

A Generalized Linear Model Approach to Spatial Data Analysis and Prediction

Author(s): C. A. Gotway and W. W. Stroup

Source: *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 2, No. 2 (Jun., 1997), pp. 157-178

Published by: [International Biometric Society](#)

Stable URL: <http://www.jstor.org/stable/1400401>

Accessed: 29/09/2014 18:13

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

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.



International Biometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Agricultural, Biological, and Environmental Statistics*.

<http://www.jstor.org>

A Generalized Linear Model Approach to Spatial Data Analysis and Prediction

C. A. GOTWAY and W. W. STROUP

The theory of generalized linear models and quasi-likelihood provides a flexible framework for analyzing non-normal data. In this article, we demonstrate how this theory can be extended to include the analysis of discrete and categorical spatial data. This theory can be used to estimate parameters and test treatment effects in a designed experiment involving discrete or categorical spatial responses. It also provides a flexible method for spatial prediction using non-normal data and includes universal kriging and indicator kriging as special cases. Examples are given, including one where the focus is on comparing treatments in a designed experiment in which spatial correlation is present, and two others where spatial prediction or mapping is the desired goal. The methods presented here provide an additional set of tools for the analysis of spatial data that will be useful to researchers in a variety of disciplines, including hydrology, soil science, entomology, agronomy, and ecology.

Key Words: Generalized Linear Models; Geostatistics; Kriging; Quasi-likelihood; Spatial Modeling; Spatial Prediction.

1. INTRODUCTION

The field of geostatistics encompasses a broad methodology for analyzing and modeling spatial data. The work of many applied researchers in demonstrating the utility of geostatistical methods, in conjunction with prolific education and the increasing availability of software, has resulted in the widespread use of these methods in a variety of scientific disciplines, including hydrology, geology, soil science, agronomy, and ecology. Working with scientists in these areas can often bring to light opportunities for novel applications of geostatistical techniques. These can afford a greater understanding of the spatial process under study, but can also present situations where current geostatistical methodology falls short of providing a comprehensive set of models and methods. One such example is the lack of geostatistical methodology for analysis and prediction of discrete and categorical spatial data.

Assuming the data have a Gaussian distribution, geostatistical methods for analyzing spatially correlated data are well known. In the case of designed experiments, estimates

C. A. Gotway is a Mathematical Statistician, Centers for Disease Control and Prevention, MS F62, 1600 Clifton Road NE, Atlanta, GA 30333. W. W. Stroup is Professor, Department of Biometry, University of Nebraska, Lincoln, NE 68583-0712.

©1997 American Statistical Association and the International Biometric Society
Journal of Agricultural, Biological, and Environmental Statistics, Volume 2, Number 2, Pages 157–178

of treatment effects and differences that account for the spatial correlation can be obtained using maximum likelihood or restricted maximum likelihood (Zimmerman and Harville 1991). In the case of spatial prediction, ordinary or universal kriging (Journel and Huijbregts 1978; Cressie 1986) are frequently used.

When the data are non-normal, the situation is much less clear. One approach used for predicting non-Gaussian data is to transform the data to normality, make predictions on the transformed scale, and then back-transform to the original scale. However, transformation back to the original scale either results in biased predictions or requires approximation, except in the case of lognormal kriging (Dowd 1982). Moreover, transformation to normality is really only an option for data on a continuous scale of measurement. Another approach, based on indicator kriging (Journel 1983), has also been advocated for spatial predictions based on non-normal data. This technique utilizes indicator variables obtained by truncating the original data at several threshold values and then using ordinary kriging to predict each of the indicator variables. Indicator kriging provides an approximation to the conditional probability distribution of the variable of interest given the data, and the conditional mean of this empirical conditional distribution is known as the "E-type estimate" (Deutsch and Journel 1992). Most often, however, indicator kriging is limited to use with binary data to construct a probability map.

Although much work has been done on the development of nonlinear spatial predictors, little attention has been given to other spatial data analysis techniques whose focus is not on prediction or mapping, but on the estimation of fixed effects when the observations are spatially dependent. Researchers increasingly recognize the need to account for spatial correlation in the analysis of field experiments. Zimmerman and Harville (1991) have effectively illustrated the utility of geostatistical methods in the analysis of spatial field experiments. Stroup, Baenziger, and Mulitze (1994) showed that adjusting for spatial correlation can lead to conclusions that are much more intuitive and biologically meaningful than those otherwise obtained when this dependence is ignored. However, most analyses of data from field experiments that do incorporate spatial correlation depend on likelihood-based methods for estimation and inference. With spatial data, use of these methods effectively requires an assumption of multivariate normality, because, in practice, other assumptions prove intractable or place restrictions on the covariance structure. Unfortunately, the assumption of multivariate normality is not really appropriate for analysis of spatial counts or spatial categorical variables often of interest in ecology, geology, agronomy, and entomology. There remains a need for geostatistical methods that can incorporate these types of spatial data.

The generalized linear model (Nelder and Wedderburn 1972) is a systematic extension of linear model theory for non-normal data. In its original form, the generalized linear model involved maximum likelihood estimation for data from the exponential family of distributions. Wedderburn (1974) and McCullagh (1983) extended, via quasi-likelihood, the generalized linear model to cases where the first and second moments of the data could be described, but a parametric representation of the joint distribution could not be written explicitly. Most quasi-likelihood methods are based on the assumption of independent observations, but under certain conditions a quasi-likelihood may also be defined for dependent observations with a general variance-covariance matrix (McCullagh and Nelder 1989, Sec. 9.3). However, these conditions may be difficult to check for

a particular nontrivial correlation structure. To circumvent this problem, Liang and Zeger (1986) and Zeger and Liang (1986) obtained generalized estimating equations based on the quasi-likelihood approach that can be used for serially correlated data from repeated measures designs. Their approach was recently extended by Albert and McShane (1995) to the case of spatially correlated binary data that arise in neuroimaging.

The purpose of this article is to extend the methodology of generalized linear models to provide methods for data analysis, estimation, and prediction with discrete and categorical spatial variables. Universal kriging, indicator kriging, the current approaches for estimating treatment effects from designed experiments, and the methods described in Albert and McShane (1995) are special cases of the approach presented in this article. In the following sections, we present an overview of the theory of generalized linear models and then give several examples that illustrate how this theory can be used to enrich the analysis of spatial data.

2. OVERVIEW OF GENERALIZED LINEAR MODELS AND QUASI-LIKELIHOOD

2.1 GENERALIZED LINEAR MODELS

The linear model

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.1)$$

where \mathbf{Z} is the $n \times 1$ data vector, \mathbf{X} is an $n \times p$ matrix of explanatory variables or a design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown, fixed parameters, $E(\epsilon_i) = 0$, $i = 1, \dots, n$, and $\text{var}(\mathbf{Z}) = \sigma^2 I$, is prevalent in statistics and forms the basis for many types of analyses including regression analysis and the analysis of designed experiments. When working with such a model, one common goal is to estimate linear combinations of $\boldsymbol{\beta}$ that have important interpretations relevant to the physical or biological process under study. Exact inference with this model usually relies on an assumption of a Gaussian distribution for the data. However, when working with counts, probabilities, or binary data, a Gaussian assumption is not realistic, and in such cases it may also be more meaningful to model a function of the mean, rather than the mean itself, as a linear function of unknown parameters. Loglinear and logit models are prototypical examples that illustrate these issues.

The generalized linear model was first presented by Nelder and Wedderburn (1972). Because this original work, the framework of generalized linear models has been developed into a rich body of methods for analyzing data from non-normal distributions. In what follows, we give an overview of some of the most important components of a general linear model and indicate how they can be used to analyze data from a variety of data distributions within the exponential family. Readers are referred to McCullagh and Nelder (1989) for a more extensive treatment of generalized linear models.

Let Z_1, Z_2, \dots, Z_n be independent random variables, not necessarily identically distributed, each having a distribution in the exponential family. Denote the corresponding data values by z_1, z_2, \dots, z_n . The contribution of any z_i to the log-likelihood can be

written as

$$L(\theta_i, \phi_i; z_i) = \frac{z_i \theta_i - b(\theta_i)}{a(\phi_i)} + c(z_i, \phi_i), \quad (2.2)$$

where θ_i , the *natural parameter*, is a function of $\mu_i = E(Z_i)$, and ϕ_i is a scale parameter. The variance of Z_i is $\text{var}(Z_i) = v(\mu_i)a(\phi_i)$, where $v(\mu_i)$, the *variance function*, is the component of $\text{var}(Z_i)$ that depends on μ_i .

The joint log-likelihood is

$$L(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{z}) = \sum_{i=1}^n L(\theta_i, \phi_i; z_i) = \mathbf{z}' A^{-1} \boldsymbol{\theta} - (\mathbf{b}_{\boldsymbol{\theta}}^{1/2})' A^{-1} (\mathbf{b}_{\boldsymbol{\theta}}^{1/2}) + \mathbf{1}' \mathbf{c}, \quad (2.3)$$

where \mathbf{z} is the $n \times 1$ data vector, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)'$ is the vector of natural parameters, $\mathbf{b}_{\boldsymbol{\theta}}^{1/2} = (\sqrt{b(\theta_1)}, \dots, \sqrt{b(\theta_n)})'$, $A = \text{diag}[a(\phi_i)]$ is an $n \times n$ matrix, \mathbf{c} is an $n \times 1$ vector with i th element $c(z_i; \phi_i)$, and $\mathbf{1}$ is an $n \times 1$ vector of ones. Note that the variance-covariance matrix of \mathbf{Z} , denoted V , can be written as

$$V \equiv \text{var}(\mathbf{Z}) = v_{\boldsymbol{\mu}}^{1/2} A v_{\boldsymbol{\mu}}^{1/2}, \quad (2.4)$$

where $v_{\boldsymbol{\mu}} = \text{diag}[v(\mu_i)]$.

The main premises of the generalized linear model are that (1) the observations are not necessarily linearly related to μ_i , and hence, fitting a linear model directly to μ_i may not yield satisfactory results or meaningful interpretation, and (2) estimation must take into account the form of the log-likelihood of the data. Accordingly, a linear model is fitted to some function of the mean, $\boldsymbol{\eta} = g(\boldsymbol{\mu})$, called the *link function*. Thus, the generalized linear model has

$$\begin{aligned} E(\mathbf{Z}) &= \boldsymbol{\mu} \\ \boldsymbol{\eta} &= g(\boldsymbol{\mu}) = X\boldsymbol{\beta}, \end{aligned} \quad (2.5)$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, X is an $n \times p$ matrix of known constants, and $\text{var}(\mathbf{Z}) = V$. Alternatively, the mean function may be written as

$$E(\mathbf{Z}) = h(X\boldsymbol{\beta}), \quad (2.6)$$

where $h(\cdot)$ is the *inverse link* function. The link function may be chosen to relate the mean vector $\boldsymbol{\mu}$ to the vector of natural parameters $\boldsymbol{\theta}$, or it may follow from some biological or physical model of the mean. While not necessary, for many generalized linear models there is a one-to-one correspondence between the link and inverse link functions. In such cases, $h(\cdot) = g^{-1}(\cdot)$.

Nelder and Wedderburn (1972) showed that maximum likelihood estimate of $\boldsymbol{\beta}$ can be obtained from the method of scoring as an iterative solution to the system of equations

$$X'W X \boldsymbol{\beta} = X'W \mathbf{z}^*, \quad (2.7)$$

where $D = \text{diag}[\partial \mu_i / \partial \eta_i]$ is an $n \times n$ matrix, $W = D' V^{-1} D$, and $\mathbf{z}^* = \boldsymbol{\eta} + D^{-1}(\mathbf{z} - \boldsymbol{\mu})$.

Although estimation with the generalized linear model is a maximum likelihood procedure, the full log-likelihood is not needed. It is sufficient to describe the relationship

between the mean and the model, either as $\eta = g(\mu) = X\beta$ or as $\mu = h(X\beta)$, the form of the variance [i.e., the scale parameters $a(\phi_i)$] and the relationship between the variance and the mean, v_μ . For many distributions, $a(\phi_i)$ are known or can be treated as nuisance parameters. Quasi-likelihood formalizes this idea.

2.2 QUASI-LIKELIHOOD: INDEPENDENT OBSERVATIONS

The quasi-likelihood function, denoted here as $Q(\mu_i; z_i)$, was defined in Wedderburn (1974) by the relation

$$\frac{\partial Q(\mu_i; z_i)}{\partial \mu_i} = \frac{z_i - \mu_i}{v(\mu_i)}. \quad (2.8)$$

Wedderburn showed that the quasi-likelihood, Q , is equivalent to the log-likelihood for members of the exponential family and satisfies conditions similar to those satisfied by the score vector and information matrix associated with ordinary log-likelihoods. Consequently, the quasi-likelihood estimator of β is the solution to the system of score equations

$$\Delta' v_\mu^{-1} (\mathbf{z} - \boldsymbol{\mu}) = \mathbf{0}, \quad (2.9)$$

where $\boldsymbol{\mu} = h(X\beta)$, Δ is an $n \times p$ matrix with i th column equal to the $n \times 1$ vector $[\partial \boldsymbol{\mu} / \partial \beta_i]$, and $v_\mu = \text{diag}[v(\mu_i)]$, where $v(\mu_i)$ is the variance function defined in the previous section. Noting that for models of the form $\eta = X\beta$, $\Delta = [(\partial \mu_j / \partial \eta_j)(\partial \eta_j / \partial \beta_i)]$ can be written as DX , where $D = \text{diag}[\partial \mu_i / \partial \eta_i]$, and applying the method of scoring, it can be shown that the quasi-likelihood estimating equations can equivalently be written as

$$X'WX\beta = X'W\mathbf{z}^*, \quad (2.10)$$

which are identical to the standard generalized linear model estimating equations given in Equation (2.7) (assuming a diagonal scale parameter matrix, A).

McCullagh (1983) showed that under appropriate limiting conditions, $\hat{\beta}_{QL}$ obtained as the solution to the quasi-likelihood estimating equations given in (2.10) is consistent, asymptotically unbiased, and normally distributed with asymptotic covariance matrix $(X'WX)^{-1}$.

2.3 QUASI-LIKELIHOOD: DEPENDENT OBSERVATIONS

In the case of dependent observations, the quasi-likelihood function defined by Equation (2.8) can be generalized to

$$\frac{\partial Q(\boldsymbol{\mu}; \mathbf{z})}{\partial \boldsymbol{\mu}} = V^{-1}(\mathbf{z} - \boldsymbol{\mu}), \quad (2.11)$$

where V is no longer a diagonal matrix as assumed in Section 2.2, but a general symmetric positive definite matrix whose elements are functions of $\boldsymbol{\mu}$. By differentiating Q with

respect to β , the score function can be written as

$$\mathbf{U} = \Delta' V^{-1}(\mathbf{z} - \boldsymbol{\mu}), \quad (2.12)$$

where Δ is an $n \times p$ matrix with elements $[\partial\mu_i/\partial\beta_j]$.

In order to guarantee a solution to Equation (2.11), V^{-1} must satisfy certain conditions not easily verified in practice (see McCullagh and Nelder 1989, Sec. 9.3). However, the score function may still be used to provide a reasonable estimator of β . Liang and Zeger (1986) and Zeger and Liang (1986) showed that, under mild regularity conditions, a consistent estimator of β can be obtained as a solution to $\mathbf{U} = \mathbf{0}$, based on nonindependent observations whose variance-covariance matrix can be written as

$$V = \text{var}(\mathbf{Z}) = v_{\boldsymbol{\mu}}^{1/2} R(\boldsymbol{\alpha}) v_{\boldsymbol{\mu}}^{1/2}, \quad (2.13)$$

where R is a correlation matrix that describes the dependence among the observations, and $\boldsymbol{\alpha}$ is a $g \times 1$ vector that fully characterizes R . The vector $\boldsymbol{\alpha}$ may include parameters as described in Equation (2.2), or may include only correlation parameters. If only correlation parameters appear in $\boldsymbol{\alpha}$, then $v_{\boldsymbol{\mu}}$ will contain any necessary scale parameters and will not be a variance function per se.

3. SPATIAL DATA ANALYSIS: ESTIMATION AND PREDICTION WITH GENERALIZED LINEAR MODELS

In geostatistics, the elements of the correlation matrix $R(\boldsymbol{\alpha})$, described previously, are obtained from semivariogram models parameterized by constants denoting the nugget effect, the partial sill, and the range. Cressie (1991) described these models and Albert and McShane (1995) demonstrated how they can be used to construct $R(\boldsymbol{\alpha})$.

3.1 ESTIMATION

For V written in the form given by Equation (2.13), with parameters that describe the spatial dependence among observations incorporated through $R(\boldsymbol{\alpha})$, $\hat{\beta}_G$ obtained as a solution to Equation (2.9) provides a consistent estimator of β , and the general equation in (2.9) is referred to as the generalized estimating equation (GEE). Liang and Zeger (1986) emphasized that $\hat{\beta}_G$ will still be consistent for β even if the correlation matrix is not correctly specified.

Under a generalized linear model formalism with $\eta = g(\boldsymbol{\mu}) = X\beta$, the derivative matrix, Δ in Equation (2.12) is $\Delta = DX$, with $D = \text{diag}[\partial\mu_i/\partial\eta_i]$. Substituting these into Equation (2.12), and applying the method of scoring as in the previous section, we again obtain

$$X'WX\beta = X'W\mathbf{z}^*, \quad (3.1)$$

where $D = \text{diag}[\partial\mu_i/\partial\eta_i]$ is an $n \times n$ matrix, $W = D'V^{-1}D$, V is given by Equation (2.13), and $\mathbf{z}^* = \boldsymbol{\eta} + D^{-1}(\mathbf{z} - \boldsymbol{\mu})$.

3.2 SPATIAL PREDICTION

Let $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$ represent the vector of random variables associated with the data values z_1, \dots, z_n , observed at spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_n$. Suppose we want to predict the values of k additional random variables, $Z(\mathbf{s}_{0,1}), \dots, Z(\mathbf{s}_{0,k})$, at locations $\mathbf{s}_{0,1}, \dots, \mathbf{s}_{0,k}$ where no observations are available. Let \mathbf{Z}_0 be the corresponding $k \times 1$ vector of random variables whose values are to be predicted. Assume that the mean function for \mathbf{Z} and \mathbf{Z}_0 can be written as

$$\begin{aligned} E(\mathbf{Z}) &= \boldsymbol{\mu}(\mathbf{s}) \\ E(\mathbf{Z}_0) &= \boldsymbol{\mu}(\mathbf{s}_0), \end{aligned} \quad (3.2)$$

where $\boldsymbol{\mu}(\mathbf{s})$ is an $n \times 1$ dimensional mean vector associated with data locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, and $\boldsymbol{\mu}(\mathbf{s}_0)$ is a $k \times 1$ dimensional mean vector associated with the prediction locations $\mathbf{s}_{0,1}, \dots, \mathbf{s}_{0,k}$. Further assume that

$$\text{var} \begin{pmatrix} \mathbf{Z} \\ \mathbf{Z}_0 \end{pmatrix} \equiv V = \begin{bmatrix} \Sigma_{ZZ} & \Sigma_{Z0} \\ \Sigma_{0Z} & \Sigma_{00} \end{bmatrix}, \quad (3.3)$$

where Σ_{ZZ} , Σ_{Z0} , and Σ_{00} are known positive definite matrices of dimensions $n \times n$, $n \times k$, and $k \times k$, respectively.

Under the assumption that the mean functions can be written as $\boldsymbol{\mu}(\mathbf{s}) = \mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\mu}(\mathbf{s}_0) = \mathbf{X}_0\boldsymbol{\beta}$, Gotway and Cressie (1993) showed that the best heterogeneous linear predictor has the form

$$p(\mathbf{Z}; \mathbf{Z}_0) = \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z} + (\mathbf{X}_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{X})\boldsymbol{\beta}.$$

This predictor is well known from the early days of optimal prediction (e.g., Wold 1938) and is referred to in geostatistics as the simple kriging predictor. Gotway and Cressie (1993) used this predictor as a basis for constructing a large class of predictors generated by different estimators of $\boldsymbol{\beta}$, given by

$$p(\mathbf{Z}; \mathbf{Z}_0) = \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z} + (\mathbf{X}_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{X})\widehat{\boldsymbol{\beta}}, \quad (3.4)$$

where $\widehat{\boldsymbol{\beta}}$ is any estimator of $\boldsymbol{\beta}$. The best linear unbiased predictor (BLUP) (Goldberger 1962), also known as the universal kriging predictor (Cressie 1986), is contained in this class and is obtained by taking $\widehat{\boldsymbol{\beta}}$ to be the generalized least squares estimator of $\boldsymbol{\beta}$. Gotway and Cressie (1993) obtained other predictors from this class and focused on predictors obtained using shrinkage estimators of $\boldsymbol{\beta}$.

One way of applying this general class of predictors to prediction with generalized linear models is to use $\widehat{\boldsymbol{\beta}}$, obtained by the iterative solution to the matrix Equation (3.1) with $V \equiv \Sigma_{ZZ}$. However, this assumes a linear function for the mean and does not allow explicit use of a link function, which forms a basic underlying premise of generalized linear model methodology. Instead, we rewrite (3.4) as

$$p(\mathbf{Z}; \mathbf{Z}_0) = \mathbf{X}_0\widehat{\boldsymbol{\beta}} + \Sigma_{0Z}\Sigma_{ZZ}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}), \quad (3.5)$$

which, more generally, can be expressed as

$$\widehat{\mathbf{Z}}_0 = \widehat{\boldsymbol{\mu}}(\mathbf{s}_0) + \Sigma_{0Z}\Sigma_{ZZ}^{-1}(\mathbf{Z} - \widehat{\boldsymbol{\mu}}(\mathbf{s})), \quad (3.6)$$

corresponding to the model in (3.2). Thus, prediction with generalized linear models can be accomplished by obtaining $\widehat{\beta}_G$ as the solution vector corresponding to Equation (3.1), estimating η as $\widehat{\eta} = X\widehat{\beta}_G$, corresponding to the data, and $\widehat{\eta}_0 = X_0\widehat{\beta}_G$, corresponding to the variables to be predicted. Then $\widehat{\mu}(s) = h(\widehat{\eta})$ and $\widehat{\mu}(s_0) = h(\widehat{\eta}_0)$ can be used with an appropriate matrix V , based on the relationship in Equation (2.13), to calculate \widehat{Z}_0 from Equation (3.6). In practice, the elements of V are obtained from Equation (2.13), and a model that expresses the correlation between two data points as a function of the distance between them is used to obtain the elements in $R(\alpha)$.

4. EXAMPLES

In this section, we illustrate how the quasi-likelihood approach may be used to analyze spatial data. The first example is based on data from a designed experiment in which the goal is to estimate treatment effects and make treatment comparisons. The last two examples illustrate how the quasi-likelihood approach can be used in spatial prediction and mapping. Note that these examples are provided solely to demonstrate the application of the methodology discussed in the previous section and do not present a full analysis of the datasets.

4.1 EXAMPLE 1: ESTIMATION OF TREATMENT EFFECTS

This example is from an agronomic field trial designed to evaluate experimental varieties (entries) of wheat for resistance to Hessian fly damage. The data are given in Appendix Table A.1. Sixteen entries were compared using a randomized complete block design with four blocks. The field plots, each 3.7×3.7 m, were arranged as an 8×8 grid and divided into four 4×4 quarters with each quarter forming one block. In each plot corresponding to the ij th treatment combination, n_{ij} plants were sampled. The plants were classified according to whether Hessian fly damage was present or not. In each plot, z_{ij} plants showed damage and $n_{ij} - z_{ij}$ did not. Thus, the sample proportion of damaged plants was $p_{ij} = z_{ij}/n_{ij}$. The objective of the experiment was to compare the π_j , the probability of an entry j plant showing damage.

Because the response variable is binomial, the data were analyzed using a generalized linear model with logit link function

$$\eta_{ij} = g(\pi_{ij}) = \log[\pi_{ij}/(1 - \pi_{ij})] = m + \rho_i + \tau_j, \quad (4.1)$$

where m is an intercept, ρ_i is the i th block effect, and τ_j is the effect of the j th treatment. To estimate the parameter vector $\beta = (m, \rho_1, \rho_2, \rho_3, \rho_4, \tau_1, \tau_2, \dots, \tau_{16})'$, the elements of the estimating equations given in (3.1), specifically, v_μ , $R(\alpha)$, and D are needed. For binomial data, the general form of the variance function is $v(\mu) = \mu(1 - \mu)$, and the scale parameter is $a(\phi) = 1/n$. Thus, v_μ can be written in terms of the variance function and the known scale parameters as

$$v_\mu = \text{diag}[\pi_{ij}(1 - \pi_{ij})/n_{ij}]. \quad (4.2)$$

The field was determined to have substantial spatial correlation, which could be modeled

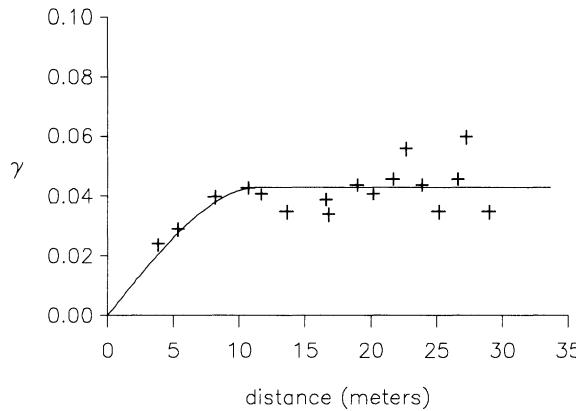


Figure 1. Empirical Semivariogram and Fitted Model for Example 1.

using a spherical semivariogram with no nugget effect and a range of 11.6 m. A plot of the semivariogram appears in Figure 1. The k th element of the correlation matrix is

$$r_{kl} = \begin{cases} 1 - 1.5(d_{kl}/11.6) + .5(d_{kl}/11.6)^3 & \text{if } d_{kl} \leq 11.6 \\ 0 & \text{if } d_{kl} > 11.6 \end{cases}, \quad (4.3)$$

where d_{kl} is the distance between the k th and l th observations. Thus, the variance-covariance matrix was modeled as $V = v_{\mu}^{1/2} R v_{\mu}^{1/2}$, where v_{μ} is given in Equation (4.2) and R is a correlation matrix with elements $[r_{ij}]$, given by (4.3). Finally, $D = \text{diag}[\partial \pi_j / \partial \eta_j]$. From the link function given in (4.1), it follows that the inverse link is

$$\pi_{ij} = h(\eta_{ij}) = \exp(\eta_{ij})/[1 + \exp(\eta_{ij})],$$

$\partial \pi_{ij} / \partial \eta_{ij} = \pi_{ij}(1 - \pi_{ij})$, and D is therefore $\text{diag}[\pi_{ij}(1 - \pi_{ij})]$.

Because the X matrix for model (4.1) is singular, a generalized inverse was required to compute solutions to the quasi-likelihood estimating equations given in (3.1). Using the sweep procedure of Goodnight (1979), a solution for β is given in Table 1. The estimates of π_j can be obtained from $\hat{\beta}$ by first estimating $\eta_j = m + \rho_+ + \tau_j$ and then applying the inverse link. For example, for treatment 1, $\eta_1 = m + \rho_+ + \tau_1 = -1.67 + (1/4)(-.45 + .04 - .13 + 0) + 4.07 = 2.27$, and the estimated probability of a favorable outcome for treatment 1 is $\hat{\pi}_1 = \exp(2.27)/[1 + \exp(2.27)] = .91$. Estimates for the other treatments can be obtained similarly.

Interval estimation and hypothesis testing proceed along the lines of standard generalized linear models. That is, for any estimable function of the parameter vector, $\mathbf{k}'\beta$, the estimate $\mathbf{k}'\hat{\beta}$ is distributed approximately $N(\mathbf{k}'\beta, \mathbf{k}'(X'WX)^{-1}\mathbf{k})$. Thus, the standard error of an estimable function is approximately $(\mathbf{k}'(\hat{W}X)^{-1}\mathbf{k})^{1/2}$, where \hat{W} is the estimate of W obtained from the fitted model.

Confidence intervals for the predicted π_j can be obtained by applying the inverse link to the upper and lower confidence bounds of the estimable function $\eta_j = m + \rho_+ + \tau_j$, that is,

$$\begin{aligned} & \exp(\eta_j - z_{\alpha/2} \text{SE}(\eta_j))/[1 + \exp(\eta_j - z_{\alpha/2} \text{SE}(\eta_j))], \\ & \exp(\eta_j + z_{\alpha/2} \text{SE}(\eta_j))/[1 + \exp(\eta_j + z_{\alpha/2} \text{SE}(\eta_j))], \end{aligned}$$

Table 1. Estimates of the Parameter Vector β for Example 1

Parameter	Estimate
m	-1.67
ρ_1	-.45
ρ_2	.04
ρ_3	-.13
ρ_4	0
τ_1	4.07
τ_2	3.47
τ_3	2.77
τ_4	2.20
τ_5	2.55
τ_6	2.14
τ_7	2.54
τ_8	1.85
τ_9	1.74
τ_{10}	2.19
τ_{11}	1.29
τ_{12}	2.21
τ_{13}	-.44
τ_{14}	1.35
τ_{15}	.76
τ_{16}	0

where $z_{\alpha/2}$ is the appropriate value from the standard normal distribution for a $(1 - \alpha)\%$ confidence level. Standard errors for the estimated π_j can be obtained using the Delta rule (Agresti 1990). The general form is

$$\text{var}(h(\eta)) = \left[\frac{\partial h(\eta)}{\partial \eta} \right]^2 \text{var}(\eta).$$

For the estimated π_j , the standard error is thus

$$[((\hat{\pi}_i)(1 - \hat{\pi}_j))^2 \mathbf{k}'(X' \hat{W} X)^{-1} \mathbf{k}]^{1/2},$$

where \mathbf{k} contains the elements corresponding to the estimable function $\eta_{.j} = m + \rho_{.} + \tau_{.j}$. For example, for treatment 1, the standard error of $\hat{\pi}_1$ is $[((.91)(1 - .91))^2 (2.107)]^{1/2} = .124$, where 2.107 is the standard error of $\eta_{.j}$.

Hypotheses of the form $K'\beta = \theta$ can be tested using the Wald Statistic

$$(K'\hat{\beta} - \theta)' [K'(X'WX)^{-1} K]^{-1} (K'\hat{\beta} - \theta).$$

This statistic has an approximate χ^2_ν distribution, where $\nu = \text{rank}(K)$.

Table 2 gives estimates and standard errors of the π_j for the 16 entry means. Estimates are given for three methods: (1) the generalized linear model accounting for spatial correlation, described previously, (2) a "standard" generalized linear model assuming independent plots, and (3) a normal errors model accounting for spatial correlation. Estimates (2) and (3) correspond to results one would obtain using a generalized linear model procedure (e.g., SAS PROC GENMOD) and a normal errors linear model procedure capable of correlated errors (e.g., SAS PROC MIXED), respectively. Table 2 also

Table 2. Predicted Probabilities, Standard Errors, and Test Statistics for Three Methods of Analysis

Entry	<i>GLM with Spatial correlation</i>		<i>GLM with independent units</i>		<i>GLS-Normal errors, spatial correlation</i>	
	\hat{p}	$SE(\hat{p})$	\hat{p}	$SE(\hat{p})$	\hat{p}	$SE(\hat{p})$
1	.91	.124	.81	.152	.98	.108
2	.84	.119	.79	.145	.85	.108
3	.72	.116	.73	.146	.75	.109
4	.60	.103	.52	.132	.63	.107
5	.68	.125	.79	.154	.68	.111
6	.58	.117	.64	.153	.58	.105
7	.68	.115	.70	.151	.68	.108
8	.51	.119	.46	.148	.55	.110
9	.48	.107	.52	.133	.50	.110
10	.60	.131	.70	.167	.67	.108
11	.37	.113	.51	.144	.37	.108
12	.60	.118	.51	.150	.58	.108
13	.09	.130	.11	.167	.07	.109
14	.39	.121	.27	.156	.37	.110
15	.26	.110	.35	.137	.23	.110
16	.14	.112	.19	.145	.13	.109
<i>Wald χ^2:</i>						
Block		3.07		4.12		.72
Entry		252.33		103.42		126.00

contains Wald statistics for the block and overall entry effects. For entries particularly affected by spatial dependence (e.g., entries 2, 6, and 11 through 16), the estimates for the spatial GLM show adjustment for spatial correlation similar to the normal errors model, whereas the GLM assuming independence often gives quite different estimates on the entry probabilities. The Wald statistics for the spatial GLM are quite high. This suggests future research into the sampling distributions of statistics generated by spatial GLM's and the determination of appropriate modifications, if needed.

4.2 EXAMPLE 2: PREDICTION OF WEED COUNTS.

The data discussed in this example are based on 1992 velvetleaf weed counts collected from an eastern Nebraska corn and soybean field (Johnson, Mortensen, and Gotway 1996). A total of 821 counts were taken every 7 meters over an area 189×224 m, with a small portion of the southwest part of the field not sampled. For this example, the original weed count dataset was sampled every 14 meters to obtain a subset of 212 counts. A three-dimensional scatter diagram of the sampled weed count dataset is shown in Figure 2.

Because the data are counts, a Poisson-type generalized linear model with link function

$$\eta_i = g(\mu_i) = \log(\mu_i) \quad (4.4)$$

and variance function

$$v(\mu_i) = \mu_i, \quad (4.5)$$

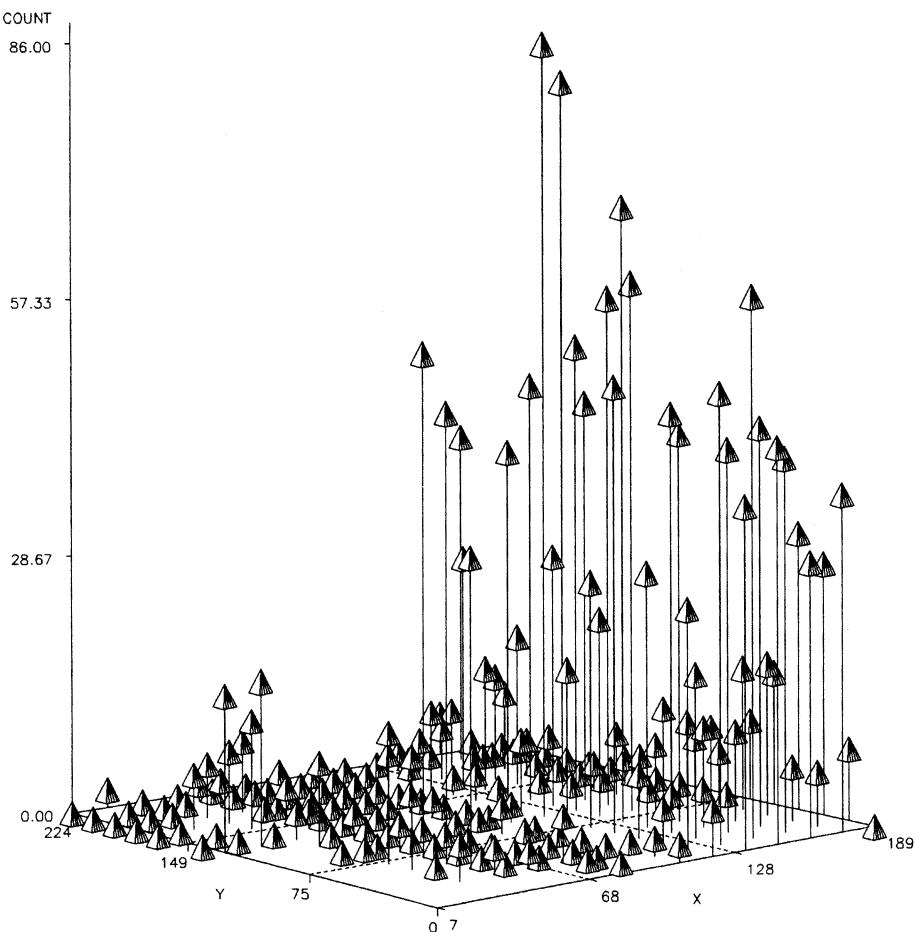


Figure 2. Scatter Diagram of Weed Count Data for Example 2.

seemed reasonable. To account for the large-scale trend apparent from Figure 2, the link function was modeled as a linear combination of trend surface parameters $\mathbf{x}'_i \boldsymbol{\beta}$, where \mathbf{x}_i is a vector of functions of the spatial coordinate \mathbf{s}_i that reflects the large-scale trend. Classical multiple regression of the data $\{z_i\}$ on the coordinates $\{(u_i, v_i)\}$ that designate locations $\{\mathbf{s}_i\}$ in the plane suggested

$$\mathbf{x}'_i = (1 \quad u_i \quad v_i \quad u_i v_i \quad u_i^2) \quad (4.6)$$

as a plausible functional form for this trend surface. Thus, the large-scale mean structure was described by $\eta = X\boldsymbol{\beta}$, where X is an $n \times p$ matrix of trend surface functions with j th column \mathbf{x}_j , and $\boldsymbol{\beta}$ is a 5×1 vector of unknown parameters that reflect the contribution of each trend surface function to the overall large-scale structure of the weed count process.

The residuals from the classical regression were used to check for additional spatial variation. The semivariogram of the residuals was used to infer the range of correlation for the small-scale error process. An exponential model,

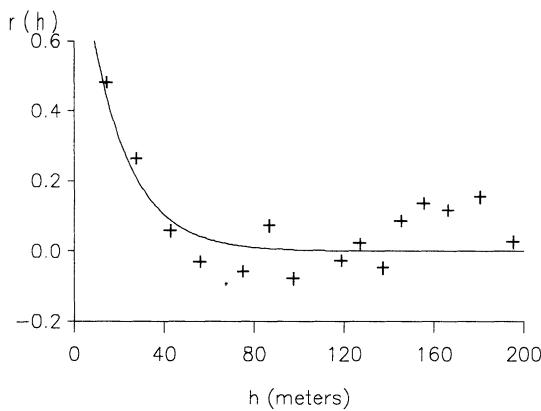


Figure 3. Empirical Residual Semivariogram and Fitted Model for Example 2.

$$r(d_{ij}) = 1 - \exp(-d_{ij}/17.6), \quad (4.7)$$

was used to model the correlation function of the process where the range of 17.6 meters was obtained using the weighted least squares method of semivariogram model fitting described in Cressie (1985). The residual correlation function and fitted model are shown in Figure 3. Thus, the variance of the weed count process was modeled as in Equation (2.13) using $v\mu = \text{diag}[\mu_i]$, and correlation matrix R having elements $r(d_{ij})$ given in Equation (4.7).

An alternative model based on a negative binomial-type process was also considered for the weed count data. This distribution is commonly used in weed ecology to describe the distribution of weed counts (Wiles, Oliver, York, Gold, and Wilkerson 1992). With this model, the link function was taken to be

$$\eta_i = g(\mu_i) = \log\left(\frac{\mu_i}{\mu_i + k}\right), \quad (4.8)$$

and variance function was specified as

$$v(\mu_i) = \mu_i + \frac{\mu_i^2}{k}, \quad (4.9)$$

where k is the aggregation parameter associated with the negative binomial distribution. Wiles et al. (1992) discussed the interpretation of this parameter and methods for estimating it from sample data. For the purpose of this example, the value $k = .32$ was obtained from the sample data using a method of moments estimator.

The variance of the weed count process was also modeled as in Equation (2.13) with the correlation matrix R having elements $r(d_{ij})$ given in Equation (4.7), but with $v\mu = \text{diag}[\mu_i + \mu_i^2/k]$. The functional form of the trend surface was the same as that designated by Equation (4.6).

The method of scoring was used to estimate β for each of the models using Equation (3.1). In addition, both of the generalized linear models and universal kriging (corresponding to estimation of β using generalized least squares) were used to map the weed count

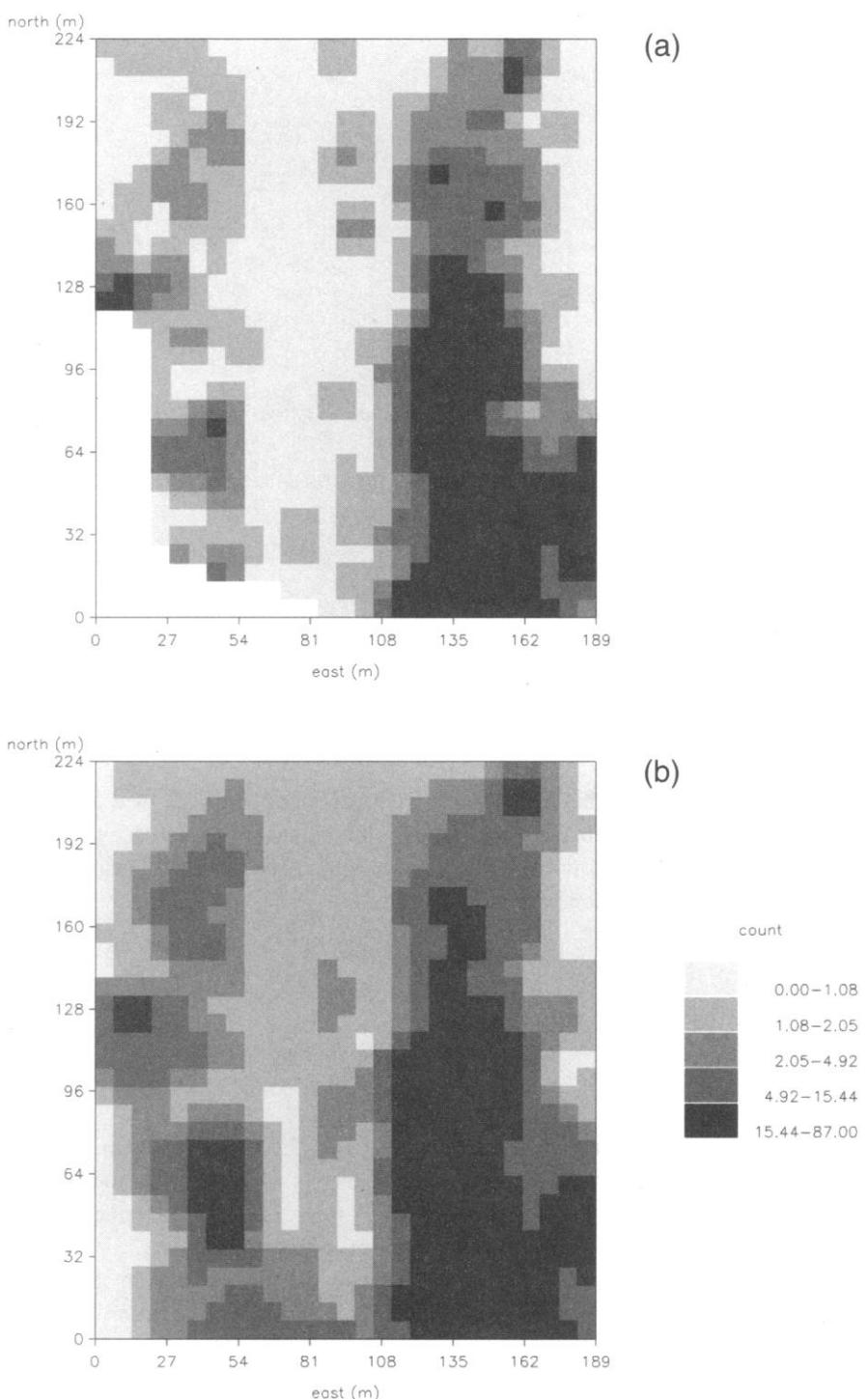


Figure 4. Weed Count Maps: (a) "True" map, based on all 821 weed counts, and (b) universal kriging map of the same field predicted from 212 weed counts.

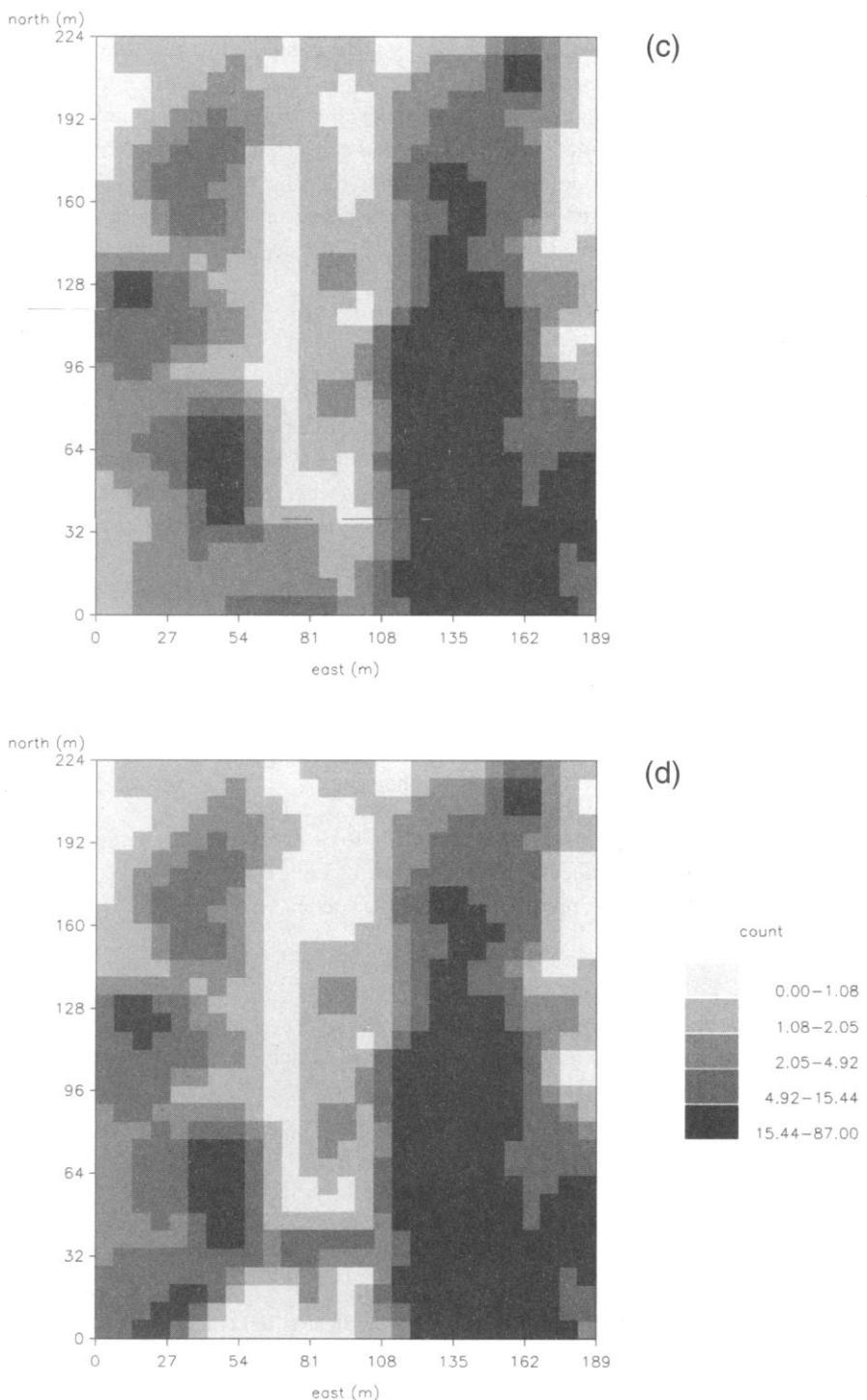


Figure 4. Weed Count Maps: (c) Poisson and (d) negative binomial maps, respectively, of the same field predicted from 212 weed counts.

process over the 189×224 m area using a grid spacing of 7 meters. Equation (3.6) was used to make all predictions. All three maps were compared to the true map based on all 812 original data and are shown in Figure 4.

The general features represented in all maps appear very similar and seem to represent the coarse features of the spatial distribution of weeds. The three predicted maps are somewhat different from one another, particularly in the southwest portion of the field where the process was extrapolated. The maps based on the generalized linear model approaches seem to capture the trough of low values in the center portion of the field that is not reflected in the universal kriging map.

4.3 EXAMPLE 3: MAPPING INDICATOR VARIABLES

One important application in geostatistics is concerned with predicting probabilities of the form $\Pr(Z(\mathbf{s}_0) > z)$, for some threshold value z . Such probabilities are of interest in many application areas, particularly in mining (where they are used in predicting the quantity of metal above a certain cutoff grade) and in environmental risk assessment (in which they are used to infer hot spots where contamination levels have exceeded a critical threshold). To illustrate how the generalized linear model approach can be used in these situations, we transformed data on potentiometric head (level of groundwater in a well measured in feet above sea level) from the Wolfcamp aquifer in West Texas (Harper and Furr 1986; Cressie 1991) into binary observations using

$$I(\mathbf{s}) = \begin{cases} 1 & \text{if } Z(\mathbf{s}) > 2,000 \\ 0 & \text{otherwise} \end{cases}. \quad (4.10)$$

The cutoff value of 2,000 was chosen arbitrarily to designate relatively high groundwater levels. A posting of the indicator data is shown in Figure 5.

The indicator kriging approach assumes the model

$$I(\mathbf{s}) = \mu + \epsilon(\mathbf{s}), \quad (4.11)$$

where μ is a constant that is independent of location but that is not constrained to be in $[0, 1]$. Ordinary kriging is used with the indicator variables and the indicator semivariogram to provide an approximation to $E(I(\mathbf{s})|\mathbf{Z}) = P(Z(\mathbf{s}) > 2,000|\mathbf{Z})$. The indicator semivariogram calculated from the indicator values is shown in Figure 6. Theoretically, the sill of this semivariogram should be about $p(1 - p)$ where p is the proportion of 1's in the dataset (in this case $p(1 - p) = .25$). Clearly, the indicator semivariogram does not reach a sill at .25, reflecting the effect of large-scale trend (considered to be linear by both Harper and Furr (1986) and Cressie (1991)) on the indicator data that is also evident in the data posting in Figure 5.

To account for the binary nature of the indicator data and the constraint on the variance, we used a generalized linear model approach based on a Bernoulli model with a link function

$$\eta_i = g(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right), \quad (4.12)$$

and variance function

$$v(\pi_i) = \pi_i(1 - \pi_i), \quad (4.13)$$

where $\pi_i = E(I(s_i))$. The large scale trend was incorporated as in the previous example using $\eta_i = x'_i\beta$, with $x'_i = (1 \ u_i \ v_i)$.

The correlation was modeled with a spherical model, but the range parameter was estimated from the data using the following iterative approach:

Step 1: β was estimated by iteratively solving Equation (3.1) based on the generalized linear model specified previously. The variance-covariance matrix, V , was modeled as in Equation (2.13), with $v_{\mu} = \text{diag}[\pi_i(1 - \pi_i)]$, and R having the same general form as that of Equation (4.3), but with the value of 3.1 replaced by the more general range parameter, a . A range of $a = 46$ miles, borrowed from the analysis in Cressie (1991), was used as a starting value.

Step 2: The empirical semivariogram of the residuals was then calculated and a was estimated using weighted least squares (Cressie 1985).

Step 3: The estimated range obtained in Step 2 was used to re-estimate β as in Step 1.

Steps 4–N: Steps 3–N were repeated N times until the estimates of β and a remained unchanged.

After $N = 18$ iterations (using a fairly strict convergence criterion), $\hat{\beta} = (-7.39 \ 10.08)'$ and $\hat{a} = 53.9$ miles. The semivariogram of the residuals is shown in Figure 7.

The predictor given in Equation (3.6) was used to map the area using estimated means $x'_i\hat{\beta}$ for the data locations, and means $x'_{0,j}\hat{\beta}$ for the predicted variables at locations $\{s_{0,j}\}$, and covariance structure $v_{\mu}\hat{R}(\hat{\alpha})v_{\mu}$. Indicator kriging was also used to construct a comparable map, based on a generalized least squares estimate of μ in the model given in Equation (4.11), and a spherical semivariogram model fit to the indicator semivariogram displayed in Figure 6. Both of these maps are shown in Figure 8. Again, the maps reflect the same large-scale features, but clearly differ in regions where data were sparse, because of the differences in the underlying models associated with the two different prediction

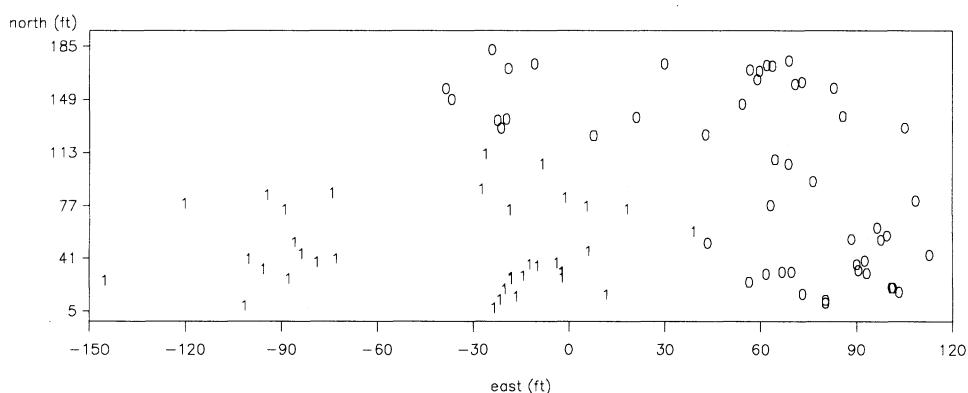


Figure 5. Posting of Wolfcamp Aquifer Indicator Data for Example 3.

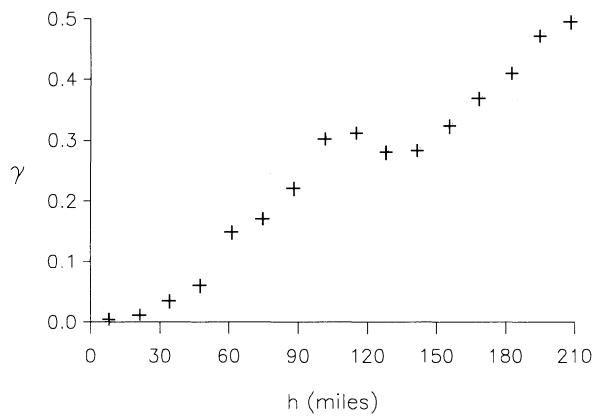


Figure 6. Empirical Indicator Semivariogram for the Wolfcamp Aquifer Data of Example 3.

approaches. The map based on the generalized linear model approach seems to be more consistent with the data when compared to the post plot shown in Figure 5.

5. DISCUSSION

The generalized linear model, in conjunction with quasi-likelihood, provides a framework for integrating geostatistics with spatially correlated observations that are inherently non-normal. For many applications, the choice of an appropriate link function not only results in an analysis consistent with the distribution of the data, but guarantees estimates that are within the parameter space. For example, agronomists have often fit linear models directly to the observed proportions (i.e., $\pi_j = m + \tau_j$.) Estimated π_j from such a model may be less than 0 or greater than 1. On the other hand, the logit link guarantees the estimated π_j will be between 0 and 1.

The methods discussed in this article may be applied to a wide variety of potential problems. Assuming that a reasonable link function and a plausible model of the spatial

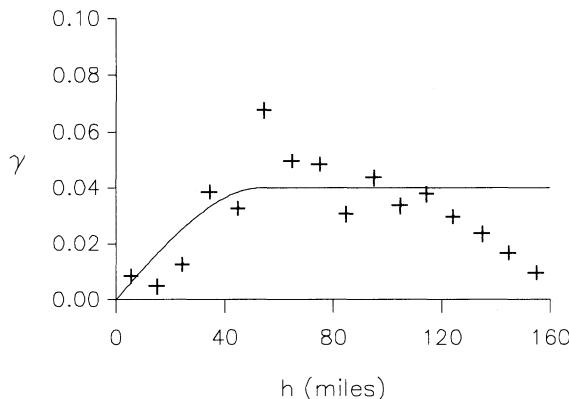


Figure 7. Empirical Residual Indicator Semivariogram and Fitted Model for Example 3.

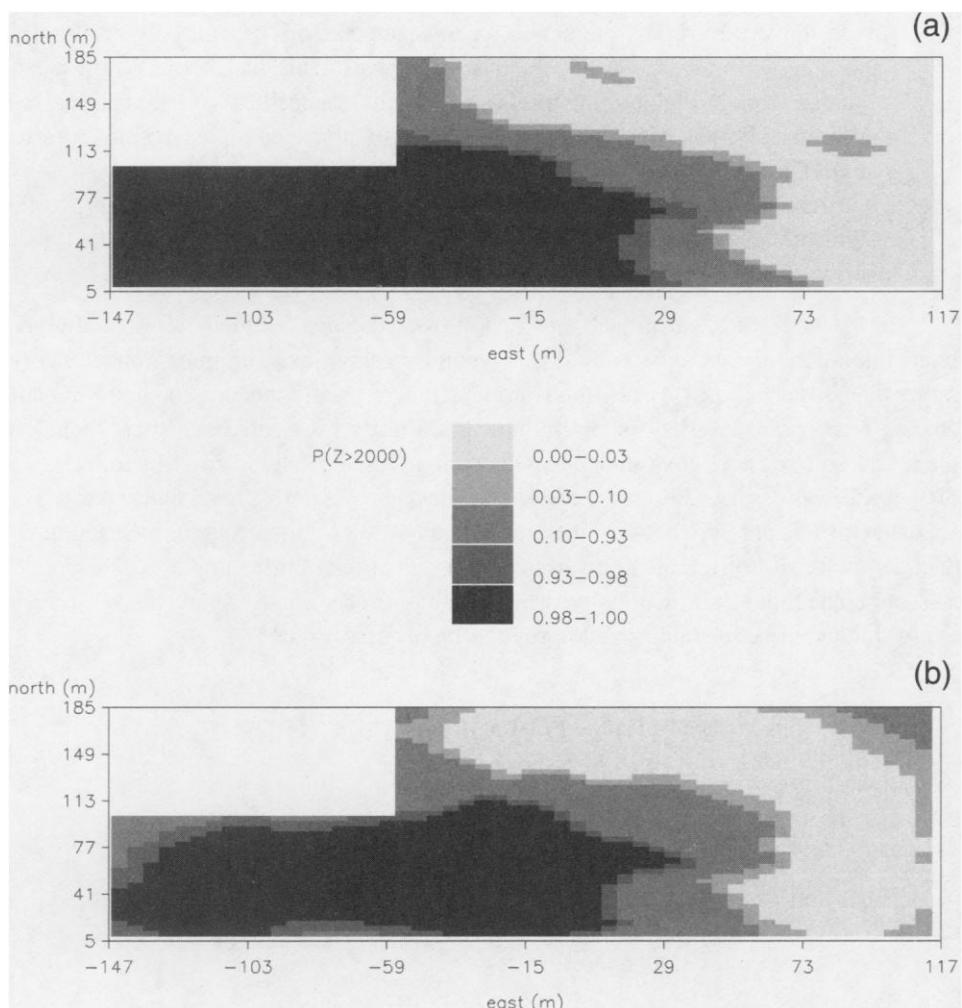


Figure 8. Wolfcamp Aquifer Probability Maps Created Using (a) Generalized Linear Model Approach and (b) Indicator Kriging.

correlation can be specified, the generalized linear model can be used to estimate parameters, test treatment effects, and predict unobservables. It also provides a flexible method for mapping non-normal data, which includes indicator kriging and universal kriging as special cases.

There are several areas for future research suggested by these methods. These include:

1. The methods for interval estimation and hypothesis testing assume asymptotic normality. What are their small-sample properties? When are they well or poorly behaved?
2. When $\hat{\beta}$ is nonlinear, assessing the uncertainty associated with predictions based on Equation (3.6) uncertainty can be difficult. Even when $\hat{\beta}$ is linear in the data,

as in universal kriging, prediction mean-squared errors depend only on the data locations and not on the data values themselves. This makes prediction mean-squared errors difficult to interpret as a measure of uncertainty in predictions, and unfortunately, has caused many researchers to stop mentioning prediction error at all. Clearly, a better assessment of prediction error is needed.

3. Are there experimental designs or sampling strategies that would improve the efficiency of desired estimates or hypothesis tests with spatially-correlated, non-normal data?

The generalized linear model approach allows statistical methods for spatially correlated non-normal data to be viewed as a natural aspect of existing linear model theory, rather than a "case apart." The methods presented here greatly increase the options available to workers in a variety of disciplines that already use spatial statistics, including plant and soil science, environmental science, agricultural engineering, hydrology, geology, and ecology. Additionally, areas that may not currently make extensive use of spatial statistics, possibly because they make heavy use of non-normal (e.g., categorical) data can make use of the methods discussed in this article. For example, for disciplines such as epidemiology or sociology, where spatial effects almost surely play a role in many studies, these methods should prove to be of great value.

APPENDIX: DATA FOR EXAMPLE 1

Table A.1. Data for Example 1

<i>Block</i>	<i>Entry</i>	<i>LAT</i>	<i>LNG</i>	<i>N_{ij}</i>	<i>z_{ij}</i>
1	14	1	1	8	2
1	16	1	2	9	1
1	7	1	3	13	9
1	6	1	4	9	9
1	13	2	1	9	2
1	15	2	2	14	7
1	8	2	3	8	6
1	5	2	4	11	8
1	11	3	1	12	7
1	12	3	2	11	8
1	2	3	3	10	8
1	3	3	4	12	5
1	10	4	1	9	7
1	9	4	2	15	8
1	4	4	3	19	6
1	1	4	4	8	7
2	15	5	1	15	6

(Cont'd next page)

Table A.1. (Cont'd) Data for Example 1

<i>Block</i>	<i>Entry</i>	<i>LAT</i>	<i>LNG</i>	N_{ij}	z_{ij}
2	3	5	2	11	9
2	10	5	3	12	5
2	2	5	4	9	9
2	11	6	1	20	10
2	7	6	2	10	8
2	14	6	3	12	4
2	6	6	4	10	7
2	5	7	1	8	8
2	13	7	2	6	0
2	12	7	3	9	2
2	16	7	4	9	0
2	9	8	1	14	9
2	1	8	2	13	12
2	8	8	3	12	3
2	4	8	4	14	7
3	7	1	5	7	7
3	13	1	6	7	0
3	8	1	7	13	3
3	14	1	8	9	0
3	4	2	5	15	11
3	10	2	6	9	7
3	3	2	7	15	11
3	9	2	8	13	5
3	6	3	5	16	9
3	1	3	6	8	8
3	15	3	7	7	0
3	12	3	8	12	8
3	11	4	5	8	1
3	16	4	6	15	1
3	5	4	7	12	7
3	2	4	8	16	12
4	9	5	5	15	8
4	4	5	6	10	6
4	12	5	7	13	5
4	1	5	8	15	9
4	15	6	5	17	6
4	6	6	6	8	2
4	14	6	7	12	5
4	7	6	8	15	8
4	13	7	5	13	2
4	8	7	6	13	9
4	3	7	7	9	9
4	10	7	8	6	6
4	2	8	5	12	8
4	11	8	6	9	7
4	5	8	7	11	10
4	16	8	8	15	7

NOTE: LAT and LNG give the spatial position of each plot. N_{ij} is the number of plants observed on the ij th plot and z_{ij} is the number of damaged plants on the ij th plot.

[Received April 1996. Revised February 1997.]

REFERENCES

- Agresti, A. (1990), *Categorical Data Analysis*, New York: Wiley.
- Albert, P. A. and McShane, L. M. (1995), "A Generalized Estimating Equations Approach for Spatially Correlated Data: Applications to the Analysis of Neuroimaging Data," *Biometrics*, 51, 627–638.
- Cressie, N. (1985), "Fitting Variogram Models by Weighted Least Squares," *Journal of the International Association for Mathematical Geology*, 17, 563–586.
- (1986), "Kriging Nonstationary Data," *Journal of the American Statistical Association*, 81, 625–634.
- (1991), *Statistics for Spatial Data*, New York: Wiley.
- Dowd, P. A. (1982), "Lognormal Kriging—the General Case," *Journal of the International Association for Mathematical Geology*, 14, 475–499.
- Deutsch, C. V., and Journel, A. G. (1992), *GSLIB: Geostatistical Software Library and User's Guide*, New York: Oxford University Press.
- Goldberger, A. S. (1962), "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Statistical Association*, 57, 369–375.
- Goodnight, J. H. (1979), "A Tutorial on the Sweep Operator," *American Statistician*, 33, 149–158.
- Gotway, C. A., and Cressie, N. (1993), "Improved Multivariate Prediction Under a General Linear Model," *Journal of Multivariate Analysis*, 45, 56–72.
- Harper, W. V., and Furr, J. M. (1986), "Geostatistical Analysis of Potentiometric Data in the Wolfcamp Aquifer of the Palo Duro Basin, Texas," Technical Report BMI/ONWI-587, Battelle Memorial Institute, Columbus, OH.
- Johnson, G. A., Mortensen, D. A., and Gotway, C. A. (1996), "Spatial and Temporal Analysis of Weed Seedling Populations Using Geostatistics," *Weed Science*, 44, 704–710.
- Journel, A. G. (1983), "Nonparametric Estimation of Spatial Distributions," *Journal of the International Association for Mathematical Geology*, 15, 445–468.
- Journel, A. G., and Huijbregts, C. J. (1978), *Mining Geostatistics*, London: Academic Press.
- Liang, K., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.
- McCullagh, P. (1983), "Quasi-Likelihood Functions," *Annals of Statistics*, 11, 59–67.
- McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), New York: Chapman & Hall.
- Nelder, J. A., Wedderburn, R. W. M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society, Ser. A*, 135, 370–384.
- Stroup, W. W., Baenziger, P. S., and Mulitze, D. K. (1994), "Removing Spatial Variation From Wheat Yield Trials: A Comparison of Methods," *Crop Science*, 34, 62–66.
- Wedderburn, R. W. M. (1976), "Quasi-likelihood Functions, Generalized Linear Models and the Gauss–Newton Method," *Biometrika*, 61, 439–447.
- Wiles, L. J., Oliver, G. W., York, A. C., Gold, H. J., and Wilkerson, G. G., (1992), "Spatial Distribution of Broadleaf Weeds in North Carolina Soybean (*Glycine max*) Fields," *Weed Science*, 40, 554–557.
- Wold, H. (1938), *A Study in the Analysis of Stationary Time Series*, Uppsala: Almqvist & Wiksell.
- Zeger, S. L., and Liang, K-Y. 1986, "Longitudinal Data Analysis for Discrete and Continuous Outcomes," *Biometrics*, 42, 121–130.
- Zimmerman, D. L., and Harville, D. A. (1991), "A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments," *Biometrics*, 47, 223–239.