Springer

Autologistic Model of Spatial Pattern of Phytophthora Epidemic in Bell Pepper: Effects of Soil Variables on Disease Presence

Author(s): Marcia L. Gumpertz, Jonathan M. Graham and Jean B. Ristaino

Source: *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 2, No. 2 (Jun., 1997), pp. 131-156
Published by: Springer
Stable URL: https://www.jstor.org/stable/1400400
Accessed: 13-03-2020 16:17 UTC

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/1400400?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

JSTOR

Springer is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Agricultural, Biological, and Environmental Statistics*

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

Autologistic Model of Spatial Pattern of Phytophthora Epidemic in Bell Pepper: Effects of Soil Variables on Disease Presence

Marcia L. GUMPERTZ, Jonathan M. GRAHAM, and Jean B. RISTAINO

The autologistic model is a flexible model for predicting presence or absence of disease in an agricultural field, based on soil variables, while taking spatial correlation into account. In the autologistic model, the log odds of disease in a particular quadrat are modeled as a linear combination of disease presence or absence in neighboring quadrats and the soil variables. Neighboring quadrats can be defined as adjacent quadrats within a row, quadrats in adjacent rows, quadrats two rows away, and so forth. The regression coefficients give estimates of the increase in odds of disease if neighbors within a row or in adjacent rows show disease symptoms; thus, we obtain information about the degree of spread in different directions. The coefficients for the soil variables give estimates of the increase in odds of disease as soil water content or pathogen population density increase. In this problem, the soil variables may also be highly correlated over quadrats, and disease incidence in within-row neighbors may be highly correlated with disease incidence in adjacent-row neighbors. This collinearity makes estimation and interpretation of the parameters of the autologistic model more difficult. We discuss fitting the autologistic model and tools for evaluating the aptness of the model.

Key Words: Binary response variable; Disease incidence; Markov random field; Pseudolikelihood estimation; Spatial correlation.

1. INTRODUCTION

Statistical models of the spatial patterns of disease in an agricultural field can be useful for understanding dispersal mechanisms and for developing methods of control of disease. This paper describes and demonstrates the use of the autologistic model for spatial pattern of Phytophthora epidemics in bell pepper. There are three features of the autologistic model that make it well suited to the study of spatial patterns of disease: (1) it applies specifically to binary response variables such as disease presence or absence; (2) explanatory variables can be incorporated into the model; and (3) the probability of disease in a quadrat depends explicitly on whether the neighboring plots are diseased.

Marcia L. Gumpertz is Associate Professor, Department of Statistics, and Jean B. Ristaino is Associate Professor, Department of Plant Pathology, North Carolina State University, Raleigh, NC 27695-8203 and 27695-7616, respectively. Jonathan M. Graham is Assistant Professor, Department of Mathematical Sciences, University of Montana, Missoula, MT 59812.

©*1997 American Statistical Association and the International Biometric Society*
Journal of Agricultural, Biological, and Environmental Statistics, Volume 2, Number 2, Pages 131–156

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

Several different kinds of statistical models have been proposed for describing spatial patterns of disease. The ones that have received the most exposure in the phytopathology literature have been models designed for continuous response variables.

For example, spatio-temporal autocorrelation analysis, described in Reynolds and Madden (1988), and kriging (see Lecoustre et al. 1989) have been proposed for describing spatial correlations in disease epidemics. These methods involve modeling the spatial or spatio-temporal covariance structure of the disease data using autoregressive-moving average type models and variogram models, respectively. They are used to study the effects of distance (and time) on the spread of the epidemic and to map the disease. These statistical methods were developed for response variables that are measured on a continuous scale, such as yield. Although designed for data measured on a continuous scale, they have also been applied to binary and categorical data.

Logistic regression is often used to model nonspatial binary data. Spatial correlation can be incorporated into logistic regression models in two general ways: (1) by explicitly specifying both the spatial correlation structure and the model for the expected probability of disease; or (2) by modeling the probability of disease in a given quadrat as depending on the disease status of neighboring quadrats. Albert and McShane (1995) presented a type of logistic regression model called a marginal model, in which the spatial autocorrelation pattern is modeled explicitly, not by conditioning on the response of the neighbors. They modeled lesion incidence in brain scans using a method of estimation that takes advantage of having separate brain scans for a large number of individuals. This article takes the second approach, focusing on the autologistic model, which was originally developed by physicists to model electron spin at each site in a magnetic field (Cressie 1991). Besag (1972, 1974) developed much of the statistical theory of autologistic models and gave some examples involving plant disease. Some autologistic models incorporating time have also been proposed (Besag 1977; Chadoeuf et al. 1992). For example, Besag (1977) demonstrated the use of the autologistic model to describe incidence of foot rot in endive as a function of disease in neighboring plots at both the current and the previous times. The autologistic model has also been extended to ordered categorical data, such as disease ratings on a scale of 1 to 4 (Strauss 1992). Smyth et al. (1992) applied a similar model to progression of anthracnose in tropical pasture legumes. Wu and Huffer (1996) used an autologistic model with covariates to model the distribution of plant species in Florida.

Most of the studies just cited modeled spatial pattern of disease as a function solely of proximity of diseased plants or of spatial and temporal relationships, but did not incorporate other environmental information. Measures of pathogen population density in the soil and environmental covariates, such as soil water content, microclimate variables, elevation gradient of the field, soil nutrient concentrations, and soil compaction, should provide additional information for predicting the presence or absence of disease and elucidating the conditions under which the epidemic spreads. Concomitant information can theoretically be incorporated into any of the spatio-temporal autocorrelation, kriging, and logistic regression models.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

In the autologistic model presented here, the log odds of disease presence in a particular quadrat, also called the logit,

$$\text{logit}(p) = \ln\left(\frac{\Pr(\text{disease present})}{\Pr(\text{disease absent})}\right),$$

is modeled as a linear combination of soil water content and pathogen population density in the quadrat, and disease presence in neighboring quadrats. In the next section we describe the autologistic model with covariates in more detail. In Section 3 we briefly discuss available methods of estimation of the parameters. Section 4 demonstrates the fitting and interpretation of this model for a Phytophthora epidemic in two naturally infested fields of bell pepper, and Section 5 discusses some issues that arise in assessing the fit of the model and the interpretation of the results.

The focus of this article is on demonstrating the use of the autologistic model using methods that are currently available to practitioners. The aim is to demonstrate real applications of the model and interpretation of the parameters. The interpretations are the same regardless of the method of estimation employed.

2. THE AUTOLOGISTIC MODEL

The autologistic model is a simple generalization to spatial data of the standard logistic model for independent binary data. In the standard logistic model for binary data, the log odds of disease are modeled as a linear function of some regressor variables, $X_1, \ldots, X_r$:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_r X_r = \sum_{k=0}^{r} \beta_k X_k, \qquad (2.1)$$

where p is the probability of a "success" (e.g., the probability of disease being present). For a good exposition, see Hosmer and Lemeshow 1989. The standard model is appropriate in any experimental setup in which all observations are independent. For example, if plants were grown in pots and inoculum were introduced into each pot individually, the observations would be independent of each other. In this situation, the standard logistic model (2.1) would be appropriate and the parameters of model (2.1), $\beta_0, \ldots, \beta_r$, could be estimated by maximum likelihood using standard software such as SAS PROC LOGISTIC (SAS Inst. 1996).

In logistic regression models, the regression coefficients quantify the effects of changes in the covariate values on the odds of disease. In model (2.1), the log odds of disease in a quadrat appears as a function of k regressor variables, $X_1, \ldots, X_k$. If X_1 increases by one unit while the other variables remain constant, the log odds of disease increases by β_1 units. The odds ratio for an increase of one unit is defined as

$$\text{odds ratio} = \frac{\text{odds of disease if } X_1 = x + 1}{\text{odds of disease if } X_1 = x}.$$

Thus, the odds ratio is e^{β_1}. For example, $\beta_1 = .4055$ corresponds to the odds ratio $e^{\beta_1} = 1.5$ for increasing X_1 by one unit, and the odds of disease increase 50% for every unit increase in X_1.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

In a disease epidemic in an agricultural field, we would expect quadrats to be spatially correlated. The autologistic model incorporates the spatial autocorrelation by conditioning the probability of occurrence of disease in a quadrat on disease presence in neighboring quadrats. The definition of neighboring quadrats is very flexible and can be tailored to the particular situation under study. For rectangular lattices there are some standard systems of neighbors. A first-order system includes only the four adjacent quadrats in the set of neighbors, two within the row and two in adjacent rows; a second-order system includes the four diagonal neighbors in addition to the quadrats of the first-order system; a third-order system includes quadrats two rows or columns away (Besag 1974):

First-order Second-order Third-order

However, it is not necessary to use the standard neighborhood systems and it is not necessary that the lattice have any regular shape. What is required is that a set of neighbors be defined for each quadrat in the lattice, and if quadrat i is a neighbor of quadrat j, the converse is also true.

For binary data, if the response at site i depends in a pairwise fashion on the observed number of neighbors with disease present and on r covariates, then the conditional probability of a particular response, $y_i = 1$ (disease present) or $y_i = 0$ (disease absent), is

$$\Pr(Y_i = y_i | x_i, y_j, j \in N_i) = \frac{\exp\{\sum_{k=0}^{r} \beta_k x_{ik} y_i + \sum_{j \in N_i} \gamma_j y_i y_j\}}{1 + \exp\{\sum_{k=0}^{r} \beta_k x_{ik} + \sum_{j \in N_i} \gamma_j y_j\}}, \tag{2.2}$$

where the set of neighbors of the ith site is denoted N_i. Since y_i takes the value 1 if disease is present, the log of the odds of disease being present is:

$$\text{logit}(\Pr(Y_i = 1) | x_i, y_j, \text{for } j \in N_i) = \sum_{k=0}^{r} \beta_k x_{ik} + \sum_{j \in N_i} \gamma_j y_j,$$

which looks just like the standard logistic model (2.1), with the addition of terms for disease in neighboring quadrats (Y_j). For example, an autologistic model with two soil covariates X_1 and X_2, and first-order spatial dependence where the spatial dependence is the same magnitude down rows and across rows, would be

$$\text{logit}(\Pr(Y_i = 1) | x_{i1}, x_{j2}, s_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma_1 s_i \,,$$

where $s_i = \sum_{j \in N_i} y_j$ is the sum of the disease values in the four neighbors. In this model the parameters β_1 and β_2 quantify the effects of the covariates given the disease status of the neighbors. For example, if soil moisture is a covariate, its parameter would measure the log of the increase in odds of disease that was due to increasing moisture after accounting for the effect of disease in any neighboring quadrats. This type of model has flexibility in that neighbors may be defined in any way that makes sense, the observations

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

do not necessarily need to be taken on a rectangular lattice, and covariates (X_{ik}) can be incorporated. The covariate terms in the autologistic model are not necessarily required to enter in a linear fashion.

What is the effect of fitting model (2.1) when in fact the autologistic model is appropriate? We can gain some insight by looking at a linear autoregressive model for continuous response variables, which is analogous to the autologistic model for binary response variables. In the linear conditional autoregressive model, $\mathbf{y} = \gamma\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{W}$ is a symmetric matrix defining neighbor relationships and $\mathbf{W}\mathbf{y}$ represents the response in neighboring quadrats, the effect of ignoring the neighbors depends on whether the covariates $\mathbf{X}$ are spatially correlated. The ordinary least squares estimator obtained by regressing $\mathbf{y}$ on $\mathbf{X}$ is $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + \gamma(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}$. For any given covariate matrix $\mathbf{X}$, this has conditional expectation $\boldsymbol{\beta} + \gamma(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{W}(\mathbf{I} - \gamma\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}$. The magnitude of the second term depends on the spatial pattern of $\mathbf{X}$. If each covariate is set up in a spatially balanced design so that $\mathbf{W}\mathbf{X} = \mathbf{0}$, the second term is 0. On the other hand, if the covariates are spatially correlated, the effect of regressing on $\mathbf{X}$ and ignoring the spatial dependence is usually to produce estimates of the effects of the covariates that are too large, because some of the effect that is due to disease in neighboring quadrats is attributed to the covariates. We see a similar effect when fitting autologistic regression models. A small simulation demonstrating this type of effect is presented in Section 5.

Autologistic models for infinitely large lattices exhibit an unusual feature, called "long-range dependence," when the spatial dependence parameters, γ, are large. Under long-range dependence, the correlation between two sites decreases with distance but never becomes zero, no matter how far apart the sites are. Infinitely large lattices that exhibit long-range dependence consist almost entirely of either zeroes or ones, with a speckling of the other. This feature of autologistic models has made them useful for modeling physical processes that undergo phase transition, such as magnetism and crystallization.

This property of long-range dependence occurs when the spatial dependence parameter(s) exceed a particular "critical value." Critical values have been derived for the simplest autologistic models—those with 50:50 odds of zeroes to ones, no covariates, and a first-order system of neighbors (Pickard 1977). For example, for the model logit $(\Pr(Y_i = 1 \mid s_i) = \alpha + \gamma s_i$, where s_i is the sum of the four neighbors, a field exhibits long-range dependence if the average correlation between neighboring sites, ρ, is .71 or higher. This corresponds to $\gamma = 1.76$ and $\alpha = -2\gamma$. For a similar model—which allows the spatial dependence to be different in two directions, with s_1 = sum of two within-row neighbors and s_2 = sum of two adjacent-row neighbors, logit $\Pr(Y_i = 1 \mid s_{i1}, s_{i2}) = \alpha + \gamma_1 s_{i1} + \gamma_2 s_{i2}$ and $\alpha = -(\gamma_1 + \gamma_2)$—some pairs of critical correlations and parameter values are given in Table 1 (computed based on Pickard 1977). The restriction that $\alpha = -(\gamma_1 + \gamma_2)$ ensures that the overall odds of zeroes to ones is 50:50. It is unclear whether phase transition occurs if $\alpha \neq (\gamma_1 + \gamma_2)$ or covariates are included in the model.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

Table 1. Critical Values Above Which Long Range Dependence Occurs in the Model logit $\Pr(Y_i = 1 \mid s_{i1}, s_{i2}) = \alpha + \gamma_1 s_{i1} + \gamma_2 s_{i2}$ with $\alpha = -(\gamma_1 + \gamma_2)$

Spatial autocorrelations		*Spatial dependence parameters*	
ρ_1	ρ_2	γ_1	γ_2
.64	.89	.40	4.64
.65	.81	.81	3.23
.67	.76	1.21	2.45
.71	.71	1.76	1.76

3. METHODS OF ESTIMATION

The method of choice for estimating parameters of the ordinary logistic regression model is maximum likelihood. In the autologistic model for spatially correlated responses, the observations are not independent and it is not possible to write the likelihood function in closed form. A new method of estimation, Markov chain Monte Carlo maximum likelihood (MCML), has recently been developed which approximates the likelihood function using Monte Carlo sampling (Geyer 1991). Monte Carlo maximum likelihood provides estimates that are consistent and asymptotically normal; however, special conditions are necessary for asymptotic normality to hold when the spatial dependence parameters are at the critical values (Gidas 1993). Currently, however, commercial software is not available for computing Monte Carlo maximum likelihood estimates.

In this article, all parameters have been estimated using a method called maximum pseudolikelihood. Besag (1975) coined the term "pseudolikelihood" (PL) for a function that would be the likelihood were the data independent, the product of the conditional probabilities in expression (2.2). Maximization of the pseudolikelihood function as though it were a true likelihood function is simple to implement because it just involves fitting the autologistic model using standard software for ordinary logistic regression (Strauss 1992). For example, to fit an autologistic model with covariates X_1 and X_2, and first-order spatial dependence that is the same magnitude within and across rows, the procedure is simply to compute the sum of four neighbors for each site, $s_i = \sum_{j \in N_i} y_j$, and then fit the model logit $(\Pr(Y_i = 1 \mid x_{i1}, x_{i2}, s_i)) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma s_i$ using a program such as SAS PROC LOGISTIC. Note, however, that the standard errors that are printed out by ordinary logistic regression software are not appropriate for correlated data, and the usual likelihood ratio-type statistics do not have asymptotic Chi-square distributions (Graham 1995). We present a method of computing parametric bootstrap standard errors in Section 4.

Parameter estimation for the autologistic model is more difficult when the spatial dependence parameters approach the critical values than if the spatial dependence is smaller. Both pseudolikelihood and MCML estimates are consistent regardless of whether the parameters are at the critical value or not (Geman and Graffigne 1986; Gidas 1993), but near the critical value, variability of pseudolikelihood estimates can be much higher than MCML estimates (Geman and Graffigne 1986; Geyer 1991). If the within-row or across-row correlations are near or above these values, pseudolikelihood estimates are not recommended. Otherwise, pseudolikelihood estimates are almost as efficient as

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

MCML estimates (Odencrantz 1988; Geyer 1991; Graham 1995). The big advantage of pseudolikelihood estimation is its computational simplicity and availability, with little sacrifice of precision.

4. APPLICATION TO PHYTOPHTHORA EPIDEMIC

4.1 BACKGROUND

The pathogen *Phytophthora capsici* Leonian causes lesions on the crown, stem, and leaves of bell pepper, and rapidly causes the plant to die. Ristaino et al. (1993, 1994) and Larkin et al. (1995) described the spatial pattern of Phytophthora epidemics in six naturally infested commercial bell pepper fields. Ristaino et al. (1994) found that the spatial correlations of disease symptom types down rows and across rows contain information about whether the disease is spread by water, root-to-root contact, or aerially. They demonstrated that disease generally tends to form longer clusters along rows than across rows, and from this concluded that movement of surface water within furrows is important in spreading inoculum. The spatial and temporal order in which wilt occurs and stem and crown lesions develop provide further clues to the methods of dispersal (Ristaino et al. 1994). It appears that root infections spreading to the crown are the most frequent paths of infection. Stem, leaf, and fruit infections were much rarer or nonexistent, indicating that splash dispersal was not as important for spreading inoculum in the field studied. Larkin et al. (1995) presented spatial correlograms and crosscorrelograms for disease severity and soil variables in the Phytophthora study. The correlograms show that the distance over which quadrats are correlated increases steadily over time and reveal some correspondence between disease and the soil variables.

Soil water content measurements and leaf disk assays of soil pathogen population were collected for two of the naturally infested commercial bell pepper fields that were studied by Ristaino et al. (1993) (described in detail therein). Each field was a square lattice of 20 rows by 20 quadrats with 2 to 3 bell pepper plants per quadrat. The response variable of interest was presence or absence of disease in a quadrat. If any plant was wilted, dead, or had lesions on stem, crown, or leaves, disease was considered to be present in the quadrat. Disease presence or absence was recorded for each quadrat on nine dates throughout the growing season, from 6/16/92 to 8/5/92. Soil water content (%) was measured in each quadrat of field 1 on 7/2/92 and field 2 on 6/22/92. The soil pathogen population, assayed as the number of leaf disks colonized out of five, was measured in each quadrat on two dates: 6/29/92 and 7/29/92 for field 1, and 6/19/92 and 8/5/92 for field 2.

For one of the fields (field 2), the initial pattern of disease appeared to follow the soil water content pattern (Fig. 2). For the other field (field 1), no such patterns are obvious from the maps of soil water content and number of leaf disks colonized (Fig. 1). The patterns of soil water content were quite different in these two fields. Field 2 had a distinct wet corner and disease was present in most of the quadrats in this corner from the first sampling date. Field 1 was wetter overall (mean water content = 10.8%, compared to field 2 mean water content = 8.8%) but more homogeneous, with no distinct wet or dry areas (field 1 std. dev. = 1.82, field 2 std. dev. = 2.39).

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

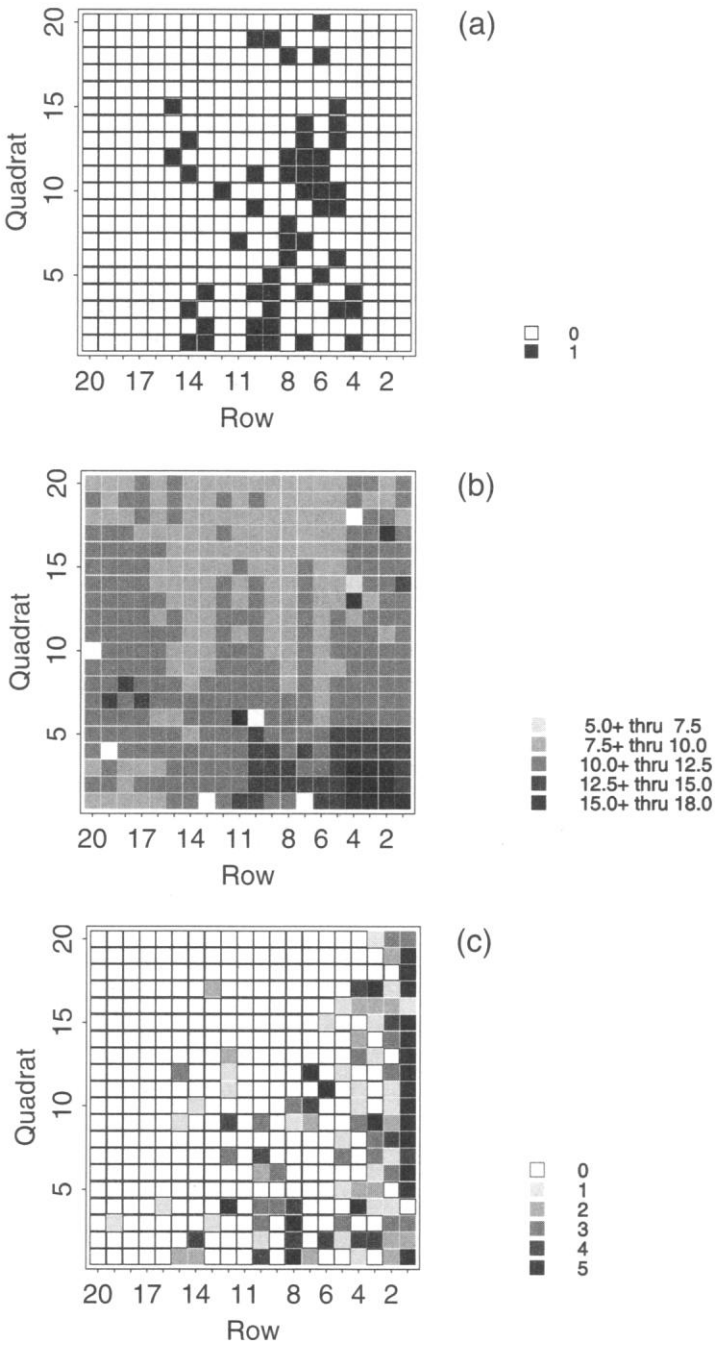

Figure 1. Field 1 (a) Disease Incidence 7/6/92, (b) Soil Moisture (%) 7/2/92, and (c) Leaf Disk Assay 6/29/92.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

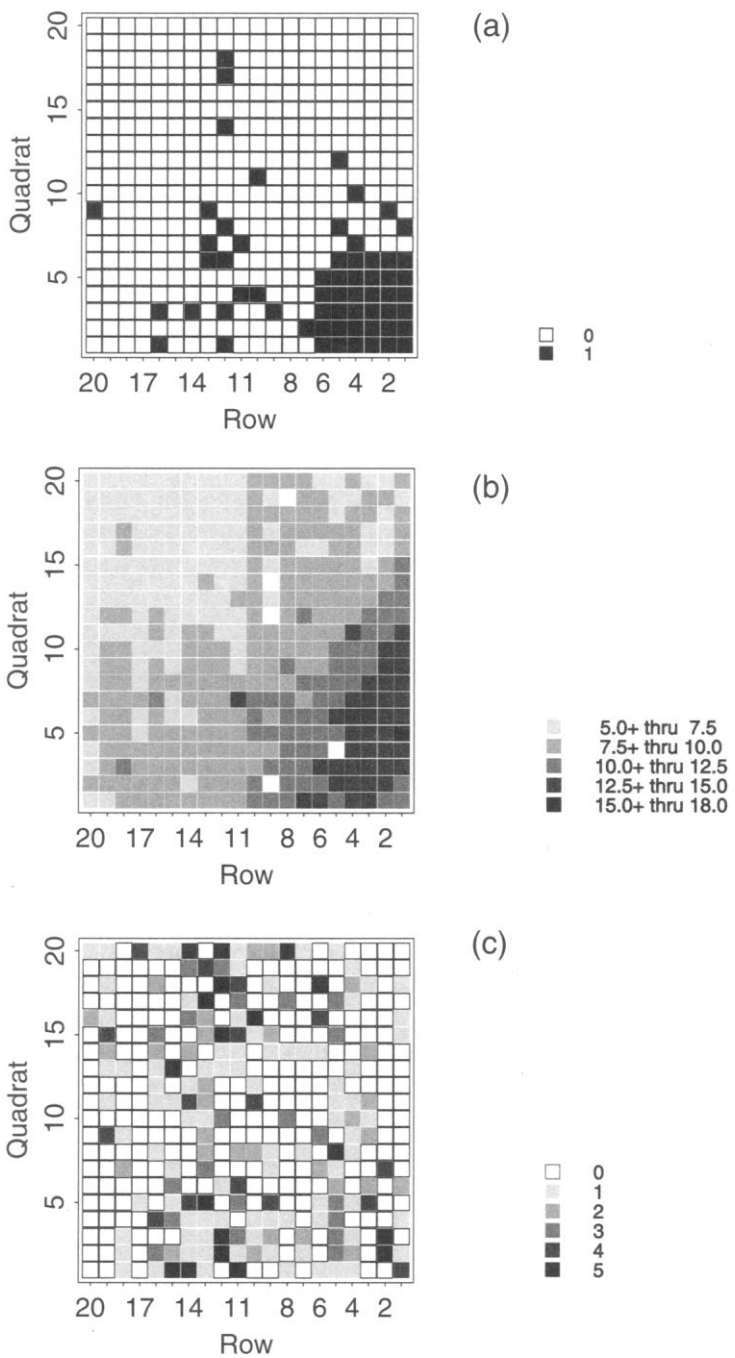

Figure 2. Field 2 (a) Disease Incidence 6/25/92, (b) Soil Moisture (%) 6/22/92, and (c) Leaf Disk Assay 6/19/92.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

4.2 MODELS

For each field we fit three models to the data for the first date after soil water content was measured, 7/6/92 and 6/25/92, respectively. The first model (MODEL1) is a logistic regression model of the effects of soil water content and soil pathogen density on disease presence, ignoring spatial correlation. This is a preliminary model used to check whether the soil covariates alone can explain the spatial patterns. If spatial correlation is still present in the residuals after fitting MODEL1, we conclude that the soil covariates alone are not sufficient to account for the observed spatial variability. In some settings, spatial correlations can be completely eliminated by regression on covariates. In the present application, however, disease is actually spread from one plant to another, so it is likely that, even after taking the soil variables into account, the disease status of the neighboring quadrats could be an important predictor of disease presence. The second model (MODEL2) is a second-order autologistic model with soil water content and the first leaf disk assay as covariates. This model is constructed by adding terms for disease in adjacent quadrats within the row (W), in adjacent rows (A), and diagonally (D) to MODEL1. The third model (MODEL3) is a pure autologistic model without covariates. Under MODEL3 predictions are based solely on disease in the neighboring quadrats; this model is used to check whether the soil covariates can be dropped from MODEL2.

The three models are:

$$\text{MODEL1: } \operatorname{logit}(p_{ij}) = \beta_0 + \beta_1 M_{ij} + \beta_2 L_{ij}, \tag{4.1}$$

$$\text{MODEL2: } \operatorname{logit}(p_{ij}) = \beta_0 + \beta_1 M_{ij} + \beta_2 L_{ij} + \gamma_1 W_{ij} + \gamma_2 A_{ij} + \gamma_3 D_{ij1} + \gamma_4 D_{ij2}, \tag{4.2}$$

$$\text{MODEL3: } \operatorname{logit}(p_{ij}) = \beta_0 + \gamma_1 W_{ij} + \gamma_2 A_{ij} + \gamma_3 D_{ij1} + \gamma_4 D_{ij2}. \tag{4.3}$$

In the three models, M = soil water content, L = number of leaf disks colonized in June, and the subscripts i and j indicate row and quadrat, respectively. The numbers of diseased neighbors are indicated by $W_{ij} = Y_{i,j-1} + Y_{i,j+1}$, the number of diseased quadrats of the two adjacent quadrats within the same row; A_{ij} = number of diseased quadrats of the two adjacent quadrats in neighboring rows; D_{ij1} = number of diseased quadrats of the two diagonal quadrats in the (1,1) and (−1,−1) direction; and D_{ij2} = number of diseased quadrats of the two diagonal quadrats in the (−1,1) and (1,−1) direction. The types of neighbors of site T_{ij} are diagrammed in Table 2; notice that the rows are numbered from right to left to match the actual field layout.

Including four separate terms for neighbors allows us to examine whether correlations across rows are as strong as those within rows and whether there are any diagonal

Table 2. Types of Neighbors of Site T_{ij}

		Row		
		$i+1$	i	$i-1$
	$j+1$	D_{ij1}	W_{ij}	D_{ij2}
Quadrat	j	A_{ij}	T_{ij}	A_{ij}
	$j-1$	D_{ij2}	W_{ij}	D_{ij1}

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

Table 3. Field 1 on 7/6/92: Parameter Estimates and Proportion of Quadrats Misclassified From Fitted Models

Model	Intercept	Soil water content	6/29 leaf disk	Within-row disease	Across-rows disease	Diagonal (1,1)	Diagonal (−1,1)	Missclass.†
MODEL1	−2.70	.0918	.178					42/254
MODEL2	−3.91	.103	.0646	1.33	−.739	.805	1.08	40/254
MODEL3	−2.81			1.36	−.670	.730	1.09	40/256

† Total number of quadrats is 16 × 16 = 256. Models that involve soil water content were fitted to 254 quadrats because 2 water content values were missing from the inner 16 × 16 lattice.

gradients in the field. If we thought that there was a gradient in a different direction, such as the (1,2) direction, we could add terms to the model to capture the expected pattern.

All models were fit to the inner 16 × 16 lattice of 256 quadrats so that models involving adjacent quadrats and quadrats two spaces away could be accommodated. In each field, one quadrat had a measured water content value greater than 25%, which was much higher than the surrounding water content values, and so was omitted from all regression and correlation computations. In addition, in field 1, five water content values were missing, and in field 2, four water content values were missing.

Spatial correlations were computed for the original data and for Pearson residuals after fitting each model. The lag one autocovariance between adjacent quadrats within a row is computed as the covariance between the "tail" and "head" variables, where the tail variable is the response in quadrats 3 through 17 and the head variable is the response in quadrats 4 through 18:

$$C(1) = \frac{1}{16 \cdot 15} \sum_{i=3}^{18} \sum_{j=3}^{17} y_{ij} y_{i,j+1} - \left(\frac{1}{16 \cdot 15} \sum_{i=3}^{18} \sum_{j=3}^{17} y_{ij}\right)\left(\frac{1}{16 \cdot 15} \sum_{i=3}^{18} \sum_{j=4}^{18} y_{ij}\right).$$

The spatial autocorrelation is computed by dividing $C(1)$ by the product of the "head" and "tail" standard deviations. Pearson residuals,

$$\chi_{ij} = \frac{y_{ij} - \hat{p}_{ij}}{\sqrt{\hat{p}_{ij}(1 - \hat{p}_{ij})}},$$

Table 4. Field 2, on 6/25/92: Parameter Estimates and Proportion of Quadrats Misclassified From Fitted Models

Model	Intercept	Soil water content	6/19 leaf disk	Within-row disease	Across-rows disease	Diagonal (1,1)	Diagonal (−1,1)	Missclass.†
MODEL1	−9.10	.706	.460					23/253
MODEL2	−6.15	.295	.387	.292	.226	1.20	.837	23/253
MODEL3	−3.26			.550	.199	1.39	1.22	22/256

† Total number of quadrats is 16 × 16 = 256. Models that involve soil water content were fitted to 253 quadrats because 3 water content values were missing from the inner 16 × 16 lattice.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

Table 5. Parametric Bootstrap Standard Errors of Pseudolikelihood Estimates for Field 1 on 7/6/92, MODEL3; and Field 2 on 6/25/92, MODEL2

	Intercept	*Soil water content*	*Leaf disk assay*	*Within-row disease*	*Across-row disease*	*Diagonal (1,1)*	*Diagonal (−1,1)*
Field 1 SE†	.40			.55	.69	.53	.54
Field 1 naive est.	.32			.32	.42	.36	.30
Field 2 SE††	1.71	.20	.18	.75	.78	.65	.60
Field 2 naive est.	1.35	.14	.16	.47	.50	.43	.46

NOTE: The naive standard errors were computed in SAS PROC LOGISTIC.

† Standard deviation of pseudolikelihood estimates from 487 Monte Carlo runs; 13 runs did not converge.

†† Standard deviation of 500 Monte Carlo runs.

are standardized differences between the observed response ($Y = 1$ for disease present, $Y = 0$ otherwise) and the predicted probability of disease.

The parameter estimates $\hat{\beta}$, and the percent of quadrats misclassified for each fitted model are shown in Tables 3 and 4. For any given quadrat, if $\hat{p}_{ij} > .5$, disease was predicted to be present. The misclassification rate is the proportion of quadrats for which the predictions do not match the disease status actually observed.

4.3 RESULTS

Before doing any regression, disease incidence showed moderate correlations ($r = .30$ and $r = .38$ in fields 1 and 2, respectively) between adjacent quadrats within a row on the last sampling date. After fitting the logistic model for the effects of soil water content and soil pathogen population level, ignoring spatial correlations (MODEL1), the Pearson residuals showed correlations between neighboring quadrats within a row of .27 and .12 for fields 1 and 2, respectively. The reduction in correlation from .38 to .12 in field 2 indicates that a large part of the spatial correlation in disease incidence may be attributable to the environmental variables of soil water content and Phytophthora population level. In field 1, on the other hand, the correlation between adjacent quadrats does not appear to be explained by the soil variables.

Parameter estimates for MODEL1, MODEL2, and MODEL3 are given in Tables 3 and 4, and standard errors are in Table 5. In field 1, substantial within-row dependence was seen and some diagonal trends across the field may also have been present. The estimated odds of disease were nearly four times higher if one neighboring quadrat within the row was diseased than if the two neighbors were disease-free (MODEL3, Tables 3 and 5, $\hat{\gamma}_1 = 1.36, s(\hat{\gamma}_1) = 0.55$, odds ratio $e^{\hat{\gamma}_1} = 3.9$), with all other variables held constant. The soil water content and leaf disk data did not appear to be helpful in predicting disease presence or absence, as the odds ratios for water content and leaf disk assay were both close to one ($\hat{\beta}$ close to 0) in MODEL2.

In field 2 there was a clear visual correspondence between the maps of soil water content and disease incidence (Fig. 2), with the southeast corner having both high soil water content and high disease incidence. On 6/25/92, the estimated odds of disease

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

Table 6. Field 2: Parameter Estimates for MODEL2 on Three Sampling Dates

Date	Intercept	6/22/92 water content	6/19/92 leaf disk assay	8/5/92 leaf disk assay	Within-row disease	Across-rows disease	Diagonal (1,1)	Diagonal (−1,1)
6/25/92	−6.28	.30	.39	.13	.26	.26	1.23	.79
7/13/92	−4.42	.13	.35	.11	.84	−.55	.95	1.41
8/04/92	−4.51	.11	.17	.49	.33	1.10	.22	1.50

increased 337% with an increase of 5 percentage points in soil water content (Tables 4 and 5, $\hat{\beta}_1 = .295$, $s(\hat{\beta}_1) = .20$, $e^{5\hat{\beta}_1} = 4.37$). The estimated odds of disease was 50% higher if one leaf disk was colonized and nearly six times higher if five leaf disks were colonized than the odds if no leaf disks were colonized. At the 6/25 sampling date, the soil population assay had the largest effect relative to its standard error. All the spatial parameter estimates except for the diagonal (1, 1) direction were either smaller than their standard errors or similar in magnitude (Table 5).

A second-order autologistic model with three covariates was fitted to the field 2 data for two sampling dates in addition to 6/25/92: 7/13/92 and 8/4/92 (Table 6). The relationship with water content showed up most strongly on 6/25/92, which was just three days after the soil water content measurements were taken, and decreased as time went on. The relationship between disease and the soil pathogen level in the 6/19/92 assay, as indicated by the number of leaf disks colonized, was fairly strong early in the season and weak late in the season. At the end of the season, the leaf disk assay of 8/5/92 was a better predictor of disease than the early leaf disk assay, and the odds of disease were estimated to increase about 60% if the number of leaf disks colonized was increased by one. By the end of the season, 42% of quadrats were diseased, the spatial dependence was quite high, and the number of diseased neighbors in any one direction was highly associated with the number in every other direction, making it difficult to estimate the spatial dependence parameters accurately.

4.4 STANDARD ERRORS

Parametric bootstrap standard errors were obtained by generating 500 lattices from the autologistic model with parameters set equal to the pseudolikelihood estimates from field 1, MODEL3 or field 2, MODEL2. The values of soil water content and the leaf disk assay were fixed at the values observed in the field.

Pseudolikelihood estimates were computed for each of the 500 lattices, and the standard deviations of these 500 estimates are reported in Table 5. For comparison, the standard error reported by SAS PROC LOGISTIC is also given. In these two examples, the "naive" estimate produced by standard logistic regression software always underestimates the standard errors for the spatial dependence parameters.

The 500 lattices for each model were generated by Gibbs sampling (e.g., see Gilks et al. 1996). This is computer intensive, but can be done fairly easily in a matrix programming language such as SAS PROC IML. One simple procedure is as follows. Starting

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

with the observed data lattice, cycle through the quadrats, updating each quadrat in turn. To update a particular quadrat, assign it disease status $y = 1$ or $y = 0$, according to the probability function in expression (2.2), using the values of the covariates and the current disease status of the neighbors. One complete sweep has been performed after every quadrat has been updated once. Five hundred and fifty complete sweeps were generated, and the last 500 sweeps were included in the sample. The 500 lattices are correlated, but the sample variance converges to the true variance as the length of the chain increases. This is because Gibbs sampling produces a sequence of lattices and a corresponding sequence of pseudolikelihood parameter estimates that cover the entire range of possible values (called an ergodic Markov chain) in such a way that the contribution of the correlations between lattices becomes negligible as the length of the chain increases [see Cressie (1991) for discussion of ergodicity].

5. DISCUSSION

5.1 EVALUATION OF FIT

The second-order autologistic model was successful in decreasing the spatial autocorrelation at most lag distances in both fields studied (Fig. 3). In field 1, however, it was not very successful in reproducing the spatial pattern of disease (Fig. 4). Only 24% of diseased quadrats were predicted to have disease. The overall percent of quadrats misclassified was 16% regardless of whether the soil variables were included in the autologistic model or not (compare MODEL2 and MODEL3). MODEL2 does a better job of predicting disease for field 2 than for field 1, with an overall misclassification rate of 9%. Examination of the map of misclassified quadrats (Fig. 4) shows that prediction of disease was generally accurate in the southeast corner with high soil moisture and concentration of diseased quadrats, but elsewhere in field 2 the model almost always predicted that quadrats would be healthy when they actually contained diseased plants.

If the autologistic model does not fit a set of data well, it is possible that an autologistic model is still appropriate, but that different covariates or functions of the covariates should be included in the model, or that the coefficients for spatial dependence vary with the soil conditions, or that a higher order neighbor system is needed,. For example, we may wonder whether neighbors two quadrats away (third-order neighbors) should be incorporated into the model. In field 2 there were 23 misclassified quadrats after fitting the autologistic model with covariates (MODEL2). The third-order model with covariates did not have any better predictive ability; it misclassified 25 of the 253 quadrats. In addition, AIC values using pseudolikelihoods from PROC LOGISTIC were computed for models from no spatial dependence terms up to third-order dependence structures with and without the soil covariates. Minimizing the Aikaike information criteria (AIC),

$$\text{AIC} = -2\,\ell n\,\text{likelihood} + 2p,$$

where p is the number of parameters, takes the number of parameters and the number of observations into account while maximizing the likelihood (or in our case, pseudolikelihood) function. The models with the lowest AIC values were MODEL3 for field 1 and MODEL2 for field 2.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

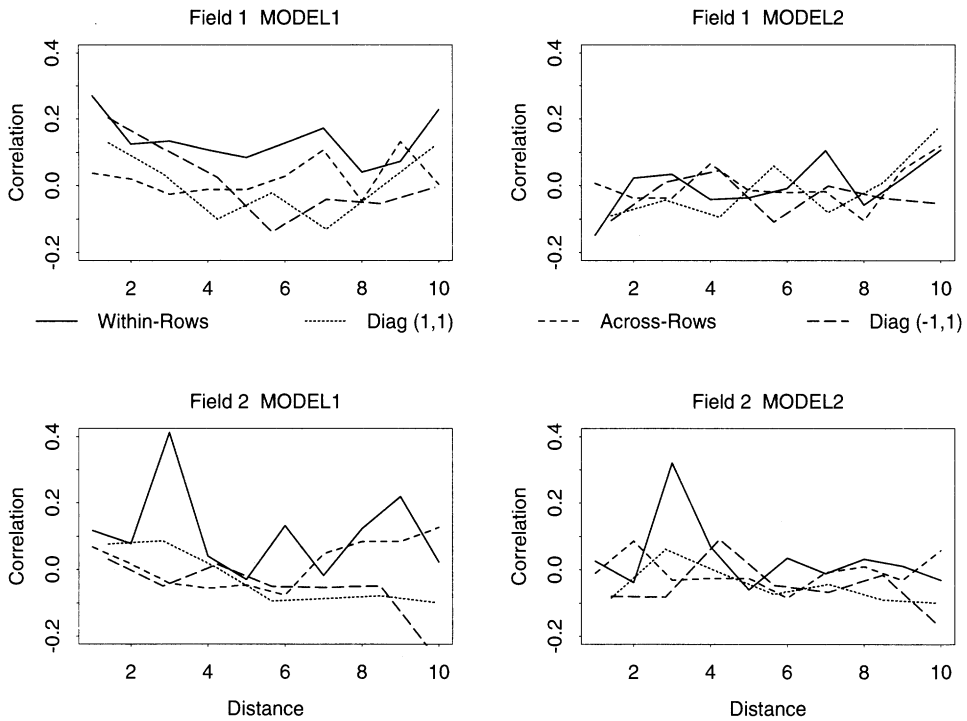

Figure 3. Correlograms of Pearson Residuals.

5.2 CROSSVALIDATION

The misclassification rates and maps of misclassified quadrats presented thus far have been based on the model fitted to the entire set of data. Crossvalidation provides another tool for evaluating the predictive ability of a model, for examining the stability of parameter estimates, and for checking for influential quadrats. For crossvalidation, we refit the autologistic model 256 times, omitting one quadrat along with its neighbors each time, and then predicted disease presence or absence for the omitted quadrat. When fitting MODEL2 for crossvalidation of the ijth quadrat, for example, quadrats $j-1$ to $j+1$ in rows $i-1$ to $i+1$ were omitted, the parameters were estimated, then disease status for quadrat T_{ij} was predicted. Thus, the dataset used to fit the model for prediction of a given quadrat was completely free of the point to be predicted. The numbers of quadrats misclassified by crossvalidation were very similar to the simple misclassification rates reported in Tables 3 and 4: 16% for field 1 MODEL3, and 8% for field 2 MODEL2. The magnitudes of the regression coefficients varied among the crossvalidation runs when different quadrats were omitted, but generally reinforced our conclusions. For example, in field 2 the parameter estimate for soil water content ranged from .20 to .41 for the crossvalidation datasets, whereas it was .30 for the complete dataset (Table 4). The parameter estimates for the effect of disease within a row or across rows were both generally small, either negative or positive, but the estimates for the diagonal terms were consistently positive—in the (1,1) direction they ranged from .61 to 1.83 and in the $(-1,1)$ direction they ranged from .48 to 1.11.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

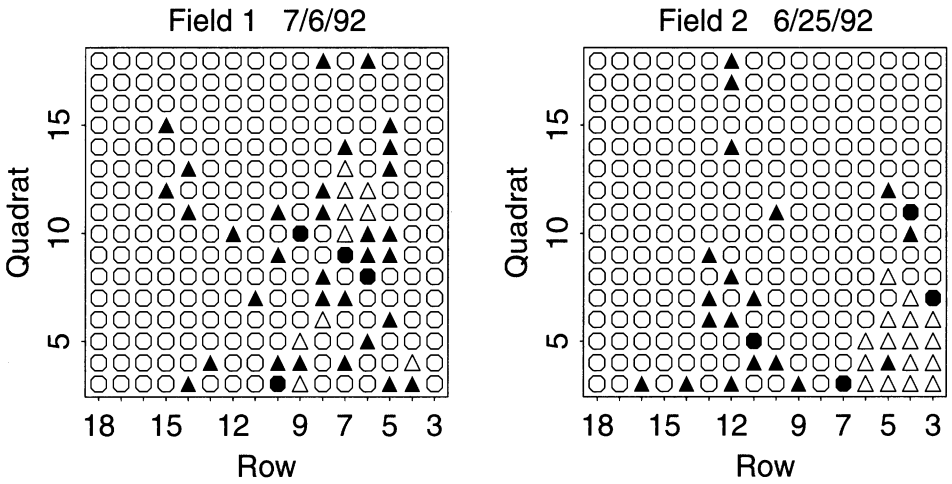

Figure 4. Misclassified Quadrants After Fitting MODEL2. Disease status (actual and predicted, respectively): ○ = Healthy, Healthy; ● = Healthy, Diseased; ▲ = Diseased, Healthy; △ = Diseased, Diseased.

5.3 SPATIALLY CORRELATED COVARIATES AND COLLINEARITY

In MODEL2 for field 2, the estimated relationships between the soil variables and disease are weaker than one might expect, whereas the coefficient for the spatial dependence in the diagonal (1, 1) direction is large (Table 4). When interpreting these results, it is important to keep in mind that the parameters represent the effects of the soil variables *given the disease state of the neighboring quadrats.* The impact of soil moisture after accounting for the effect of diseased neighbors may be very different from the unconditional effect of soil moisture. This type of distinction between interpretation of parameters in models that incorporate spatial (or other) correlation by conditioning on the neighboring responses and models that incorporate a covariance structure directly was discussed in another context by Diggle et al. (1994).

If spatial correlations are ignored in fitting the logistic regression model (MODEL1), the effects of water content and the leaf disk assays appear stronger than if spatial dependence is incorporated as in MODEL2. MODEL1 is clearly not adequate because spatial correlations remain after regressing disease incidence on soil water content and the leaf disk assay alone (Fig. 3). When terms for disease in neighboring quadrats are omitted from the model (MODEL1), the effects of soil water content and leaf disk assays are larger (Table 4). This type of effect in parameter estimates may be expected when the response depends on disease in neighboring quadrats and the covariates are also spatially correlated. In this field there is high correlation between soil water content levels in neighboring diagonal quadrats (Fig. 5). To demonstrate the effect of omitting spatial dependence parameters, MODEL 1, 2, and 3 were fitted to the 500 lattices generated for computation of the standard errors. The average differences between the pseudolikelihood estimate and the true (generated) parameter values are given in Table 7. The parameter for soil water content is overestimated if the spatial dependence parameters are left out (MODEL1), and the spatial dependence parameters are overestimated if the soil covariates are left out (MODEL3).

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

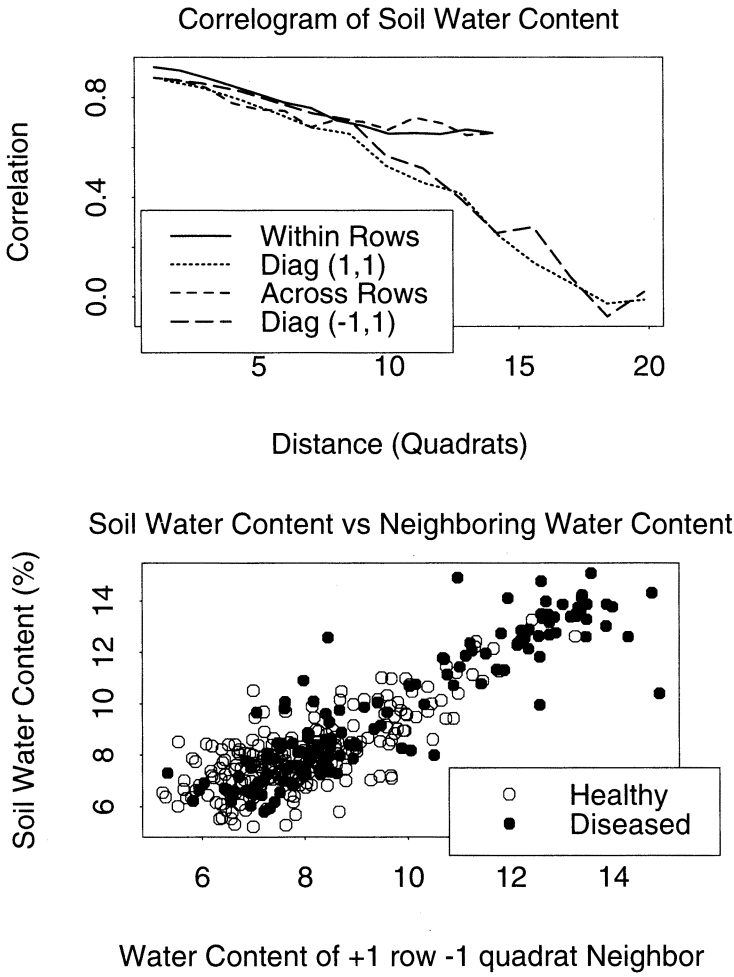

Figure 5. Field 2: Spatial Variability of Soil Water Content.

Collinearity among regressor variables also causes difficulty in estimation and interpretation of the coefficients in logistic regression models, resulting in parameter estimates with large standard errors. In autologistic regression there is the potential for the disease in neighbors in different directions to be highly correlated with each other. This effect can be seen in the results for field 2. Using the measure of association $\hat{\gamma}$, which estimates the difference between the probabilities of concordance and discordance (Agresti 1990), the number of diseased neighbors one quadrat away within the row, $W_{ij} = (y_{i,j-1} + y_{i,j+1})$, is highly associated with the number of diseased neighbors in every other direction—across rows $\hat{\gamma} = .82$, within the row two quadrats away $\hat{\gamma} = .84$, diagonal (1,1) direction $\hat{\gamma} = .73$, and diagonal $(-1,1)$ direction $\hat{\gamma} = .74$. Gamma is analogous to a correlation coefficient, measuring whether two variables such as A_{ij} and W_{ij} tend to increase together. Two quadrats are termed concordant if A_{ij} and W_{ij} are both greater in one quadrat than in the other; they are discordant if the relationship of the two quadrats is different for A_{ij}

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

Table 7. Average Difference Between PL Estimates and True Parameter Values in 500 Simulated Lattices Based on Field 2 Data, MODEL2

Parameter	Intercept	Soil water content	Leaf disk assay	Within-row disease	Across-row disease	Diagonal (1,1)	Diagonal (−1,1)
True value	−6.1544	.2948	.3872	.2924	.2261	1.1952	.8366
MODEL1 bias	−4.45	.594	.040				
MODEL2 bias	− .704	.064	.022	−.010	−.0044	.029	−.019
MODEL3 bias	2.71			.190	.182	.240	.236

than for W_{ij}. Gamma ranges from -1 to 1, taking on the value 1 if the probability of discordance is 0. The standard errors for the spatial dependence parameters were lower for field 1 than for field 2. Correspondingly, the association values between W_{ij} and the numbers of diseased neighbors in other directions are also lower for field 1 than for field 2—across rows $\hat{\gamma} = .59$, two quadrats away within the row $\hat{\gamma} = .71$, diagonal (1,1) $\hat{\gamma} = .44$, and diagonal $(-1,1)$ $\hat{\gamma} = .41$.

6. SUMMARY

Models for binary response variables are not as familiar as models for continuous response variables, and statistical theory and software for them tend not to be as well developed as for models for continuous response variables. Methods of estimation for logistic regression models are rapidly becoming more available, and commercial software should be appearing within the next several years. The aim of this article has been to demonstrate the use and interpretation of one such model for spatial binary data, the autologistic. A second aim has been to explore issues that arise in assessing the fit of the model and its usefulness for studying the pattern of disease in an agricultural field.

In this article, parameters of the autologistic model were estimated using pseudolikelihood estimation with parametric bootstrap standard errors. This is a simple method that provides consistent estimators with good precision, provided that the spatial dependence is moderate. The advantage of this method is that it is easily implemented using existing software. Standard errors were obtained by generating lattices from the fitted model using Gibbs sampling, then computing the empirical standard deviations of parameter estimates from the generated lattices. More precise parameter estimates, as well as test statistics, can be obtained using Monte Carlo maximum likelihood estimation (Geyer 1991; Graham 1995).

Autologistic models lend themselves naturally to studies of disease epidemics. In our study of Phytophthora epidemics, we used one example of a field in which soil moisture is fairly homogeneous across the field and the disease appears to be correlated down rows. This is what the researchers would expect to see if the disease were carried by water moving down rows or by root-to-root contact within the rows. Fitting the autologistic model permits us to estimate that the odds of disease is four times higher if a neighboring quadrat is diseased than if neither neighbor is diseased. In the other field there was a strong moisture gradient in the field, and this was highly associated with disease presence.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

In this case, fitting the autologistic model gave estimates of the increased odds of disease with increased soil water content. In this field, the spatial dependence parameters appeared less important than the effect of soil moisture after accounting for the effect of moisture. The effect of spatial dependence in different directions was harder to sort out in this field than in the first.

In the fitting and interpretation of autologistic models, if the spatial correlation is too high in any direction it becomes difficult to distinguish among the directions. In such cases it may be more feasible to fit some other kind of model for spatially correlated binary data, such as the marginal logistic regression model with spatially correlated errors described by Albert and McShane (1995). We have found the following plots and tables useful for studying whether the fitted model adequately captures the spatial correlation structure: (1) correlograms of the Pearson residuals; (2) measures of association such as $\hat{\gamma}$ for detecting collinearity in the regressor variables; and (3) maps of the misclassified quadrats. Crossvalidation is another useful tool, and the quadrats misclassified in crossvalidation can be mapped and crosstabulated in the same ways as the simple misclassification errors.

APPENDIX

		Field 1			Field 2		
Row	Quadrat	7/6/92 disease status	7/2/92 soil water content	6/29/92 leaf disk assay	6/25/92 disease status	6/22/92 soil water content	6/19/92 leaf disk assay
1	1	0	15.05	5	1	12.05	4
1	2	0	14.32	2	1	12.07	1
1	3	0	13.99	3	1	13.87	0
1	4	0	13.85	0	1	14.14	1
1	5	0	13.16	5	1	15.07	0
1	6	0	11.81	5	1	13.87	2
1	7	0	11.47	4	0	14.23	0
1	8	0	11.19	5	1	14.31	0
1	9	0	11.94	5	0	13.75	0
1	10	0	11.20	5	0	13.77	0
1	11	0	10.73	5	0	12.62	0
1	12	0	11.25	5	0	12.50	0
1	13	0	11.64	5	0	11.87	0
1	14	0	13.84	5	0	10.12	0
1	15	0	11.04	5	0	10.16	1
1	16	0	10.73	1	0	8.86	1
1	17	0	10.59	5	0	8.38	0
1	18	0	9.69	5	0	8.38	1
1	19	0	10.32	5	0	7.48	0
1	20	0	10.33	3	0	7.72	0
2	1	0	16.60	2	1	11.25	0
2	2	0	15.58	2	1	13.02	5
2	3	0	15.99	3	1	13.39	5
2	4	0	14.56	1	1	13.57	0
2	5	0	12.74	0	1	13.87	0
2	6	0	11.70	3	1	13.40	0
2	7	0	11.29	1	0	14.75	4
2	8	0	10.82	4	0	13.32	0
2	9	0	10.26	2	1	13.98	0
2	10	0	11.16	1	0	13.26	0
2	11	0	9.64	1	0	12.21	0
2	12	0	10.15	3	0	11.13	0
2	13	0	10.30	0	0	10.65	0
2	14	0	10.08	3	0	9.15	0

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

2	15	0	10.20	4	0	8.05	0
2	16	0	10.67	2	0	7.46	0
2	17	0	15.49	1	0	7.09	0
2	18	0	10.38	0	0	7.34	0
2	19	0	9.89	2	0	8.03	0
2	20	0	9.98	3	0	7.05	0
3	1	0	16.61	0	1	13.86	0
3	2	0	15.13	4	1	13.17	0
3	3	0	15.20	0	1	13.37	0
3	4	0	13.82	1	1	13.49	0
3	5	0	14.12	2	1	13.28	4
3	6	0	11.35	1	1	12.60	2
3	7	0	11.54	3	0	12.63	0
3	8	0	10.99	3	0	12.69	0
3	9	0	10.56	5	0	12.42	2
3	10	0	10.78	0	0	11.31	1
3	11	0	10.83	0	0	11.26	2
3	12	0	10.87	0	0	10.39	0
3	13	0	9.70	1	0	8.94	0
3	14	0	10.28	0	0	8.83	2
3	15	0	12.03	1	0	7.29	0
3	16	0	10.91	2	0	7.28	0
3	17	0	11.33	5	0	7.40	0
3	18	0	10.17	0	0	7.55	1
3	19	0	9.96	0	0	7.53	0
3	20	0	10.41	1	0	6.81	0
4	1	1	16.14	1	1	12.76	1
4	2	0	15.96	4	1	12.85	2
4	3	1	15.56	0	1	12.60	1
4	4	1	14.58	5	1	13.48	0
4	5	0	13.37	2	1	14.29	1
4	6	0	12.07	0	1	12.55	0
4	7	0	11.17	0	1	12.76	0
4	8	0	11.75	0	0	11.33	1
4	9	0	10.62	3	0	11.88	1
4	10	0	11.94	1	1	11.78	2
4	11	0	11.50	1	0	14.90	1
4	12	0	9.72	0	0	9.91	0
4	13	0	15.02	3	0	9.74	1
4	14	0	6.15	2	0	8.62	0
4	15	0	10.31	0	0	7.74	0
4	16	0	10.45	2	0	7.55	0
4	17	0	10.20	4	0	7.66	1
4	18	0	.	0	0	7.79	2
4	19	0	10.63	0	0	7.14	1
4	20	0	10.01	0	0	7.74	0
5	1	0	14.87	0	1	12.21	1
5	2	0	14.33	1	1	13.47	3
5	3	1	14.59	3	1	12.73	3
5	4	0	13.71	0	1	26.00	3
5	5	0	13.60	1	1	12.28	3
5	6	1	12.09	0	1	12.88	2
5	7	0	11.01	3	0	11.73	1
5	8	0	10.22	1	1	11.13	5
5	9	1	9.94	0	0	10.68	0
5	10	1	10.79	0	0	10.97	1
5	11	0	10.63	0	0	9.66	1
5	12	0	10.16	1	1	9.86	1
5	13	1	9.19	0	0	8.51	0
5	14	1	9.04	0	0	8.21	0
5	15	1	9.67	0	0	8.14	3
5	16	0	9.00	1	0	8.23	0
5	17	0	9.07	0	0	7.61	3
5	18	0	8.32	0	0	7.33	0
5	19	0	8.86	0	0	7.17	0
5	20	0	8.41	0	0	6.94	1
6	1	0	13.11	0	1	13.26	1
6	2	0	13.05	4	1	11.82	0
6	3	0	12.40	0	1	14.11	0
6	4	0	11.03	0	1	12.13	1
6	5	1	10.41	0	1	12.36	1
6	6	0	9.65	0	0	10.71	0

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

6	7	0	9.09	0	0	10.77	2
6	8	0	9.48	0	0	10.04	0
6	9	1	8.96	0	0	9.68	3
6	10	1	9.05	0	0	9.60	0
6	11	1	9.53	5	0	9.15	0
6	12	1	8.38	0	0	9.08	0
6	13	0	8.58	0	0	8.60	0
6	14	0	8.80	0	0	8.92	1
6	15	0	8.55	1	0	8.14	0
6	16	0	8.44	0	0	7.39	4
6	17	0	8.01	0	0	7.54	0
6	18	1	8.75	0	0	7.56	5
6	19	0	8.72	0	0	7.72	0
6	20	1	8.18	0	0	7.20	0
7	1	1	.	2	0	12.57	0
7	2	0	14.73	0	1	11.95	1
7	3	0	12.40	0	0	12.35	2
7	4	1	12.57	0	0	11.21	0
7	5	0	11.42	0	0	10.89	0
7	6	0	11.58	0	0	11.43	0
7	7	1	11.02	0	0	9.41	0
7	8	0	10.69	0	0	10.39	1
7	9	0	10.61	2	0	9.55	0
7	10	1	10.68	4	0	9.47	0
7	11	1	10.25	0	0	9.59	0
7	12	1	11.03	5	0	10.48	0
7	13	1	10.30	0	0	8.26	0
7	14	1	10.15	0	0	9.01	1
7	15	0	10.21	0	0	8.48	0
7	16	0	9.36	0	0	7.10	0
7	17	0	9.38	0	0	7.76	0
7	18	0	8.81	0	0	7.89	0
7	19	0	9.64	0	0	7.82	0
7	20	0	9.86	0	0	7.62	1
8	1	0	11.68	5	0	11.52	1
8	2	0	12.89	4	0	12.14	0
8	3	0	12.63	5	0	10.95	0
8	4	0	12.25	4	0	11.46	1
8	5	0	11.66	0	0	11.01	0
8	6	1	11.19	0	0	10.74	0
8	7	1	10.35	0	0	10.98	0
8	8	1	9.59	0	0	9.98	0
8	9	0	8.91	1	0	10.16	0
8	10	0	9.67	3	0	9.96	3
8	11	1	9.34	0	0	8.71	0
8	12	1	9.81	0	0	8.99	0
8	13	0	9.46	0	0	9.80	0
8	14	0	8.65	0	0	8.90	1
8	15	0	9.77	0	0	9.68	0
8	16	0	9.11	0	0	8.40	0
8	17	0	8.87	0	0	8.32	3
8	18	1	9.40	0	0	8.08	0
8	19	0	8.71	0	0	.	0
8	20	0	8.32	0	0	7.62	5
9	1	1	12.44	0	0	11.67	0
9	2	1	14.07	0	0	.	1
9	3	1	13.51	0	1	10.75	2
9	4	1	14.12	3	0	9.49	1
9	5	1	12.50	0	0	10.04	4
9	6	0	12.01	3	0	9.96	0
9	7	0	10.79	0	0	10.46	1
9	8	0	10.29	0	0	9.42	1
9	9	0	10.50	0	0	8.94	0
9	10	0	9.82	0	0	8.58	0
9	11	0	9.58	0	0	8.53	0
9	12	0	9.33	0	0	.	0
9	13	0	9.24	0	0	7.21	0
9	14	0	9.79	0	0	.	1
9	15	0	9.17	0	0	7.36	2
9	16	0	9.17	0	0	7.93	0
9	17	0	9.41	0	0	6.83	0
9	18	0	9.07	0	0	7.69	1

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

9	19	1	9.34	0	0	7.08	0
9	20	0	8.89	0	0	7.52	2
10	1	1	13.42	4	0	10.12	0
10	2	1	13.41	1	0	10.15	2
10	3	0	12.67	3	0	9.74	1
10	4	1	14.15	3	1	9.81	1
10	5	0	13.53	0	0	10.32	0
10	6	0	.	2	0	10.08	0
10	7	0	12.13	4	0	10.89	0
10	8	0	12.34	0	0	9.94	2
10	9	1	10.84	3	0	9.68	0
10	10	0	11.43	0	0	9.46	0
10	11	1	10.51	0	1	9.65	4
10	12	0	10.73	0	0	9.67	1
10	13	0	10.92	0	0	8.35	0
10	14	0	10.29	0	0	8.48	0
10	15	0	9.90	0	0	8.41	1
10	16	0	10.00	0	0	8.43	5
10	17	0	9.93	0	0	8.51	0
10	18	0	9.69	0	0	7.74	0
10	19	1	10.38	0	0	8.41	0
10	20	0	9.64	0	0	8.00	2
11	1	0	12.57	0	0	8.68	5
11	2	0	12.18	0	0	8.67	1
11	3	0	12.19	0	0	7.60	3
11	4	0	12.24	0	1	9.60	0
11	5	0	11.31	0	0	8.17	3
11	6	0	17.64	0	0	7.97	4
11	7	1	12.25	0	1	12.57	0
11	8	0	11.26	0	0	7.62	2
11	9	0	10.23	0	0	7.03	0
11	10	0	10.21	0	0	7.06	0
11	11	0	9.69	0	0	7.20	0
11	12	0	10.12	0	0	6.19	0
11	13	0	9.57	0	0	7.69	1
11	14	0	9.74	0	0	6.51	1
11	15	0	10.35	0	0	6.08	4
11	16	0	8.97	0	0	5.54	1
11	17	0	9.43	0	0	5.87	3
11	18	0	9.35	0	0	6.64	4
11	19	0	9.55	0	0	6.99	1
11	20	0	9.53	0	0	7.45	1
12	1	0	10.85	0	1	9.22	0
12	2	0	11.74	0	0	7.86	5
12	3	0	12.00	0	1	8.40	5
12	4	0	11.97	5	0	8.10	1
12	5	0	12.03	0	0	8.45	0
12	6	0	10.82	0	1	8.45	1
12	7	0	10.91	3	0	8.61	0
12	8	0	10.57	0	1	9.28	0
12	9	0	10.18	4	0	7.88	0
12	10	1	10.51	0	0	7.98	3
12	11	0	10.36	1	0	7.41	0
12	12	0	10.67	1	0	6.99	0
12	13	0	10.25	2	0	7.01	1
12	14	0	10.18	0	1	6.73	1
12	15	0	8.90	0	0	6.56	5
12	16	0	9.30	0	0	6.46	0
12	17	0	10.04	0	1	6.48	0
12	18	0	9.63	0	1	6.95	5
12	19	0	10.27	0	0	6.64	3
12	20	0	9.36	0	0	6.21	5
13	1	1	.	0	0	8.93	1
13	2	1	11.07	0	0	8.87	1
13	3	0	10.85	1	0	7.86	1
13	4	1	10.85	0	0	7.76	1
13	5	0	11.13	0	0	7.53	5
13	6	0	10.62	0	1	8.40	0
13	7	0	10.94	0	1	8.47	3
13	8	0	10.61	0	0	8.11	2
13	9	0	9.42	0	1	8.68	2
13	10	0	9.96	0	0	8.27	1

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

13	11	0	9.39	0	0	7.58	2
13	12	0	8.98	0	0	6.83	2
13	13	0	8.71	0	0	7.02	1
13	14	0	8.87	0	0	7.51	0
13	15	0	9.87	0	0	7.16	2
13	16	0	9.57	0	0	6.59	2
13	17	0	9.83	2	0	7.12	5
13	18	0	8.50	0	0	5.94	0
13	19	0	8.94	0	0	6.56	4
13	20	0	8.68	0	0	6.10	0
14	1	1	10.37	2	0	8.71	5
14	2	0	10.48	4	0	7.40	1
14	3	1	10.48	0	1	8.37	1
14	4	0	10.05	0	0	8.44	1
14	5	0	9.76	0	0	8.24	4
14	6	0	10.79	0	0	7.47	0
14	7	0	10.05	0	0	7.92	0
14	8	0	9.74	0	0	8.05	0
14	9	0	9.25	0	0	7.52	0
14	10	0	9.10	1	0	7.64	1
14	11	1	9.00	0	0	7.77	4
14	12	0	9.32	0	0	7.21	1
14	13	1	9.01	0	0	6.95	0
14	14	0	8.59	0	0	6.71	2
14	15	0	8.97	0	0	6.62	0
14	16	0	8.86	0	0	6.23	3
14	17	0	9.11	0	0	7.31	1
14	18	0	9.20	0	0	6.92	0
14	19	0	8.70	0	0	6.36	3
14	20	0	8.56	0	0	6.02	5
15	1	0	10.80	2	0	7.92	5
15	2	0	11.49	0	0	9.01	2
15	3	0	10.86	0	0	8.97	0
15	4	0	10.75	0	0	8.26	3
15	5	0	10.31	0	0	7.65	1
15	6	0	9.58	0	0	7.25	3
15	7	0	10.28	0	0	7.27	1
15	8	0	10.18	0	0	8.00	0
15	9	0	9.68	1	0	7.33	0
15	10	0	9.85	0	0	6.85	0
15	11	0	10.65	0	0	7.24	1
15	12	1	10.06	3	0	7.09	0
15	13	0	9.80	0	0	6.43	5
15	14	0	9.66	0	0	6.85	0
15	15	1	9.37	0	0	5.82	0
15	16	0	9.48	0	0	5.33	0
15	17	0	8.87	0	0	6.04	0
15	18	0	10.29	0	0	5.79	0
15	19	0	10.53	0	0	5.53	0
15	20	0	10.88	0	0	5.58	1
16	1	0	9.81	0	1	9.34	1
16	2	0	11.01	0	0	8.21	3
16	3	0	9.84	0	1	9.88	0
16	4	0	10.53	1	0	8.33	4
16	5	0	10.02	0	0	8.30	0
16	6	0	9.52	0	0	8.00	0
16	7	0	10.79	0	0	10.51	0
16	8	0	10.30	0	0	8.08	1
16	9	0	10.79	0	0	7.90	0
16	10	0	10.54	0	0	7.47	0
16	11	0	10.49	0	0	7.87	1
16	12	0	9.72	0	0	7.71	1
16	13	0	11.16	0	0	7.07	1
16	14	0	9.17	0	0	7.20	2
16	15	0	9.67	0	0	6.63	3
16	16	0	10.06	0	0	6.94	0
16	17	0	8.79	0	0	6.58	0
16	18	0	9.00	0	0	6.33	2
16	19	0	9.38	0	0	6.38	0
16	20	0	9.15	0	0	6.55	1
17	1	0	9.75	0	0	8.38	0
17	2	0	10.03	0	0	8.14	0

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

17	3	0	9.95	0	0	8.49	0
17	4	0	10.89	0	0	8.17	0
17	5	0	11.11	0	0	7.46	1
17	6	0	11.23	0	0	7.00	0
17	7	0	13.57	0	0	7.60	0
17	8	0	11.81	0	0	7.18	0
17	9	0	10.86	0	0	6.75	0
17	10	0	11.42	0	0	5.81	0
17	11	0	11.61	0	0	6.40	0
17	12	0	11.11	0	0	6.37	0
17	13	0	12.20	0	0	6.19	0
17	14	0	11.38	0	0	5.88	0
17	15	0	10.58	0	0	6.77	0
17	16	0	10.27	0	0	7.05	0
17	17	0	9.87	0	0	5.69	0
17	18	0	10.03	0	0	5.29	0
17	19	0	10.43	0	0	5.23	0
17	20	0	11.96	0	0	5.92	4
18	1	0	9.98	0	0	8.25	1
18	2	0	9.61	0	0	8.64	1
18	3	0	9.78	0	0	10.06	1
18	4	0	10.70	0	0	8.79	0
18	5	0	11.10	0	0	8.79	0
18	6	0	12.27	0	0	8.37	0
18	7	0	11.34	0	0	9.45	2
18	8	0	13.84	0	0	8.01	0
18	9	0	11.40	0	0	8.66	1
18	10	0	11.72	0	0	7.79	0
18	11	0	11.85	0	0	6.92	0
18	12	0	11.07	0	0	7.77	0
18	13	0	11.78	0	0	6.84	1
18	14	0	10.59	0	0	7.12	0
18	15	0	10.85	0	0	7.37	0
18	16	0	10.12	0	0	7.81	0
18	17	0	10.45	0	0	7.62	0
18	18	0	9.90	0	0	6.99	0
18	19	0	10.49	0	0	7.25	0
18	20	0	9.61	0	0	6.38	0
19	1	0	9.67	0	0	7.39	0
19	2	0	10.54	0	0	7.61	0
19	3	0	10.75	1	0	8.42	0
19	4	0	.	0	0	8.06	0
19	5	0	10.06	0	0	8.59	0
19	6	0	10.98	0	0	8.18	0
19	7	0	13.14	0	0	8.22	0
19	8	0	10.95	0	0	7.74	0
19	9	0	11.15	0	0	8.22	4
19	10	0	11.15	0	0	8.03	0
19	11	0	11.41	0	0	7.41	1
19	12	0	10.53	0	0	8.12	0
19	13	0	10.66	0	0	7.49	1
19	14	0	11.02	0	0	6.17	2
19	15	0	10.43	0	0	6.44	4
19	16	0	10.29	0	0	6.67	1
19	17	0	10.07	0	0	6.52	0
19	18	0	9.76	0	0	6.45	1
19	19	0	9.17	0	0	7.01	0
19	20	0	9.33	0	0	6.83	1
20	1	0	9.17	0	0	7.02	0
20	2	0	10.17	0	0	9.63	0
20	3	0	9.73	0	0	6.92	0
20	4	0	10.19	0	0	6.98	0
20	5	0	10.63	0	0	7.97	0
20	6	0	10.95	0	0	6.73	0
20	7	0	12.20	0	0	7.53	0
20	8	0	12.09	0	0	6.38	0
20	9	0	11.66	0	1	6.73	0
20	10	0	25.84	0	0	7.16	0
20	11	0	11.22	0	0	6.69	0
20	12	0	11.28	0	0	6.34	0
20	13	0	12.11	0	0	6.27	0
20	14	0	11.60	0	0	6.11	0

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

20	15	0	10.55	0	0	5.64	0
20	16	0	11.45	0	0	6.64	2
20	17	0	10.43	0	0	6.56	0
20	18	0	9.45	0	0	5.65	0
20	19	0	10.24	0	0	5.94	0
20	20	0	9.17	0	0	5.59	1

ACKNOWLEDGMENTS

This research was funded in part by grant number 92-37303-7715 from the National Research Initiatives Competitive Grants Program of the U.S. Department of Agriculture. We thank the referees for their thoughtful reviews.

[Received September 1994. Revised April 1997.]

REFERENCES

Agresti, A. (1990), *Categorical Data Analysis,* New York: Wiley.

Albert, P. S., and McShane, L. M. (1995), "A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data," *Biometrics,* 51, 627–638.

Besag, J. (1972), "Nearest-Neighbour Systems and the Auto-Logistic Model for Binary Data," *Journal of the Royal Statistical Society,* Ser. B, 34, 75–83.

——— (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society,* Ser. B, 36, 192–225.

——— (1975), "Statistical Analysis of Non-Lattice Data," *The Statistician,* 24, 179–195.

——— (1977), "Some Methods of Statistical Analysis for Spatial Data," *Bulletin of the International Statistical Institute Proceedings of the 41st Session,* Vol. XLVII, Book 2, 77–92.

Chadoeuf, J., Nandris, D., Geiger, J. P., Nicole, M., and Pierrat, J. C. (1992), "Modelisation Spatio-Temporelle d'une Epidemic Par un Processus de Gibbs: Estimation et Tests," *Biometrics,* 48, 1165–1175.

Cressie, N. A. C. (1991), *Statistics for Spatial Data,* New York: Wiley.

Diggle, P. J., Liang, K., and Zeger, S. L. (1994), *Analysis of Longitudinal Data,* Oxford: Oxford Press.

Geman, S., and Graffigne, C. (1986), "Markov Random Field Image Models and Their Applications to Computer Vision," *Proceedings of the International Congress of Mathematicians,* Berkeley, CA, 1496–1517.

Geyer, C. J. (1991), "Markov Chain Monte Carlo Maximum Likelihood," in *Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface,* pp. 156–163.

Gidas, B. (1993), "Parameter Estimation for Gibbs Distribution From Fully Observed Data," in *Markov Random Fields: Theory and Applications,* eds., R. Chellapa and R. Jain, New York: Academic Press.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996,) *Markov Chain Monte Carlo in Practice,* London: Chapman & Hall.

Graham, J. M. (1995), "Markov Chain Monte Carlo Inference Procedures for Discrete Spatial Lattice Models," unpublished Ph.D. dissertation, North Carolina State University, Raleigh, NC.

Hosmer, D. W., and Lemeshow, S. (1989), *Applied Logistic Regression,* New York: Wiley.

Larkin, R. P., Gumpertz, M. L., and Ristaino, J. B. (1995), "Geostatistical Analysis of Phytophthora Epidemic Development in Commercial Bell Pepper Fields," *Phytopathology,* 84, 191–203.

Lecoustre, R., Fargette, D., Fauquet, C., and de Reffye, P. (1989), "Analysis and Mapping of the Spatial Spread of African Cassava Mosaic Virus Using Geostatistics and the Kriging Technique," *Phytopathology,* 79, 913–920.

Odencrantz, J. R. C. (1988), "Parametric Inference for Binary and Colored Gibbs Lattices," unpublished Ph.D. dissertation, University of California, Riverside, CA.

Pickard, D. K. (1977), "Asymptotic Inference for an Ising Lattice. II," *Advances in Applied Probability,* 9, 476–501.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms

Reynolds, K. M., and Madden, L. V. (1988), "Analysis of Epidemics Using Spatio-Temporal Autocorrelation," *Phytopathology,* 78, 240–246.

Ristaino, J. B., Larkin, R. P., and Campbell, C. L. (1993), "Spatial and Temporal Dynamics of Phytophthora Epidemics in Commercial Bell Pepper Fields," *Phytopathology,* 83, 1312–1320.

Ristaino, J. B., Larkin, R. P., and Campbell, C. L. (1994), "Spatial Dynamics of Disease Symptom Expression During Phytophthora Epidemics in Bell Pepper," *Phytopathology,* 84, 1015–1024.

SAS Institute Inc. (1996), *SAS/STAT Software: Changes and Enhancements Through Release 6.11,* Cary, NC: Author.

Smyth, G. K., Chakraborty, S., Clark, R. G., and Petitt, A. N. (1992), "A Stochastic Model for Anthracnose Development in Stylosanthes Scabra," *Phytopathology,* 82, 1267–1272.

Strauss, D. (1992), "The Many Faces of Logistic Regression," *The American Statistician,* 46, 321–326.

Wu, H., and Huffer, F. (1997), "Modeling the Distribution of Plant Species Using the Autologistic Regression Model," *Environmental and Ecological Statistics,* 4, 49–64.

This content downloaded from 150.131.78.77 on Fri, 13 Mar 2020 16:17:56 UTC
All use subject to https://about.jstor.org/terms