, moderate, and high dependence) seem to affect the accuracy of these inference procedures.

5.2 Simulation Study: Power Comparison of Test Statistics

A standard way of comparing hypothesis testing procedures is through the power of a test. To calculate the power for a given test, say $H: \beta = 0.5$ vs. $A: \beta \neq 0.5$, we would need to compute the power function $P(\beta_0) = Pr(\text{Reject } H \mid \beta_0 \text{ is the true value})$ for different values of β_0 . Since this function is computed through simulation for autologistic models, this entails the generation of sample lattices via the Gibbs sampler, for *each* value of β_0 . Unfortunately, this MCMC calculation of $P(\beta_0)$, even with a moderately small sample size of 500 over a small range of β_0 values is prohibitively computationally intensive. In an effort to circumvent this problem and still make a meaningful comparison of the power for the inference procedures, an alternative procedure is employed.

Instead of performing power calculations for a fixed hypothesis through the generation of different lattice samples for each β_0 , the power is calculated at a *fixed* β_0 (where the data are generated under β_0) for *different hypotheses*. As an example, for $\beta_0 = 0.5$ ($\alpha_0 = -1.0$) as above, 500 lattice samples are generated from the AL (α_0, β_0) distribution. In addition to computing the various $PLRT$, $MCLRT$, $MCWT$, and $MCLMT$ test statistics for H above, these test statistics are computed for many other hypotheses, such as ($H: \beta = 0.4$ vs. $A: \beta \neq 0.4$) or ($H: \beta = 0.6$ vs. $A: \beta \neq 0.6$). This is simple to do since the test statistics are computed from the *same* set of generated data; only separate maximizations of the pseudolikelihood and Monte Carlo likelihoods are required for each case.

This procedure is performed using the β -values from the first three of the five cases from the previous simulation study as true values, at lattice sizes of 20x20, 40x40, and 60x60. For each lattice-case combination, 15 tests of the form $H_i: \beta = \beta_i$ vs. $A_i: \beta \neq \beta_i$, $i = 1, \dots, 15$ were performed for each of the 500 samples generated from the autologistic

(α, β) distribution. The β_i values were chosen at varying distances both above and below β_0 to give a wide range of power values. This resulted in 15 test statistics corresponding to the 15 β_i values for each testing method, for each of the 500 samples.

These 500 test values are compared to the proper 95% critical value for each test, and the number of significant values (values larger than the critical value) is recorded. Naturally, with a significance level of 0.05 for testing $H: \beta = \beta_0$ vs. $A: \beta \neq \beta_0$, we expect 25 of the 500 values to be significant for each method if the approximate χ^2 -distributions are accurate. These proportions of significant values are simply estimates of the size of the test, or power at β_0 , $P(\beta_0) = Pr(\text{Reject } H_i \mid \beta_0)$, for the 15 tests (H_1 vs. A_1), \dots , (H_{15} vs. A_{15}). These power values, referred to as rejection probabilities, are plotted against the β_i , $i = 1, \dots, 15$, the test values for β used in the 15 tests. It is important to note that the resulting curve is not a standard power curve, but gives the power for rejecting various hypotheses H_i given that the data are distributed according to $\beta = \beta_0$.

Figure 2 contains four of these rejection probability plots. The first three plots represent the three cases (or levels of dependence) studied: $\beta = 0.25, 0.50$, and 1.00 for the 20×20 lattice, and the last examines the differences in power for different lattice sizes for the MCLRT method. Plots for other lattice sizes and cases not shown here are qualitatively the same. It is worth emphasizing that these plots are *not* empirical power curves, but do enable a worthwhile comparison between the various tests by examining each test's probability of rejecting the null hypothesis when it is false, for different types of tests.

In viewing these plots, a number of observations can be made. First, it is clear from each of the first three plots that the MCMC testing procedures have greater power for rejecting each test $H_i: \beta = \beta_i$ vs. $A_i: \beta \neq \beta_i$, than the PLRT procedures for larger β_i . This difference is especially noticeable in the moderately high dependence test case: $H: \beta = 1.00$ vs. $A: \beta \neq 1.00$. The same general result holds true for the larger lattice sizes, although the effects were less pronounced. One might conclude from this that there may be significant gains in power to using the MCMC procedures over the PLRT procedures when spatial dependence is large, although many other types of tests should be examined.

A second point worth emphasizing from these plots is an apparent bias in the MCLMT method. Although the MCLMT method appears to have greater power than the MCLRT and MCWT methods, at $\beta = \beta_0$, the power of rejection for the MCLMT statistic is greater than the expected 0.05 significance level. Hence, this test is too liberal under the null hypothesis, possibly giving the rejection probability curve an upward bias. The source of this bias is unclear, and was not present in the other two MCMC test procedures.

Third, the fourth plot in Figure 2 demonstrates the tremendous gains in power available with larger lattice sizes for tests on the dependence parameter. As an example, if the true $\beta = 0.5$, and we test $H_i: \beta = 0.4$ vs. $A_i: \beta \neq 0.4$, we only have about a 10% chance of rejecting H_i with a 20×20 lattice, 30% with a 40×40 lattice, and 60% with a 60×60 lattice. As a general conclusion from these two simulation studies, the MCLRT and MCWT inference procedures seem to perform favorably to the other test procedures considered in terms of accuracy and precision under different levels of dependence and different lattice sizes. In addition, lattice size has a profound effect on the variability inherent with these test statistics and hence on

the power of the various tests.

6. Application: Phytophthora Disease Incidence Data in Bell Pepper Plants

The primary objective of this section is to demonstrate the use of the MCMC inference procedures developed in answering specific physical questions concerning the direction and magnitude of disease spread in the data. The use of these methods in performing stepwise model selection is also indicated. One question of interest to a plant pathologist might be whether or not the spread of the pathogen *P. capsici* is more prevalent within a row of pepper plants than it is between rows. This might indicate that surface water within a row is acting as an important mechanism of spread as compared to root-to-root contact. Using only the 20x20 lattice of data shown in Figure 1, a test of the presence of spatial dependence in this data using these procedures indicates ($p < 0.0001$) that some form of spatial dependence is present. To then test the proposed question above, a bi-directional first-order $AL(\alpha, \beta, \gamma)$ model is fit to the data, where β and γ represent the within-row and across-row first-order dependence, respectively. Table 2 contains the parameter estimates, estimated standard errors, and test statistics for testing the hypothesis $H: \beta = \gamma$ vs. $A: \beta \neq \gamma$, namely whether the first-order dependence is the same across rows as it is within rows.

In viewing Figure 1, the disease spread appears to be more prevalent within rows; so we might expect there to be a difference in the two directions of dependence, with β being larger. The results from Table 2 support this belief for the most part, although the $PLRT_P$ method performs terribly. The Monte Carlo tests agree fairly well here (especially the first two), all giving p-values less than or equal to 0.002, indicating that the within-row and across-row dependence are different at a 0.05 significance level. As we noted in the simulation studies, the MCLMT statistic is biased toward the alternative hypothesis, and so the smaller p-value is not surprising. One final point worth mentioning is the rather large standard errors associated with the parameter estimates. Larger lattice sizes or less discretized data would reduce this variation, but such large variability is common with binary data on a lattice of size 20x20. The large standard errors may seem troublesome to the practitioner interested in precise estimates of the model parameters, but are not so large as to weaken the utility of the inference procedures.

7. Summary

The simulation studies and application to the *Phytophthora* data in this paper indicate that the MCMC methodology for parameter estimation in AL models can be extended to include formal methods of inference. Both the MCLRT and MCWT perform favorably in terms of accuracy of the asymptotics and power in comparison to the other methods considered, and are the recommended methods of inference for use with autologistic models based on this study. The MCLMT appears to be too liberal, and neither of the PL-based tests are reliable in the presence of spatial dependence. In practice, the PL estimation method has been preferred to the MCML method due primarily to computational requirements. However, even with AL models of up to seven parameters, less than ten minutes on a Sparc-10 are necessary to obtain parameter estimates, their estimated variances, and the test statistics using the MCMC procedures.

Acknowledgements

As much of this research was conducted as part of the author's doctoral work, the author thanks Dr. Marcia Gumpertz and Dr. B.B. Bhattacharyya for their many helpful suggestions, and Dr. Jean Ristaino for the use of her data.

References

- [1] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36:192–236, 1974.
- [2] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179–195, 1975.
- [3] J. Besag and P. Green. Spatial statistics and bayesian computation. *Journal of the Royal Statistical Society, Series B*, 55(1):25–37, 1993.
- [4] N. Cressie. *Statistics for Spatial Data*. John Wiley, New York, 1993.
- [5] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
- [6] S. Geman and C. Graffigne. Markov random field image models and their application to computer vision. *Proceedings of the International Congress of Mathematics*, pages 1496–1517, 1986.
- [7] C. J. Geyer. On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society, B*, 56(1):261–274, 1994.
- [8] C. J. Geyer and E. A. Thompson. Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, B*, 54(3):657–699, 1992.
- [9] C. J. Geyer and E. A. Thompson. Annealing Markov chain Monte Carlo with applications to ancestral inference. Technical Report 589, University of Minnesota, October 1994.
- [10] B. Gidas. *Parameter Estimation for Gibbs Distributions from Fully Observed Data*. In *Markov Random Fields: Theory and Applications (R. Chellapa and R. Jain, eds.)*, chapter 17, pages 471–498. Academic Press, Inc., 1993.
- [11] J. Graham. *Markov chain Monte Carlo inference procedures for Discrete Spatial Lattice Models*. PhD thesis, North Carolina State University, 1995.
- [12] M. L. Gumpertz, J. M. Graham, and J. B. Ristaino. Autologistic model of spatial pattern of *Phytophthora* epidemic in bell pepper: Effects of soil variables on disease presence. *Proceedings of 1994 Kansas State University Conference on Applied Statistics in Agriculture*, 1995.

Table 1: Table of Pearson χ^2 LOF Test Statistics for Testing Cases 1-5

Testing Method	Lattice Size	Case 1	Case 2	Case 3	Case 4	Case 5
		$H : \beta = 0.5$ $A : \beta \neq 0.5$	$H : \beta = 1.0$ $A : \beta \neq 1.0$	$H : \beta = 0.25$ $A : \beta \neq 0.25$	$H : \beta = \gamma$ $A : \beta \neq \gamma$	$H : \beta = \delta$ $A : \beta \neq \delta$
<i>PLRT_C</i>	20	23.84	**171.60	22.24	12.40	**136.64
	40	13.52	**146.96	17.52	15.36	**99.92
	60	*34.24	**176.00	17.68	*31.12	**110.96
<i>PLRT_P</i>	20	25.04	20.00	*30.72	19.52	22.48
	40	20.16	15.44	24.40	26.96	11.76
	60	*31.44	28.96	15.28	25.44	26.08
<i>MCLRT</i>	20	7.12	14.00	11.60	18.08	13.60
	40	15.28	17.28	12.48	15.92	28.32
	60	18.72	10.64	26.56	15.76	14.64
<i>MCWT</i>	20	10.40	15.36	12.08	21.20	24.40
	40	15.52	24.16	16.48	18.48	*32.32
	60	19.84	11.68	25.76	13.36	16.96
<i>MCLMT</i>	20	20.88	15.44	17.44	21.36	29.44
	40	20.00	18.56	17.04	22.88	27.60
	60	22.88	12.48	*30.64	18.72	20.80

* Mild Departure from the Asymptotic Distribution
 ** Severe Departure from the Asymptotic Distribution

Table 2: Table of Parameter Estimates, Estimated Standard Errors, and Test Statistics for Testing $\beta = \gamma$ in the Bi-Directional First-Order Autologistic Model.

Estim. Method	Estimates			Test Value	χ^2		Testing Method
	$\hat{\alpha}$ $\widehat{SE}(\hat{\alpha})$	$\hat{\beta}$ $\widehat{SE}(\hat{\beta})$	$\hat{\gamma}$ $\widehat{SE}(\hat{\gamma})$		d.f.	p-value	
<i>PL</i>	-2.4358	1.6443	0.6422	5.34	1.15	0.0262	<i>PLRT_C</i>
	(0.3210)	(0.2987)	(0.2537)	2.99	4.20	0.5903	<i>PLRT_P</i>
<i>MCML</i>	-2.9788	2.0950	0.8450	9.53	1.00	0.0020	<i>MCLRT</i>
	(0.2426)	(0.2701)	(0.1990)	9.97	1.00	0.0016	<i>MCWT</i>
				13.00	1.00	0.0003	<i>MCLMT</i>

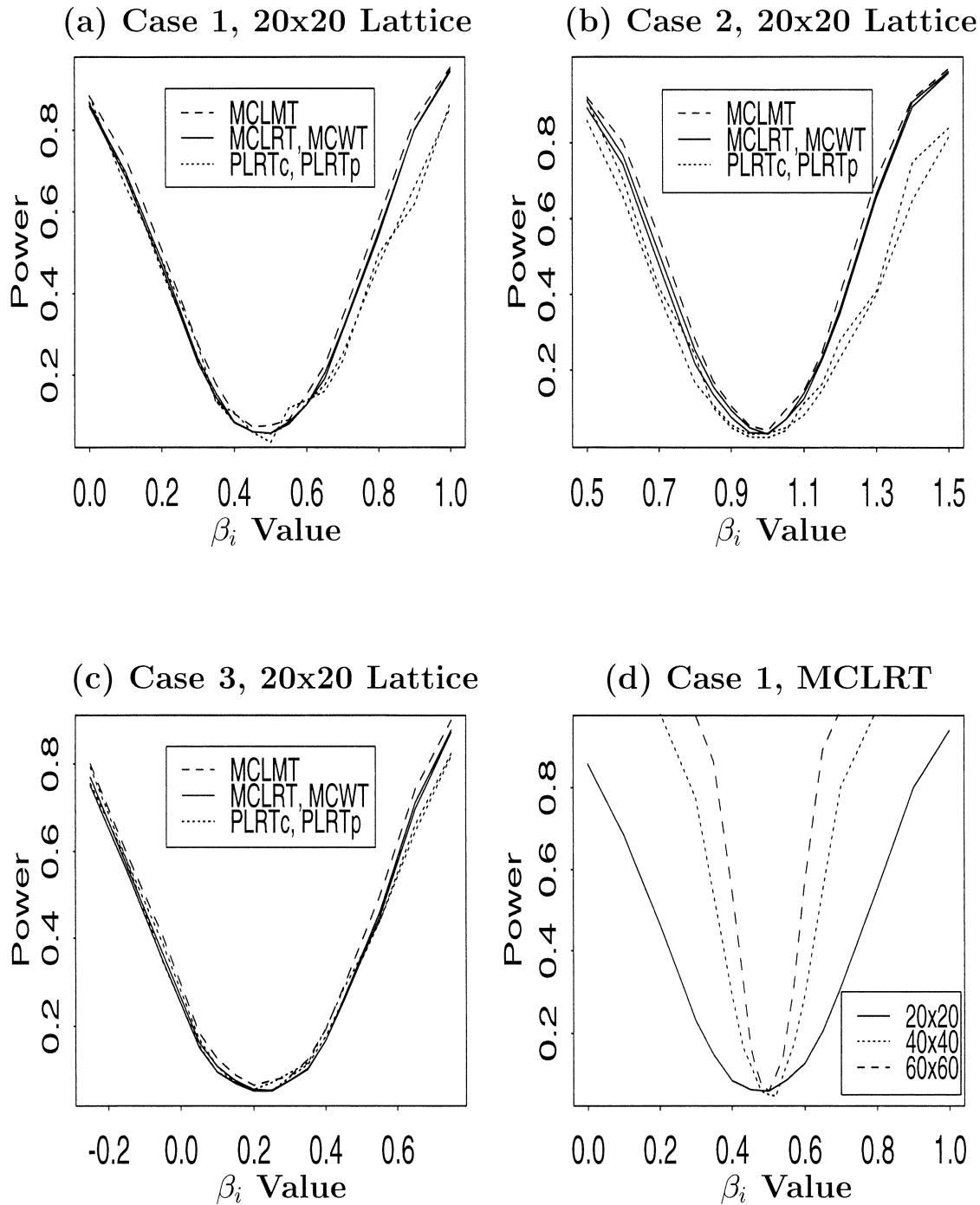


Figure 2: Rejection Probability Plots for *PLRRT*, *MCLRRT*, *MCWT*, and *MCLMT* Hypothesis Testing Methods