Supplementary materials for this article are available at http://pubs.amstat.org/toc/jabes/14/4.

Maximum Likelihood Estimation of Regression Parameters With Spatially Dependent Discrete Data

L. MADSEN

Generalized estimating equations (GEEs) have been successfully used to estimate regression parameters from discrete longitudinal data. GEEs have been adapted for spatially correlated count data with less success. It is convenient to model correlated counts as lognormal-Poisson, where a latent lognormal random process carries all correlation. This model limits correlation and can lead to negative bias of standard errors. Moreover, correlation is not the best dependence measure for highly nonnormal data. This article proposes a model which yields maximum likelihood (ML) estimates of regression parameters when the response is discrete and spatially dependent. This model employs a spatial Gaussian copula, bringing the discrete distribution into the Gaussian geostatistical framework, where correlation completely describes dependence. The model yields a log-likelihood for regression parameters that can be maximized using established numerical methods. The proposed procedure is used to estimate the relationship between Japanese beetle grub counts and soil organic matter. These data exhibit residual correlation well above the lognormal-Poisson correlation limit, so that model is not appropriate. The data and MATLAB code are available online. Simulations demonstrate that negative bias in GEE standard errors leads to nominal 95% confidence coverage less than 62% for moderate or strong spatial dependence, whereas ML coverage remains above 82%.

Key Words: Continuous extension; Correlated count data; Dependent count data; Gaussian copula; Spatial copula.

1. INTRODUCTION

Dependent discrete data arise in many disciplines. Medical studies record clustered categorical responses. Meteorological data include time series of counts of extreme events or numbers of times a threshold is exceeded. Environmental scientists observe spatially dependent counts of organisms. Dependent binary data are ubiquitous in many fields. A common inferential goal is to estimate the relationship between a clustered or otherwise de-

L. Madsen is Assistant Professor of Statistics, Oregon State University, Corvallis, OR 97331 (E-mail: *madsenL@onid.orst.edu*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 14, Number 4, Pages 375–391
DOI: 10.1198/jabes.2009.07116

pendent discrete response and a set of explanatory variables. This article proposes a maximum likelihood estimator for regression parameters from a spatially dependent discrete response.

When a spatially dependent response can reasonably be assumed to be Gaussian, the regression problem falls into the geostatistical framework, and techniques such as weighted least squares or maximum likelihood are natural procedures for estimating the regression coefficients (Schabenberger and Gotway 2005, section 6.2). When the response is discrete, researchers typically turn to the generalized estimating equation (GEE) methodology of Liang and Zeger (1986) and Zeger and Liang (1986). Some authors have adapted GEEs for nonnormal spatial data (Albert and McShane 1995; Gotway and Stroup 1997; McShane, Albert, and Palmatier 1997).

The GEE approach is unsatisfactory for three reasons. First, spatial GEE models typically assume that, conditional on a spatially correlated latent variable, the responses are independent. For count data a lognormal-Poisson latent process model is commonly used (Albert and McShane 1995; Gotway and Stroup 1997; McShane, Albert, and Palmatier 1997) wherein the latent process carries the spatial correlation. Specifically, let $\varepsilon(\cdot)$ be a lognormal isotropic stationary spatial process. Let $\mathbf{x}(s)$ denote a covariate vector at location s and let $\boldsymbol{\beta}$ be a vector of regression parameters. The ε process is designed to model spatial dependence, so it is convenient to set $E\{\varepsilon(s)\} = 1$ and to model

$$\text{cov}\{\varepsilon(s), \varepsilon(s+h)\} = \sigma_\varepsilon^2 \rho_\varepsilon(h; \boldsymbol{\theta})$$

for some spatial correlation function ρ_ε depending on distance h and a vector of covariance parameters $\boldsymbol{\theta}$. Conditional on the $\varepsilon(s)$, let $Y(s)$ be independent Poisson with mean depending on $\mathbf{x}'(s)\boldsymbol{\beta}$:

$$E\{Y(s)|\varepsilon(s)\} = \exp\{\varepsilon(s) \cdot \mathbf{x}'(s)\boldsymbol{\beta}\}.$$

These assumptions induce marginal moments:

$$\begin{aligned} E\{Y(s)\} &= \exp\{\mathbf{x}'(s)\boldsymbol{\beta}\} \equiv \mu(s), \\ \text{var}\{Y(s)\} &= \mu(s) + \sigma_\varepsilon^2 \mu^2(s), \qquad (1.1) \\ \text{corr}\{Y(s), Y(s+h)\} &= \rho_\varepsilon(h; \boldsymbol{\theta}) \left[\left\{ 1 + \frac{1}{\sigma_\varepsilon^2 \mu(s)} \right\} \left\{ 1 + \frac{1}{\sigma_\varepsilon^2 \mu(s+h)} \right\} \right]^{-1/2}. \end{aligned}$$

Marginally, this lognormal-Poisson model closely resembles a negative binomial model, as suggested by the mean-variance relationship given in Equation (1.1) which has the familiar form:

$$\text{var}\{Y(s)\} = \mu(s) + r \cdot \mu(s)^2. \qquad (1.2)$$

However, the correlation structure between the two models is very different. As discussed in Section 3, the latent process lognormal-Poisson model imposes limits on correlations that are well below the theoretical maximum for marginal negative binomial random variables. When the model limits the degree of dependence below that in the data, variance estimates will be negatively biased because the model assumes more independent information than actually exists. Section 6 illustrates a situation where the lognormal-Poisson latent variable model yields nominal 95% confidence intervals with actual coverage probabilities of less than 40%.

Yasui and Lele (1997) developed an estimating function approach for estimating regression parameters from a lognormal-Poisson latent variable model. Diggle, Tawn, and Moyeed (1998) proposed a Bayesian estimation method for this model. These methods are subject to the limits on correlation imposed by the lognormal-Poisson model.

The second problem with GEE methodology is that it models correlation, i.e., *linear* dependence. For normal random variables linear dependence is equivalent to dependence, whereas correlation may not be the right measure of dependence for highly nonnormal data. Spatial discrete data are often small-mean counts which have highly skewed distributions that cannot be approximated with a normal distribution.

Lin and Clayton (2005) developed a quasi-likelihood approach to spatially correlated binary data. Diggle, Tawn, and Moyeed (1998) introduced a binary latent variable model. Neither of these models limit correlations below the theoretical limits for binary random variables. However, both these approaches seek to model dependence between Bernoulli responses by correlation, which is not particularly appealing.

The third problem with GEE estimators is that they are not asymptotically efficient. McCullagh and Nelder (1989) gave an example where the quasi-likelihood estimator is considerably less efficient than the maximum likelihood estimator. Song (2007, section 6) gave three cases where the GEE estimator is less efficient than a vector generalized linear model.

This article introduces a maximum likelihood (ML) approach to estimation from a spatially dependent discrete response. Dependent discrete data are brought into the geostatistical framework by means of a Gaussian copula model where dependence is modeled as correlation, so that well-known geostatistical correlation structures can be used (see, e.g., Cressie 1993, section 2.3.1). The Gaussian copula places no artificial limits on dependence and can model correlations up to the theoretical maximum. Furthermore, the Gaussian copula yields a likelihood for the regression parameters which can be maximized to obtain estimates of those parameters. Simulations suggest that the ML estimators are as efficient as the latent process lognormal-Poisson GEE estimators for weak spatial dependence, and the ML estimators become more efficient than the GEE estimators as spatial dependence increases.

Copula models (Joe 2001; Nelsen 2006) are founded in the work of Hoeffding (1940) and have been used to model dependence and to construct multivariate distributions (Fisher 1997). Recent work by Pitt, Chan, and Kohn (2006) and Hoff (2007) explored Bayesian estimation using the Gaussian copula to link dependent observations with given marginal distributions. Pitt, Chan, and Kohn (2006) focused on estimating parameters of the marginal distributions whereas Hoff (2007) was concerned with estimating the dependence parameters. Song, Li, and Yuan (2008) illustrated a maximum likelihood estimation approach using the Gaussian copula model applied to longitudinal data.

Although the focus of this article is to analyze spatially dependent discrete data, the method is general and can be applied to other dependent discrete data including space-time problems and longitudinal data.

2. MOTIVATING EXAMPLE

Dalthorp (2004) discussed a study relating counts of Japanese beetle (*Popillia japonica*) grubs to soil organic matter on a golf course near Geneva, New York. A grub density exceeding 8 to 12 grubs per square foot can cause turfgrass damage. Effective control is expensive and must be done before damage occurs, therefore, modeling grub counts as a function of a covariate such as soil organic matter, a proxy for soil moisture, temperature, and thatch thickness, could be useful to turfgrass managers as well as informative for insect ecologists.

Figure 1 shows smoothed aerial views of organic matter and grub counts for the 142 observations in the dataset. A relationship between the two measurements is visible, with high grub counts generally occurring with lower organic matter levels.

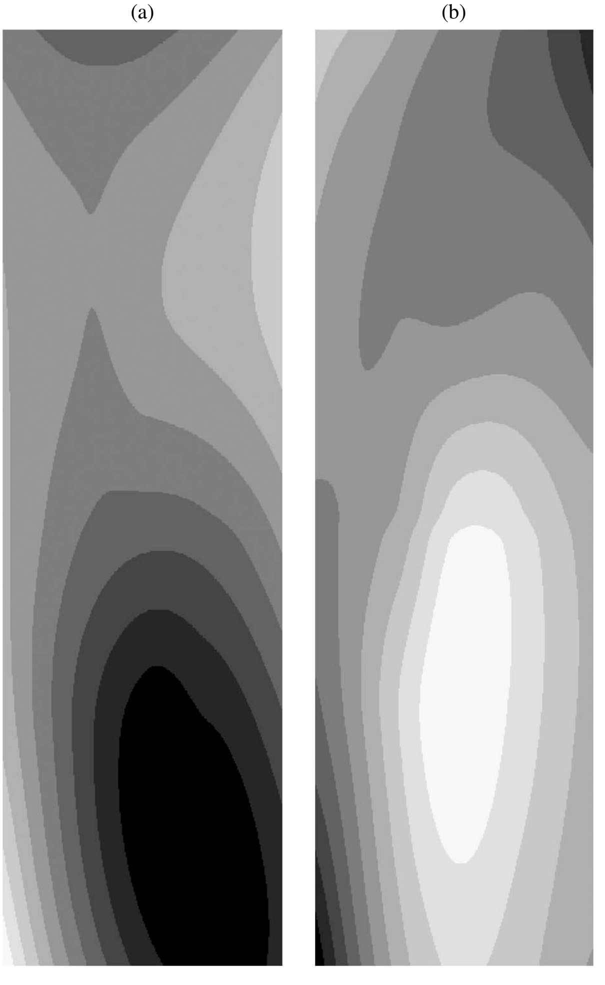

Figure 1. Smoothed aerial views of the soil organic matter (a) and grub counts (b) on the observed fairway. Dark colors represent smaller values. Higher grub counts tend to occur at lower levels of organic matter, but there is clustering among the grubs that is not explained by the spatial pattern in organic matter.

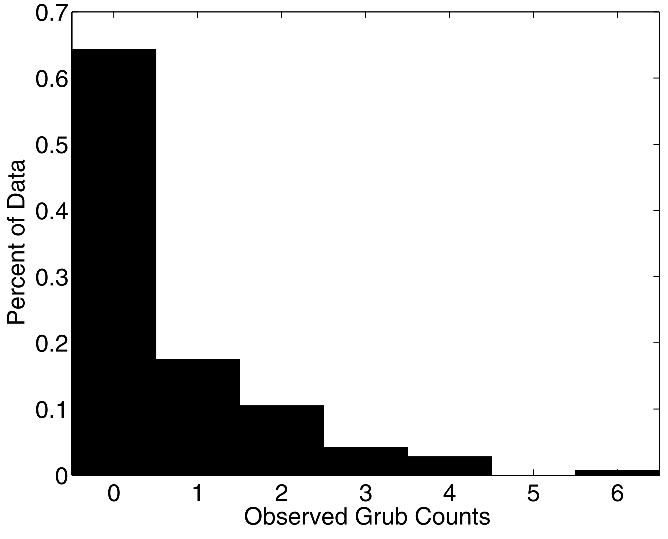

Figure 2. Histogram of observed grub counts. The distribution is quite skewed, with more than half of the observations equal to zero, so the data cannot be made symmetric with a transformation.

Figure 2 shows that the distribution of the counts is extremely skewed with zeros comprising over half of the observations, so no simple normalizing transformation is available, and a standard regression analysis is inappropriate.

The negative binomial model is a flexible probability model for overdispersed counts and can be used to model ecological count data under different biological assumptions (Solomon 1983). When the data are independent and the parameter r in Equation (1.2) is assumed constant, a generalized linear model approach can be used to estimate regression parameters (Venables and Ripley 2002).

The independence assumption is unreasonable for the grub data. Figure 1 shows clustering in the grub counts that is not explained by the spatial pattern in organic matter. This residual spatial dependence can be explained by the aggregation behavior of the adult beetles (Dalthorp, Nyrop, and Villani 2000) and should be modeled in the estimation procedure to avoid erroneous inference.

Using a GEE approach to account for the spatial correlation, Dalthorp (2004) found that the exponential of a cubic function of soil organic matter fits the observed mean grub counts. Madsen and Dalthorp (2007) observed that the sample correlations of these data are as high as 0.19, whereas the lognormal-Poisson model cannot accommodate correlations higher than 0.11. Details of the correlation limits are given in Section 3. The data are small-mean counts (the sample mean is 0.66 and the sample maximum is 6), so modeling correlation would be inappropriate regardless of correlation limits. In Section 5, the data are analyzed using the method proposed in this article. The data may be obtained from the JABES Data and Program Archive at *http://www.amstat.org/publications/jabes*.

3. LIMITS TO CORRELATION

Because correlation describes linear dependence between random variables, it is most appropriate as a measure of dependence between Gaussian random variables or those whose distribution may be well-approximated by a normal distribution. For binary random variables with means close to 0 or 1, or for small-mean count random variables, a normal approximation is not reasonable, and modeling correlation becomes complicated.

Correlation between nonnormal random variables is bounded above by a limit that may be strictly less than 1. For random variables Y_1 and Y_2 with marginal distributions $Y_1 \sim F$ and $Y_2 \sim G$, the limits on $\mathrm{corr}(Y_1, Y_2)$ can be calculated from the Fréchet–Hoeffding bounds (Nelsen 2006, p. 11). If the Y_i are discrete random variables, the upper limit is

$$\rho_U = \frac{\sum_{(y_1,y_2)\in A_1}[1 - G(y_2)] + \sum_{(y_1,y_2)\in A_2}[1 - F(y_1)] - \mu_{Y_1}\mu_{Y_2}}{\sigma_1\sigma_2}, \tag{3.1}$$

where $\mu_i = E(Y_i)$, $\sigma_i^2 = \mathrm{var}(Y_i)$, $A_1 = \{(y_1, y_2) : F(y_1) \le G(y_2)\}$, and $A_2 = \{(y_1, y_2) : F(y_1) > G(y_2)\}$.

Prentice (1988) noted that for marginally binary random variables Y_1 and Y_2 with $P(Y_i = 1) = p_i$,

$$\mathrm{corr}(Y_1, Y_2) \le \min\left[\left\{\frac{p_1(1-p_2)}{p_2(1-p_1)}\right\}^{1/2}, \left\{\frac{p_2(1-p_1)}{p_1(1-p_2)}\right\}^{1/2}\right].$$

Madsen and Dalthorp (2007) observed that latent process lognormal-Poisson random variables are maximally correlated when the latent lognormal random variables are maximally correlated. For latent process lognormal-Poisson random variables Y_1 and Y_2 with means and variances μ_{Y_i} and $\sigma^2_{Y_i}$, respectively, the upper bound is

$$\mathrm{corr}(Y_1, Y_2) \le \frac{\mu_{Y_1}\mu_{Y_2}}{\sigma_{Y_1}\sigma_{Y_2}}\left[\exp\{\sqrt{\log(c_1)\log(c_2)}\} - 1\right], \tag{3.2}$$

where $c_i = 1 + \mu_{Y_i}^{-2}(\sigma^2_{Y_i} - \mu_{Y_i})$. Madsen and Dalthorp (2007) gave two ecological datasets with sample moments that violate this bound. One of these datasets is the Japanese beetle data introduced above in Section 2.

As mentioned in Section 1, the lognormal-Poisson distribution closely resembles the negative binomial if only the marginal distributions are considered. However, marginally negative binomial random variables can achieve much higher correlations than latent process lognormal-Poisson random variables (Madsen and Dalthorp 2007). Figure 3 compares the maximum possible correlation between two latent process lognormal-Poisson random variables [Figure 3(a)] and between two negative binomial random variables [Figure 3(b)] with means and variances similar to those observed in the Japanese beetle grub data of Section 2. The correlation limit for latent process lognormal-Poisson random variables is given in Equation (3.2), whereas that for negative binomial random variables is from Equation (3.1). Empirical residual correlation in the grub data exceeds the latent process lognormal-Poisson upper bounds. Intuitively, Equation (3.2) is small because the conditionally independent Poissons add too much variance onto the correlated conditional means. When the maximum correlation allowed by the model is smaller than the correlation occurring in the data-generating process, standard errors may be too small because the model does not allow the proper correction for the lack of independence.

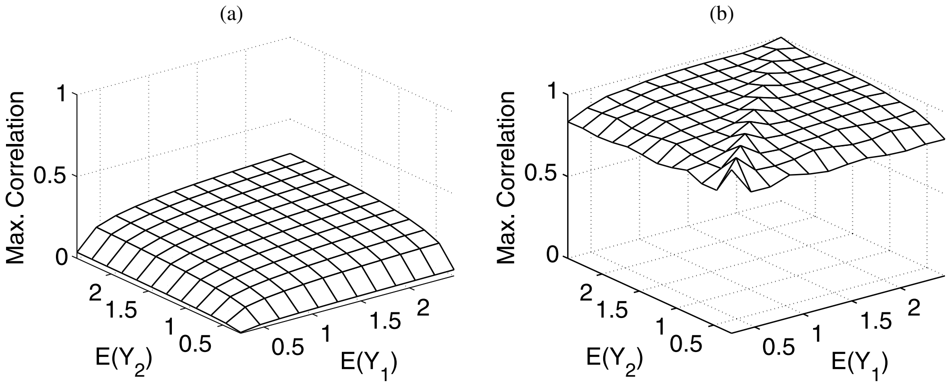

Figure 3. Maximum corr(Y_1, Y_2) for (a) lognormal-Poisson Y_i and (b) negative binomial Y_i as a function of means $E(Y_i)$. The ranges for $E(Y_i)$ are from the Japanese beetle data example of Section 2. $E(Y_i)$ ranges between 0.26 and 2.5, and variances are $\text{var}(Y_i) = 1.228\{E(Y_i)\}^{1.148}$ (Madsen and Dalthorp 2007). Observed correlations in these data exceed the lognormal-Poisson bounds. The bounds in (a) are given by (3.2) and the bounds in (b) are given by (3.1).

4. A NEW MODEL FOR DEPENDENT DISCRETE DATA

Given random variables Y_1 and Y_2 with continuous marginal distributions F_1 and F_2, respectively, the maximum possible correlation between Y_1 and Y_2 can be achieved via a Gaussian copula

$$C(y_1, y_2; \delta) = \Phi_\delta\big[\Phi^{-1}\{F_1(y_1)\}, \Phi^{-1}\{F_2(y_2)\}\big], \tag{4.1}$$

by setting $\delta = 1$, where Φ is the standard normal cdf and Φ_δ is the bivariate normal cdf with correlation δ (Joe 2001, pp. 140–141). $C(y_1, y_2; \delta)$ gives a joint distribution function of Y_1 and Y_2 with marginal distributions F_1 and F_2 and dependence determined by the parameter δ. Song (2007, p. 130) called δ the "normal scoring" between nonnormal Y_1 and Y_2 and discusses the connection between δ and two other measures of association, Spearman's ρ and Kendall's τ.

Unlike many bivariate copula models, the Gaussian copula easily generalizes to the multivariate setting. Let $Y_1, \ldots, Y_n$ be random variables with continuous marginal distributions F_i and density functions f_i. Let $\mathbf{\Sigma}$ be a nonnegative definite matrix with diagonal entries equal to 1, a valid correlation matrix. Then a joint distribution of $Y_1, \ldots, Y_n$ with the specified marginals is

$$C(\mathbf{y}; \mathbf{\Sigma}) = \Phi_{\mathbf{\Sigma}}\big[\Phi^{-1}\{F_1(y_1)\}, \ldots, \Phi^{-1}\{F_n(y_n)\}\big], \tag{4.2}$$

where $\Phi_{\mathbf{\Sigma}}$ is the multivariate normal cdf with covariance matrix $\mathbf{\Sigma}$. Differentiating $C(y_1, \ldots, y_n)$ yields joint density function

$$c(\mathbf{y}; \mathbf{\Sigma}) = |\mathbf{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{\Sigma}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \cdot \prod_{i=1}^{n} f_i(y_i), \tag{4.3}$$

where $\mathbf{z} = [\Phi^{-1}\{F_1(y_1)\}, \ldots, \Phi^{-1}\{F_n(y_n)\}]'$ and $\mathbf{I}_n$ denotes the $n \times n$ identity matrix.

If the Y_i are discrete, the joint probability mass function of $Y_1, \ldots, Y_n$ is

$$
\begin{aligned}
g(\mathbf{y}; \boldsymbol{\Sigma}) &= P(Y_1 = y_1, \ldots, Y_n = y_n) \\
&= \sum_{j_1=1}^{2} \cdots \sum_{j_n=1}^{2} (-1)^{j_1+\cdots+j_n} \Phi_{\boldsymbol{\Sigma}}\left[\Phi^{-1}\{u_{1j_1}\}, \ldots, \Phi^{-1}\{u_{nj_n}\}\right],
\end{aligned}
\tag{4.4}
$$

where $u_{i1} = F_i(y_i)$ and $u_{i2} = F_i(y_i-)$, the limit of F_i at y_i from the left (Song 2000). For count or binomial random variables, $F_i(y_i-) = F_i(y_i - 1)$.

Equation (4.4) contains 2^n terms and is intractable for n larger than 4 or 5. However, the n-fold summation in Equation (4.4) can be avoided by using a continuous extension of the Y_i proposed by Denuit and Lambert (2005). Associate with discrete Y_i a continuous random variable

$$
Y_i^* = Y_i - U_i, \tag{4.5}
$$

where U_i follows a continuous uniform distribution on (0, 1) independent of Y_i and of U_j for $j \neq i$. Then Y_i^* is a continuous random variable with distribution function

$$
F_i^*(y) = F_i([y]) + (y - [y])P(Y_i = [y + 1]),
$$

and density

$$
f_i^*(y) = P(Y_i = [y + 1]), \tag{4.6}
$$

where $[y]$ denotes the integer part of $y \in \mathbb{R}$.

Note that no information is lost by continuously extending Y_i in Equation (4.5) since Y_i can be recovered from Y_i^* as

$$
Y_i = [Y_i^* + 1].
$$

Furthermore, Denuit and Lambert (2005) proved that the continuous extension preserves Kendall's τ, and Kruskal (1958) gave Kendall's τ in terms of the parameter δ in Equation (4.1). Thus Y_i^* and Y_j^* have the same dependence relationship as Y_i and Y_j.

If $Y(\cdot)$ is an isotropic discrete random spatial process observed at locations $s_1, \ldots, s_n$, then the multivariate Gaussian copula can be used to model the joint distribution of $(Y_1, \ldots, Y_n) = [Y(s_1), \ldots, Y(s_n)]$ by giving the copula correlation matrix $\boldsymbol{\Sigma}$ a spatial form. Let $\rho(h)$ be an isotropic parametric correlogram (Cressie 1993, p. 67) depending on a vector of parameters $\boldsymbol{\theta}$ and a distance h. Define $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ to be the $n \times n$ correlation matrix with ijth element

$$
\Sigma_{ij}(\boldsymbol{\theta}) = \rho(\|s_i - s_j\|). \tag{4.7}
$$

Let Y_i^* be the continuous extension of Y_i given in Equation (4.5). The joint distribution of $Y_1^*, \ldots, Y_n^*$ can be modeled via the multivariate Gaussian copula of Equation (4.2) with correlation matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Because $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ has a spatial form, the copula model accounts for the spatial dependence among $Y_1, \ldots, Y_n$.

The density found in Equation (4.3) forms a likelihood for correlation parameters $\boldsymbol{\theta}$, regression parameters $\boldsymbol{\beta}$, and, if necessary, a scale parameter ϕ of the marginal densities f_i^*. From Equation (4.6), $\boldsymbol{\beta}$ and ϕ are exactly the parameters of the marginal distributions of $(Y_1, \ldots, Y_n)$. Because Equation (4.3) depends on the ancillary U_i, we integrate over

$\mathbf{U} = (U_1, \ldots, U_n)$ and take expected likelihood

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi; \mathbf{y}) = E\left\{ \frac{\exp[-(1/2)\mathbf{z}^{*\prime}(\boldsymbol{\Sigma}^{-1} - \mathbf{I}_n)\mathbf{z}^*] \prod_{i=1}^{n} f_i^*(y_i^*)}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \middle| \mathbf{y} \right\}, \tag{4.8}$$

where $\mathbf{z}^* = [\Phi^{-1}\{F_1^*(y_1^*)\}, \ldots, \Phi^{-1}\{F_n^*(y_n^*)\}]'$ and $\mathbf{y}$ denotes the data vector. It can be shown that Equation (4.8) is equal to the true copula joint probability mass function of Equation (4.4). Parameter vector $\boldsymbol{\xi} = (\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$ can be estimated by maximizing the log of Equation (4.8):

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\phi}) = \arg\max_{\boldsymbol{\beta}, \boldsymbol{\theta}, \phi} \log\{L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi; \mathbf{y})\}. \tag{4.9}$$

Under regularity conditions (see, e.g., Mardia and Marshall 1984), the MLEs $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\phi}$ will be consistent and asymptotically normal with asymptotic covariance matrix given by the Fisher information $-\{E(\mathbf{H})\}^{-1}$ where the ijth element of $\mathbf{H}$ is $H_{ij} = \partial^2 L/(\partial \xi_i \, \partial \xi_j)$.

Some authors noted difficulty in estimating covariance parameters (Berger, de Oliveira, and Sansó 2001; Zhang 2004; Irvine, Gitelman, and Hoeting 2007) using the Gaussian geostatistical model. Lee, Nelder, and Pawitan (2006) devoted much attention to estimating the dispersion parameter in the generalized linear model framework. Because the purpose of this article is to estimate the regression parameters $\boldsymbol{\beta}$, covariance parameters $\boldsymbol{\theta}$ and dispersion parameter ϕ are nuisance parameters, and are only used to account for spatial dependence and overdispersion, respectively. Section 6 includes a brief discussion of how well the residual dependence is estimated by $\hat{\boldsymbol{\theta}}$ and the excess variability by $\hat{\phi}$.

5. EXAMPLE

The Japanese beetle grub data, introduced in Section 2, were analyzed using the proposed method. Details of the implementation and results are given here. MATLAB code is available in the Supplemental Materials

The observed grub counts $y_1, \ldots, y_{142}$ are overdispersed counts, so we assume a marginal negative binomial distributions with means

$$\mu_i = \exp(\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3),$$

where x_i is the observed percent soil organic matter content at the ith location, as in Dalthorp (2004). The probability mass function of Y_i is

$$p(y, \phi, \mu_i) = \frac{\Gamma(y + \phi\mu_i)}{y!\Gamma(\phi\mu_i)} \cdot \frac{\phi^{\phi\mu_i}}{(1+\phi)^{y+\phi\mu_i}}, \tag{5.1}$$

where $\Gamma(\cdot)$ is the gamma function and ϕ is the "overdispersion" parameter, i.e.,

$$\text{var}(Y_i) = \mu_i \cdot \frac{1+\phi}{\phi}. \tag{5.2}$$

The usual parameterization of the negative binomial probability mass function leads to mean-variance relationship in Equation (1.2). The parameterization in Equation (5.1) is discussed in McCullagh and Nelder (1989, p. 199). Parameters ϕ and r are related as $r = \phi \cdot \mu$.

Using Taylor's Power Law (Taylor 1961), Dalthorp (2004) concluded that $\text{var}(Y_i) = a\mu_i^b$, where a is between 1.23 and 2.10, and b is between 1.15 and 1.24. The mean-variance relationship in Equation (5.2) assumes that $b = 1$. This is done for simplicity, since the focus is estimating $\boldsymbol{\beta}$. A more flexible model could be obtained by letting ϕ vary, either spatially or as a function of the covariates (Hilbe 2007).

The correlogram $\rho(h)$ is assumed to be exponential with two parameters so that the ijth element of the correlation matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ in Equation (4.7) is

$$\Sigma_{ij}(\boldsymbol{\theta}) = \begin{cases} \theta_0 \exp(-h_{ij}\theta_1), & i \neq j \\ 1, & i = j, \end{cases} \tag{5.3}$$

where h_{ij} is the distance between the locations of y_i and y_j, $0 < \theta_0 \leq 1$ is the "nugget" parameter, and $\theta_1 > 0$ is the "decay" parameter.

The inferential goal is to estimate the β_k. Maximum likelihood estimates (MLEs) of β_k, ϕ, and θ_j are obtained by numerically maximizing the log of expected likelihood of Equation (4.8) with respect to $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and ϕ. The log expected likelihood is approximated as

$$\log L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi; \mathbf{y}) \approx \log\left(\frac{1}{m}\sum_{j=1}^{m}\left[\frac{\exp\{-(1/2)\mathbf{z}_j^{*\prime}(\boldsymbol{\Sigma}^{-1} - \mathbf{I}_{142})\mathbf{z}_j^*\}\prod_{i=1}^{n} f_i^*(y_i^*)}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}\right]\right),$$

where $\mathbf{z}_j^* = [\Phi^{-1}\{F_1^*(y_1 - u_{1,j})\}, \ldots, \Phi^{-1}\{F_n^*(y_n - u_{n,j})\}]'$, and the $u_{i,j}$ are independent uniform on $(0, 1)$ for $i = 1, \ldots, 142$ and $j = 1, \ldots, m$. In general, m can be chosen by comparing estimates and standard errors from repetitions of the estimation procedure with various m. For example, if repeated estimates and standard errors are approximately equal for $m = 1000$, this suggests that $m = 1000$ is sufficient for the data. To obtain the results reported here, m was taken to be 1000, since the coefficient of variation for point estimates and standard errors were all less than 0.02 among 10 runs of the estimation algorithm with $m = 1000$.

The computational intensity of this estimation procedure increases with m. The average time for $m = 1000$ was just under 2 min on a quad core 2.4 GHz desktop computer. The computational burden is primarily in calculating the $z_{i,j}^* = \Phi^{-1}\{F_i^*(y_i - u_{i,j})\}$ and their derivatives, though these operations may be vectorized in languages such as MATLAB and R. Note that the dimension of $\boldsymbol{\Sigma}$ does not change and $\boldsymbol{\Sigma}^{-1}$ only needs to be calculated once, regardless of m. Further research may yield ways to streamline the computation. For example, faster approximations of $\Phi^{-1}(\cdot)$ and $F_i(\cdot)$ might be used initially and more precise approximations used as the optimization approaches convergence.

Variance estimates are obtained by numerically approximating the Hessian matrix $\mathbf{H}$ at the MLE and taking $\widehat{\text{var}}(\hat{\boldsymbol{\xi}}) = -\hat{\mathbf{H}}^{-1}$, so that $\widehat{\text{var}}(\hat{\boldsymbol{\xi}})$ is a numerical approximation of the observed information matrix.

Figure 4 shows the data with the fitted mean functions from ML estimation and GEE estimation. The two curves are very similar; the average squared difference in fitted values is $(142)^{-1}\sum_{i=1}^{142}(\hat{\mathbf{y}}_{ML} - \hat{\mathbf{y}}_{GEE})^2 = 0.0035$. Point estimates and standard errors of the β_k for both ML and GEE are given in Table 1. ML point estimates of the β_k are all slightly smaller than the GEE point estimates, but the GEE standard errors are only about 80% of the ML standard errors. The simulation study in Section 6 suggests that the GEE standard errors are too small.

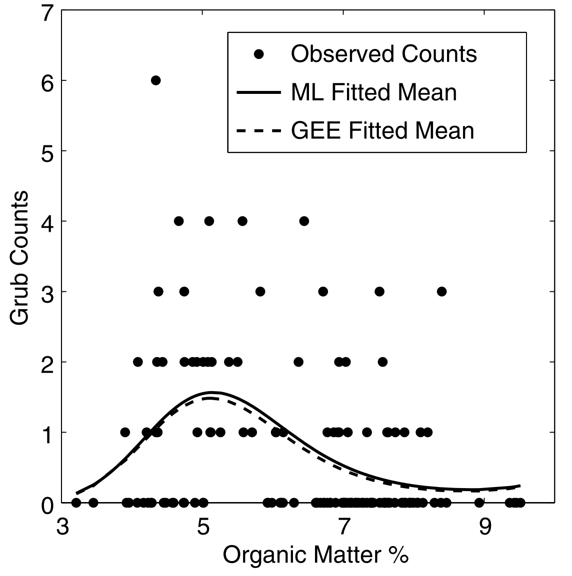

Figure 4. Plot of observed grub counts as a function of percent soil organic matter. Superimposed is the fitted mean function from both estimation procedures.

Individual hypotheses $H_0 : \beta_k = 0$ may be tested via a Wald statistic $W = \hat{\beta}_k / \mathrm{SE}(\hat{\beta}_k)$, where W follows an approximate standard normal distribution under the null hypothesis (Lee, Nelder, and Pawitan 2006, p. 22). Except for β_3 under ML estimation, all $|W| > 2$ so, with a significance level of $\alpha = 0.05$, the GEE analysis would conclude that a cubic function of organic matter is necessary, whereas the ML analysis would conclude that a quadratic function is sufficient. Of interest to turfgrass managers is the expected number of grubs given a particular percent organic matter x_0. A nominal 95% confidence interval for mean grub count for a given x_0 is

$$\hat{\mu}_0 \pm 1.96\sqrt{\mathbf{x}_0' \widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}})\mathbf{x}_0}$$

where $\mathbf{x}_0 = [1, x_0, x_0^2, x_0^3]'$ and $\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}})$ is the submatrix of $\widehat{\mathrm{var}}(\hat{\boldsymbol{\xi}})$ corresponding to $\boldsymbol{\beta}$.

Fitted correlation parameters gave a residual correlogram comparable to that in Dalthorp (2004, figure 9). The point estimate of overdispersion parameter ϕ was 51.02, which gave

Table 1. Estimates and standard errors for mean function parameters from the Japanese beetle data using both estimation procedures.

	Estimate (SE)	
Parameter	ML	GEE
β_0	−24.34 (11.63)	−25.07 (9.49)
β_1	11.96 (5.71)	12.36 (4.66)
β_2	−1.84 (0.91)	−1.91 (0.74)
β_3	0.09 (0.05)	0.09 (0.04)

$\text{var}(Y_i) = 1.02\mu_i$. For fitted means below about 0.2 this was comparable to the mean-variance relationship in Dalthorp (2004), whereas for larger fitted means, corresponding variances were up to 22% smaller than those in Dalthorp (2004). Note that underestimating the variances will tend to lead to overestimating correlations, so the model used is conservative, since we are concerned about adequately accounting for dependence.

6. SIMULATION STUDY

A simulation study was performed to assess the performance of the proposed ML estimator compared to a GEE estimator. Three sample sizes were simulated ($n = 144$, $n = 225$, and $n = 484$). Spatial locations were on a regular square grid with 1-unit spacing. For each sample size, four levels of spatial dependence (weak, low, moderate, and strong) and three sample sizes were simulated. Spatial dependence is described as *effective range*, the distance at which correlation drops to 0.05. The weakly correlated datasets have effective range $R = 1.2$, and datasets with low, moderate, and strong dependence have effective ranges $R = 3.1$, $R = 5.3$, and $R = 8.3$, respectively. Target means are taken to be constant $\exp(\beta)$ where $\beta = 1$, so all dependence in the data is due to spatial proximity, not spatial pattern of covariates. Because the GEE procedure depends on the lognormal-Poisson model which cannot model high correlations, it is expected that the ML estimator will outperform the GEE estimator for data with moderate or high spatial dependence. Table 2 gives the percentage of elements of the correlation matrix exceeding the lognormal-Poisson upper bound for each scenario. Samples with weak correlation fell within the lognormal-Poisson upper bound. A larger percentage of pairs exceeding the lognormal-Poisson upper bound is found when either dependence increases or sample size decreases.

$N = 500$ simulated datasets were generated for each scenario using the model given in Equation (5.1). ML estimates $\hat{\beta}_{ML}$ were obtained using the procedure described in Section 5. GEE estimates $\hat{\beta}_{GEE}$ were obtained using the algorithm of McShane, Albert, and Palmatier (1997). Both algorithms used the exponential correlation model from Equation (5.3), though $\boldsymbol{\Sigma}$ represents different quantities in each model. In the lognormal-Poisson model, $\boldsymbol{\Sigma}$ is the correlation matrix of the latent lognormal vector, whereas in the Gaussian copula model it is the copula correlation matrix.

Table 2. Percentage of pairs exceeding the latent process lognormal-Poisson upper bound for the twelve simulation scenarios.

Effective range	% pairs above L-P bound		
	$n = 144$	$n = 225$	$n = 484$
$R = 1.2$	0	0	0
$R = 3.1$	4.9	3.2	2.6
$R = 5.3$	19.6	15.2	7.4
$R = 8.3$	40.2	29.1	16.1

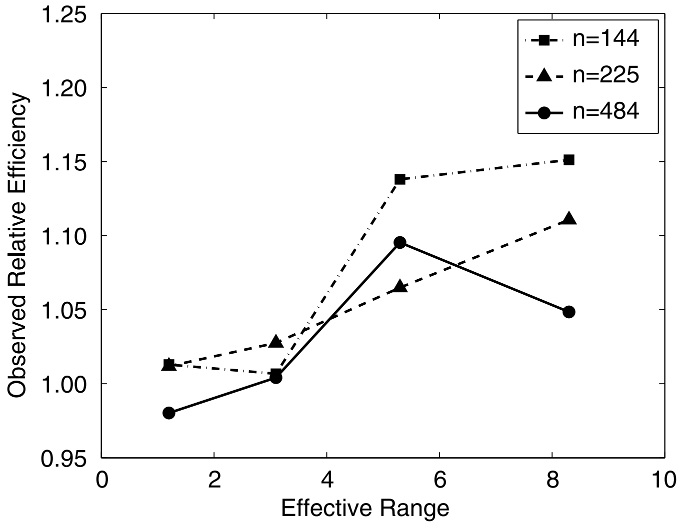

Figure 5. Observed relative efficiency of GEE estimator to MLE. Points plotted are the ratio of Monte Carlo sample variances (6.1) for each scenario. All but two of the ratios exceed one, so the ML estimator appears to be more efficient than the GEE estimator. When spatial dependence is low, the relative efficiency is approximately one.

Figure 5 plots the observed relative efficiency

$$\frac{(N-1)^{-1}\sum_{i=1}^{N}(\hat{\beta}_{GEE}-\overline{\hat{\beta}}_{GEE})^2}{(N-1)^{-1}\sum_{i=1}^{N}(\hat{\beta}_{L}-\overline{\hat{\beta}}_{ML})^2}, \tag{6.1}$$

versus the effective range. In all but 2 of the 12 simulation scenarios, the observed relative efficiency exceeds 1, suggesting that the ML estimator is more efficient than the GEE estimator. For weak spatial dependence the observed relative efficiencies are all within 0.02 of 1. As the effective range increases, the observed relative efficiency increases, though for $n = 484$ the observed relative efficiency is smaller for $R = 8.3$ than it is for $R = 5.3$.

Figure 6 gives nominal 95% confidence coverage for the two estimators under the two simulation scenarios. Both estimators perform best under weak dependence, and confidence coverage declines as dependence increases. This decline may be attributed to the fact that both estimators' standard errors are obtained from asymptotic variance estimators, and the effective sample size declines as dependence increases. ML confidence coverage exceeds GEE coverage in 11 of the 12 simulation scenarios. GEE coverage is particularly poor when spatial dependence is high. This is due to the inability of the latent process lognormal-Poisson model to account for high correlations among the data.

ML estimates of correlation parameters $\hat{\theta}_0$ and $\hat{\theta}_1$ were found to be correlated, especially for data with weak dependence. Intuitively, this is because near-independence can be modeled as in Equation (5.3) either by a large nugget parameter θ_0 or a large decay parameter θ_1. Observed correlations between $\hat{\theta}_0$ and $\hat{\theta}_1$ were between 0.83 and 0.94 for the three simulations with effective range 1.2 and dropped to between 0.25 and 0.51 for the three simulations with effective range 8.3. This correlation made estimation of these

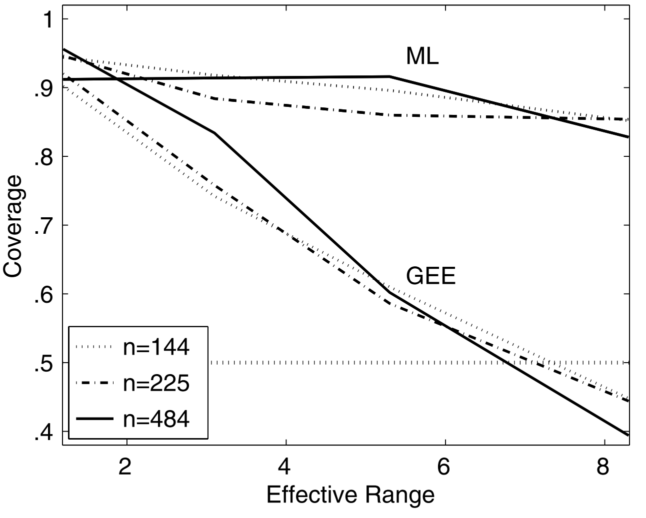

Figure 6. Comparison of nominal 95% confidence coverage for GEE and ML interval estimators. The dotted horizontal reference line is at coverage = 0.5. Coverage declines as dependence increases, particularly for GEE estimators.

parameters difficult. However, estimates of the θ_i are used here only to account for spatial dependence, characterized by $\theta_0 \exp(-h\theta_1)$.

It should be emphasized that $\theta_0 \exp(-h\theta_1)$ is not the correlation between counts separated by a distance h. Rather, this quantity is the "normal scoring," $\Sigma_{ij} = \text{corr}[\Phi^{-1}\{F_i(Y_i)\}, \Phi^{-1}\{F_j(Y_j)\}]$, where Y_i and Y_j are located h units apart. The relationship between $\text{corr}[\Phi^{-1}\{F_i(Y_i)\}, \Phi^{-1}\{F_j(Y_j)\}]$ and $\text{corr}(Y_i, Y_j)$ is not known precisely (unless Y_i and Y_j are jointly normal, see Kruskal 1958), but it is monotone and when $\Sigma_{ij} = 0$, Y_i, and Y_j are independent, and as mentioned in Section 4, when $\Sigma_{ij} = 1$, Y_i and Y_j are maximally correlated.

A comparison of $\hat{\theta}_0 \exp(-h\hat{\theta}_1)$ with the truth $\theta_0 \exp(-h\theta_1)$ reveals that normal scoring Σ_{ij} is underestimated. This underestimation has little effect on the fitted spatial dependence of the data with effective range $R = 1.2$, since these data are nearly independent. Comparison of plots of empirical variograms from the simulated data with analogous plots from data simulated with correlation parameters chosen by eye to match the observed mean of $\hat{\theta}_0 \exp(-h\hat{\theta}_1)$ showed slightly dampened spatial dependence. This may explain the drop in confidence coverage for these simulation scenarios. Improving estimates of $\mathbf{\Sigma}$ is a topic of future research.

Estimates of ϕ were highly variable, but estimates of overdispersion factor $\phi^{-1}(1+\phi)$ were stable. Observed mean squared error $\widehat{\text{MSE}} \equiv 500^{-1} \sum_{i=1}^{500} \{\hat{\phi}^{-1}(1+\hat{\phi}) - \phi^{-1}(1+\phi)\}^2$ increased with spatial dependence and decreased with sample size. For sample size $n = 484$ and effective range $R = 1.2$, $\widehat{\text{MSE}} = 0.024$, and for $n = 144$ and $R = 8.3$, $\widehat{\text{MSE}} = 0.118$.

7. CONCLUSIONS AND FUTURE RESEARCH

The spatial Gaussian copula model can be used to estimate the relationship between a set of covariates and a spatially dependent discrete response. The ML estimator presented here appears to be more efficient than the GEE estimator. Furthermore, in 11 out of 12 simulation scenarios, coverage of nominal 95% confidence intervals is closer to 95% for ML than for GEE. When the data are moderately or highly dependent, coverage of GEE intervals is well below 70%. GEE intervals' coverage is substantially lower for highly dependent data because the model cannot account for the high degree of correlation. Coverage of both ML and GEE interval estimators decline as spatial dependence increases. This decline is likely due to the reduction of effective sample size in samples with higher dependence. Further research is needed to obtain better ML estimates of spatial dependence when that dependence is strong.

Though the motivation for this research is to analyze spatially correlated discrete data, particularly count data, the model is general and can be applied to other correlated discrete data including space-time problems and longitudinal data. In the spatial setting, the Gaussian copula correlation matrix is modeled using a spatial correlation function. In other settings, other correlation models would be appropriate. Furthermore, without the continuous extension, the method applies to nonnormal continuous distributions as well.

APPENDIX

The first derivatives of log expected likelihood

$$l(\boldsymbol{\beta},\boldsymbol{\theta},\phi;\mathbf{y}) = \log\left(E\left\{\frac{\exp[-(1/2)\mathbf{z}^{*\prime}(\boldsymbol{\Sigma}^{-1}-\mathbf{I}_n)\mathbf{z}^*]\prod_{i=1}^n f_i^*(y_i^*)}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}\bigg|\mathbf{y}\right\}\right) \tag{A.1}$$

are given.

For simplicity, ignore the constant $(2\pi)^{n/2}$. Note that Equation (A.1) can be rewritten as

$$l(\boldsymbol{\beta},\boldsymbol{\theta},\phi;\mathbf{y}) = \log\left(E\left[\exp\left\{-\frac{1}{2}\mathbf{z}^{*\prime}(\boldsymbol{\Sigma}^{-1}-\mathbf{I}_n)\mathbf{z}^*\right\}\bigg|\mathbf{y}\right]\right) + \sum_{i=1}^n \log[f_i^*(y_i^*)] - \frac{1}{2}\log|\boldsymbol{\Sigma}|.$$

Let ξ be a parameter of the marginal distribution of Y_i, e.g., $\xi = \boldsymbol{\beta}_j$ or $\xi = \phi$ for the model in this article. Then

$$\frac{\partial l}{\partial \xi} = \frac{E[\exp\{-(1/2)\mathbf{z}^{*\prime}(\boldsymbol{\Sigma}^{-1}-\mathbf{I}_n)\mathbf{z}^*\}\{-(\partial \mathbf{z}^{*\prime}/\partial\xi)(\boldsymbol{\Sigma}^{-1}-\mathbf{I}_n)\mathbf{z}^*\}|\mathbf{y}]}{E[\exp\{-(1/2)\mathbf{z}^{*\prime}(\boldsymbol{\Sigma}^{-1}-\mathbf{I}_n)\mathbf{z}^*\}|\mathbf{y}]} + \sum_{i=1}^n \frac{\partial f_i^*(y_i^*)/\partial\xi}{f_i^*(y_i^*)},$$

where

$$\frac{\partial z_i^*}{\partial \xi} = \frac{2\pi}{\exp(-(1/2)z_i^2)}\cdot\frac{\partial}{\partial\xi}F_i^*(y_i^*).$$

For correlation parameter θ_i,

$$\frac{\partial l}{\partial \theta_i} = \frac{E[\exp\{-(1/2)\mathbf{z}^{*\prime}(\boldsymbol{\Sigma}^{-1} - \mathbf{I}_n)\mathbf{z}^*\}\{(1/2)\mathbf{z}^{*\prime}(\boldsymbol{\Sigma}^{-1}(\partial \boldsymbol{\Sigma}/\partial \theta_i)\boldsymbol{\Sigma}^{-1})\mathbf{z}^*\}|\mathbf{y}]}{E[\exp\{-(1/2)\mathbf{z}^{*\prime}(\boldsymbol{\Sigma}^{-1} - \mathbf{I}_n)\mathbf{z}^*\}|\mathbf{y}]} - \frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\right).$$

SUPPLEMENTAL MATERIALS

Functions for estimation: Contains main function “GrubEstimation.m” and negative log expected likelihood function “NegLogEL.m” used to estimate parameters in Section 5. The data analyzed are also included in a text file. (Grubs.zip)

ACKNOWLEDGMENTS

This research was partially funded by the U.S. Environmental Protection Agency grant CR-829095, the Science To Achieve Results (STAR) Program. It has not been subjected to the Agency’s review and, therefore, does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

This research was partially funded by Oregon Watershed Enhancement Board (OWEB) grant 208-922-6791. It has not been subjected to the OWEB’s review and, therefore, does not necessarily reflect the views of Board and no official endorsement should be inferred.

The author thanks Dan Dalthorp for providing the grub data, and Dave Birkes and the reviewers for providing valuable feedback.

[Received January 2008. Revised December 2008.]

REFERENCES

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.

Berger, J. O., de Oliveira, V., and Sansó, B. (2001), “Objective Bayesian Analysis of Spatially Correlated Data,” *Journal of the American Statistical Association*, 96, 1361–1374.

Cressie, N. A. C. (1993), *Statistics for Spatial Data* (revised ed.), New York: Wiley.

Dalthorp, D. (2004), “The Generalized Linear Model for Spatial Data: Assessing the Effects of Environmental Covariates on Population Density in the Field,” *Entomologia Experimentalis et Applicata*, 111, 117–131.

Dalthorp, D., Nyrop, J., and Villani, M. (2000), “Spatial Ecology of the Japanese Beetle, *Popillia japonica*,” *Entomologia Experimentalis et Applicata*, 96, 129–139.

Denuit, M., and Lambert, P. (2005), “Constraints on Concordance Measures in Bivariate Discrete Data,” *Journal of Multivariate Analysis*, 93, 40–57.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), “Model-Based Geostatistics,” *Applied Statistics*, 47, 299–350.

Fisher, N. (1997), “Copulas,” in *Encyclopedia of Statistical Sciences*, Vol. 1, New York: Wiley, pp. 159–163.

Gotway, C. A., and Stroup, W. W. (1997), “A Generalized Linear Model Approad to Spatial Data Analysis and Prediction,” *Journal of Agricultural, Biological, and Envirnomental Statistics*, 2, 157–178.

Hilbe, J. M. (2007), *Negative Binomial Regression*, New York: Cambridge University Press.

Hoeffding, W. (1940), “Masstabinvariante Korrelationstheorie,” *Schriften des Matematischen Instituts und des Instituts für Angewandte Matematik de Universität Berlin*, 5, 179–233.

Hoff, P. D. (2007), "Extending the Rank Likelihood for Semiparametric Copula Estimation," *Journal of Computational and Theoretical Nanoscience*, 1, 265–283.

Irvine, K. M., Gitelman, A. I., and Hoeting, J. A. (2007), "Spatial Designs and Properties of Spatial Correlation: Effects on Covariance Estimation," *Journal of Agricultural, Biological, and Environmental Statistics*, 12, 450–469.

Joe, H. (2001), *Multivariate Models and Dependence Concepts*, London: Chapman & Hall/CRC.

Kruskal, W. H. (1958), "Ordinal Measures of Association," *Journal of the American Statistical Association*, 53, 814–861.

Lee, Y., Nelder, J. A., and Pawitan, Y. (2006), *Generalized Linear Models With Random Effects, Unified Analysis via H-Likelihood*, Boca Raton, FL: Chapman & Hall/CRC.

Liang, K.-Y., and Zeger, S. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Lin, P.-S., and Clayton, M. K. (2005), "Analysis of Binary Spatial Data by Quasi-Likelihood Estimating Equations," *The Annals of Statistics*, 33, 542–555.

Madsen, L., and Dalthorp, D. (2007), "Simulating Correlated Count Data," *Environmental and Ecological Statistics*, 14, 129–148.

Mardia, K. V., and Marshall, R. J. (1984), "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Statistics," *Biometrika*, 71, 135–146.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), Boca Raton, FL: Chapman & Hall/CRC.

McShane, L. M., Albert, P. S., and Palmatier, M. A. (1997), "A Latent Process Regression Model for Spatially Correlated Count Data," *Biometrics*, 53, 698–706.

Nelsen, R. B. (2006), *An Introduction to Copulas* (2nd ed.), New York: Springer.

Pitt, M., Chan, D., and Kohn, R. (2006), "Efficient Bayesian Inference for Gaussian Copula Regression Models," *Biometrika*, 93, 537–554.

Prentice, R. L. (1988), "Correlated Binary Regression With Covariates Specific to Each Binary Observation," *Biometrics*, 44, 1033–1048.

Schabenberger, O., and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Boca Raton, FL: Chapman & Hall/CRC.

Solomon, D. L. (1983), "The Spatial Distribution of Butterfly Eggs," in *Life Science Models*, Vol. 4, eds. H. Roberts and M. Thompson, New York: Springer-Verlag, pp. 350–366.

Song, P. X.-K. (2000), "Multivariate Dispersion Models Generated From Gaussian Copula," *Scandanavian Journal of Statistics*, 27, 305–320.

——— (2007), *Correlated Data Analysis*, New York: Springer.

Song, P. X.-K., Li, M., and Yuan, Y. (2009), "Joint Regression Analysis of Correlated Data Using Gaussian Copulas," *Biometrics*, 65, 60–68.

Taylor, L. R. (1961), "Aggregation, Variance and the Mean," *Nature*, 189, 732–735.

Venables, W., and Ripley, B. (2002), *Modern Applied Statistics With S* (4th ed.), New York: Springer.

Yasui, Y., and Lele, S. (1997), "A Regression Method for Spatial Disease Rates: An Estimating Function Approach," *Journal of the American Statistical Association*, 92, 21–32.

Zeger, S., and Liang, K.-Y. (1986), "Longitudinal Data Analysis for Discrete and Continuous Outcomes," *Biometrics*, 42, 121–130.

Zhang, H. (2004), "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics," *Journal of the American Statistical Association*, 99, 250–261.