Analyzing Spatially Distributed Binary Data Using Independent-Block Estimating Equations

Samuel D. Oman,[*] **Victoria Landsman,**[**]

Department of Statistics, Hebrew University of Jerusalem, Mount Scopus 91905 Jerusalem, Israel
[*]*email:* oman@mscc.huji.ac.il
[**]*email:* msvika@mscc.huji.ac.il

Yohay Carmel,

Faculty of Agricultural Engineering, Technion Israel Institute of Technology 32000 Haifa, Israel
email: yohay@techunix.technion.ac.il

and

Ronen Kadmon

Department of Evolution, Systematics and Ecology, Institute of Life Sciences,
Hebrew University of Jerusalem, Givat Ram 91904 Jerusalem, Israel
email: kadmon@vms.huji.ac.il

Summary. We estimate the relation between binary responses and corresponding covariate vectors, both observed over a large spatial lattice. We assume a hierarchical generalized linear model with probit link function, partition the lattice into blocks, and adopt the working assumption of independence between the blocks to obtain an easily solved estimating equation. Standard errors are obtained using the "sandwich" estimator together with window subsampling (Sherman, 1996, *Journal of the Royal Statistical Society, Series B* **58,** 509–523). We apply this to a large data set describing long-term vegetation growth, together with two other approximate-likelihood approaches: pairwise composite likelihood (CL) and estimation under a working assumption of independence. The independence and CL methods give similar point estimates and standard errors, while the independent-block approach gives considerably smaller standard errors, as well as more easily interpretable point estimates. We present numerical evidence suggesting this increased efficiency may hold more generally.

Key words: Binary variables; Composite likelihood; Generalized estimating equations; Probit regression; Spatial dependence.

1. Introduction

Many applications require modeling binary variables Y_i, observed at points in a spatial lattice, in terms of corresponding vectors of explanatory variables; often the points are regularly spaced and represent small areas or volumes. As examples, Albert and McShane (1995) model the presence of lesions at different locations in the brain; Heagerty and Lele (1998) examine defoliation in different grid cells within a geographical area; Weir and Pettitt (2000) study the presence of the Finnish common toad in different squares of the Finnish national grid; and Pettitt, Weir, and Hart (2002) analyze forest biodiversity data. Our application relates the predominance of trees in grid cells within a region of natural forest to environmental and anthropogenic variables measured at the cells.

One may model spatial dependencies among the Y_i using a hierarchical generalized linear model (Clayton and Kaldor,

1987; Breslow and Clayton, 1993; Waller et al., 1997; Diggle, Tawn, and Moyeed 1998), in which the Y_i are conditionally independent, given a latent Gaussian field defined on the lattice. Dependence among the Y_i then arises from dependence among the components of the field.

Bayesian estimation may be carried out using Markov chain Monte Carlo to approximate the posterior distributions (Besag, York, and Mollie 1991; Waller et al., 1997; Diggle et al., 1998; Weir and Pettitt, 2000; Christensen and Waagepetersen, 2002; Pettitt et al., 2002; Christensen, Roberts, and Skold 2006). For a non-Bayesian analysis, maximum likelihood estimation is computationally prohibitive for moderately large lattices; hence, a number of approximate likelihood methods have been proposed. In particular, Heagerty and Lele (1998) used a composite likelihood (CL) approach (Lindsay, 1988) to approximate the log-likelihood

function by a sum of pairwise log-likelihoods. Standard errors were estimated using the "sandwich estimator," together with window subsampling (Sherman, 1996) to estimate the covariance matrix of the estimating function. As a computationally simpler (but less efficient) alternative, Heagerty and Lumley (2000) proposed modeling the marginal probabilities $P(Y_i = 1)$ assuming independence, while estimating standard errors as in Heagerty and Lele (1998); this is similar to the independence estimating equation (IEE) approach of Zeger and Liang (1986) and Liang and Zeger (1986).

The approach we present here is in the spirit of the generalized estimating equations (GEE) of Zeger and Liang (1986) and Liang and Zeger (1986), and was in part motivated by computational difficulties in applying Heagerty and Lele's (1998) CL method to our set of vegetation data. Assuming a probit link function, we construct an estimating equation with a "working covariance matrix" based on dividing the lattice into disjoint blocks and assuming independence between the observations in different blocks; hence, the matrix is block-diagonal. We choose the blocks sufficiently large to reflect the underlying covariance structure, while still allowing inversion of the corresponding covariance matrices. Within each block, we use Pearson's (1901) approximation to the true covariances of $\mathbf{Y}$ when an explicit covariance function is assumed for the latent Gaussian field. The approximation involves the standard normal probability density function and cumulative distributive function, and does not require computing bivariate normal integrals. The covariance function of the underlying field is modeled parametrically and estimated using a separate estimating equation. The resulting independent-block estimating equation (IBEE) for the regression vector is easy to program and solve; the covariance of the resulting estimator is estimated as in Heagerty and Lele (1998) using the sandwich estimator and window subsampling.

In the next section our model and method are defined, while Section 3 states the asymptotic properties of the resulting estimator. In Section 4, we compute the IBEE estimate for a moderately large set of 6000 points, together with the IEE and CL estimates. The IEE and CL methods give essentially the same point estimates and standard errors, while the IBEE estimated standard errors are considerably smaller. The IBEE point estimates differ as well, and lead to qualitatively different (and more easily interpretable) conclusions than obtained using IEE or CL. In Section 5, we present numerical evidence suggesting that the IBEE estimator may be more efficient in other cases as well. Section 6 concludes with a brief discussion. Web Appendix A gives the proofs of the results of Section 3, while Web Appendix B contains the software code used for the computations in Section 4.

2. Statistical Model and Estimation Technique

Let $S = \{s_1, \ldots, s_N\}$ denote the lattice, with $\mathbf{x}_i$ the vector of covariate values observed at s_i, and denote the distance between two points s_i and s_j by $\|s_i - s_j\|$. As in Heagerty and Lele (1998), we assume a hierarchical generalized linear model having a latent zero-mean Gaussian field $\mathbf{Z}^* = (Z_i^*)$ defined on the lattice, with var $(Z_i^*) \equiv \tau^2$ and a continuous isotropic covariance function cov $(Z_i^*, Z_j^*) = c(\|s_i - s_j\|)$.

Conditional on $\mathbf{Z}^*$, the Y_i are independent, with

$$g[P(Y_i = 1|\mathbf{Z}^*)] = \mathbf{x}_i^T \boldsymbol{\gamma} + Z_i^*, \tag{1}$$

where g is the probit link function and $\boldsymbol{\gamma}$ gives the conditional effects of the covariates. Because equation (1) is equivalent to $Y_i = 1_{\{W_i - Z_i^* \leq \mathbf{x}_i^T \boldsymbol{\gamma}\}}$, where the W_i are N(0, 1) variables independent of one another and of $\mathbf{Z}^*$, we obtain the unconditional probabilities

$$\begin{aligned} p_i &\equiv P(Y_i = 1) \\ &= P(Z_i \leq \eta_i), \end{aligned} \tag{2}$$

where the Z_i are correlated latent $N(0, 1)$ variables and $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ for a regression vector $\boldsymbol{\beta} = (1 + \tau^2)^{-1/2} \boldsymbol{\gamma}$, which now gives global effects. Similarly, we obtain

$$\text{cov}(Y_i, Y_j) = P(Z_i \leq \eta_i, Z_j \leq \eta_j) - p_i p_j. \tag{3}$$

The latent field $\mathbf{Z}$ has an isotropic correlation function

$$\rho(\|s_i - s_j\|) = \begin{cases} 1, & \text{if} s_i = s_j \\ c(\|s_i - s_j\|)/(1 + \tau^2), & \text{if } s_i \neq s_j. \end{cases} \tag{4}$$

Because $\lim_{d \to 0} \rho(d) = \tau^2/(1 + \tau^2)$, there is a "nugget effect" (Matheron, 1962) of $1/(1 + \tau^2)$.

Forming the $N \times p$ matrix $\mathbf{X}$ of covariate values and arranging the Y_i and p_i into corresponding N-vectors $\mathbf{Y}$ and $\mathbf{p} = \mathbf{p}(\boldsymbol{\beta})$, the optimal moment-based estimating equation for $\boldsymbol{\beta}$ is

$$\mathbf{D}^T \text{cov}^{-1}(\mathbf{Y})\{\mathbf{Y} - \mathbf{p}(\boldsymbol{\beta})\} = \mathbf{0}, \tag{5}$$

where $\mathbf{D} = \partial \mathbf{p}/\partial \boldsymbol{\beta}$. Equivalently, denote $\Phi_i = \Phi(\eta_i)$ and $\Phi_i = \phi(\eta_i)$, where Φ and ϕ are the standard normal Cumulative distributive function and probability density function, let

$$f_i = \phi_i/\{\Phi_i(1 - \Phi_i)\}^{1/2} \tag{6}$$

and define the $N \times N$ matrices

$$\mathbf{H} = \mathbf{H}(\boldsymbol{\beta}) = \text{diag}(\phi_i), \qquad \mathbf{F} = \mathbf{F}(\boldsymbol{\beta}) = \text{diag}(f_i). \tag{7}$$

Then, as $\mathbf{D} = \mathbf{HX}$ and var $(Y_i) = \Phi_i(1 - \Phi_i)$, we may write equation (5) in terms of correlations as

$$\mathbf{X}^T \mathbf{F} \text{corr}^{-1}(\mathbf{Y}) \mathbf{FH}^{-1}\{\mathbf{Y} - \mathbf{p}(\boldsymbol{\beta})\} = \mathbf{0}. \tag{8}$$

Because inverting corr $(\mathbf{Y})$ is infeasible for large N, we instead construct a "working correlation" matrix by partitioning the lattice into subsets or blocks $B_1, \ldots, B_m$ of contiguous sites and assuming independence between Y_j and Y_k if s_j and s_k are in different blocks. For sites s_j and s_k within the same block, we use Pearson's (1901) approximation to the covariances implied by equations (3) and (4),

$$\text{cov}(Y_j, Y_k) \approx \phi_j \phi_k \arcsin\{\rho(\|s_j - s_k\|)\}, \tag{9}$$

and replace ρ in equation (9) by a parametric working correlation function $r_{\boldsymbol{\alpha}}$, assumed to be isotropic and continuous in $\boldsymbol{\alpha}$. For example, graphical analysis of the data in Section 4 suggested the choice

$$r_{\boldsymbol{\alpha}}(d) = \begin{cases} 1, & \text{if d} = 0 \\ \alpha_1 \alpha_2^d, & \text{if d} > 0. \end{cases} \tag{10}$$

This gives working correlations

$$a_{jk} = \begin{cases} 1, & j = k \\ f_j f_k \arcsin\{r_\alpha(\|s_j - s_k\|)\}, & j \neq k \end{cases} \quad (11)$$

when s_j and s_k are in the same block. Assuming the observations in the vector $\mathbf{Y}$ have been grouped according to the blocks $B_1, \ldots, B_m$, let

$$\mathbf{A} = \mathbf{A}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \text{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_m) \quad (12)$$

denote the resulting block-diagonal working correlation matrix, with each $\mathbf{A}_i$ having entries as in equation (11). Replacing corr($\mathbf{Y}$) in equation (8) by our working correlation matrix and normalizing by N then gives the estimating function

$$\mathbf{U}_N(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{N}\mathbf{X}^T\mathbf{F}(\boldsymbol{\beta})\mathbf{A}^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha})\mathbf{F}(\boldsymbol{\beta})\mathbf{H}^{-1}(\boldsymbol{\beta})\{\mathbf{Y} - \mathbf{p}(\boldsymbol{\beta})\}. \quad (13)$$

As in Liang and Zeger (1986), we may solve $\mathbf{U}_N(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{0}$ by a modified Fisher scoring algorithm in which at each iteration we differentiate equation (13) only with respect to $\boldsymbol{\beta}$ in $\mathbf{p}(\boldsymbol{\beta})$. Because $\mathbf{A}$ is block diagonal, the update at the jth step is of the form

$$\boldsymbol{\beta}_{j+1} = \boldsymbol{\beta}_j + \{\sum_{i=1}^{m}(\mathbf{X}_i^T\mathbf{F}_i\mathbf{A}_i^{-1}\mathbf{F}_i\mathbf{X}_i)\}^{-1}$$

$$\sum_{i=1}^{m}\mathbf{X}_i^T\mathbf{F}_i\mathbf{A}_i^{-1}\mathbf{F}_i\mathbf{H}_i^{-1}\{\mathbf{Y}_i - \mathbf{p}_i(\boldsymbol{\beta}_j)\}; \quad (14)$$

here $\mathbf{H}_i$, $\mathbf{F}_i$, $\mathbf{X}_i$, and $\mathbf{p}_i$ are submatrices corresponding to block i, all evaluated at the current $\boldsymbol{\beta}_j$, while $\mathbf{A}_i$ is evaluated at $\boldsymbol{\beta}_j$ and $\boldsymbol{\alpha}$. Equation (14) is simple to program, and computation is quick for appropriate block sizes.

In practice we replace $\boldsymbol{\alpha}$ in equation (13) by $\hat{\boldsymbol{\alpha}}_N$, a consistent estimate of $\boldsymbol{\alpha}$. (More precisely, as $\boldsymbol{\alpha}$ is a "working parameter," we require that $\{\hat{\boldsymbol{\alpha}}_N\}$ be statistics which converge in probability to a constant $\boldsymbol{\alpha}_0$.) To compute $\hat{\boldsymbol{\alpha}}_N$ we form the squared differences $W_{jk} \equiv (Y_j - Y_k)^2$; from equation (9) we have

$$E(W_{jk}) \approx \Phi(\eta_j) + \Phi(\eta_k) - 2\phi(\eta_j)\phi(\eta_k)$$

$$\arcsin\{r_\alpha(\|s_j - s_k\|)\} - 2\Phi(\eta_j)\Phi(\eta_k). \quad (15)$$

Replacing the η_i by $\hat{\eta}_i = \mathbf{x}_i^T\hat{\boldsymbol{\beta}}$ for an appropriate estimate $\hat{\boldsymbol{\beta}}$, we may use equation (15) to construct an estimating equation for $\boldsymbol{\alpha}$.

3. Asymptotic Properties

Let $\hat{\boldsymbol{\beta}}_N$ denote a solution to $\mathbf{U}_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}_N) = \mathbf{0}$. Consistency of $\hat{\boldsymbol{\alpha}}_N$, as discussed preceding equation (15), allows us to show that under appropriate conditions, similar to those of Heagerty and Lele (1998) and Heagerty and Lumley (2000), $\hat{\boldsymbol{\beta}}_N$ is consistent and asymptotically normal as $N \to \infty$. More precisely, if $\boldsymbol{\beta}_0$ denotes the true value of $\boldsymbol{\beta}$, if $\hat{\boldsymbol{\alpha}}_N \to \boldsymbol{\alpha}_0$ in probability and if we define $\mathbf{V}_N = \text{cov}_{\beta_0}\{\mathbf{U}_N(\boldsymbol{\beta}_0, \boldsymbol{\alpha})\}$, then $N^{1/2}\mathbf{V}_N^{-1/2}\mathbf{D}_N(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0)$ converges in distribution to a $N(\mathbf{0}, \mathbf{I})$ variable, where $\mathbf{V}_N$ is evaluated at $\boldsymbol{\alpha}_0$ and

$$\mathbf{D}_N = N^{-1}\mathbf{X}^T\mathbf{F}(\boldsymbol{\beta}_0)\mathbf{A}^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\mathbf{F}(\boldsymbol{\beta}_0)\mathbf{X}.$$

The proof of this result, together with a precise statement of the required conditions, is in Web Appendix A. The proof uses asymptotic results of Crowder (1986), Doukhan (1994), and Guyon (1995) for random fields.

We then estimate the covariance matrix of $N^{\frac{1}{2}}\boldsymbol{\beta}_N$ using the "sandwich" estimator

$$\hat{\text{cov}}(\sqrt{N}\hat{\boldsymbol{\beta}}_N) = [\mathbf{D}\mathbf{U}(\hat{\boldsymbol{\beta}}_N, \hat{\boldsymbol{\alpha}}_N)]^{-1}$$

$$[\hat{\text{cov}}\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)][\mathbf{D}\mathbf{U}(\hat{\boldsymbol{\beta}}_N, \hat{\boldsymbol{\alpha}}_N)]^{-T}. \quad (16)$$

As no replicates are available to estimate the inner part of the sandwich, we follow Heagerty and Lele (1998) and Heagerty and Lumley (2000), in using the window subsampling method of Sherman (1996). This approach estimates cov ($\mathbf{U}$) using a rescaled average of $\mathbf{U}_k(\hat{\beta}, \hat{\alpha})\mathbf{U}_k(\hat{\beta}, \hat{\alpha})^T$, where the estimating functions $\mathbf{U}_k$ are computed over subsampled windows S_k of the research region.

4. An Application

To illustrate, we now analyze factors affecting long-term changes in tree cover in a 1125 m $\times$ 1200 m region of natural forest in the Galilee mountains of northern Israel, using binary grid maps available for 1964 and 1992. Each 15 m $\times$ 15 m grid cell in the maps was classified as "tree" or "nontree" based on image processing of aerial photographs (Carmel and Kadmon, 1999). We define $Y_i = 1$ if site s_i was a tree cell in the 1992 map. The percentage of cells in which trees were the dominant growth form increased from 1% in 1964, to 70% in 1992 (Figures 1a and 1b). Ecological considerations suggested that these changes were influenced by both
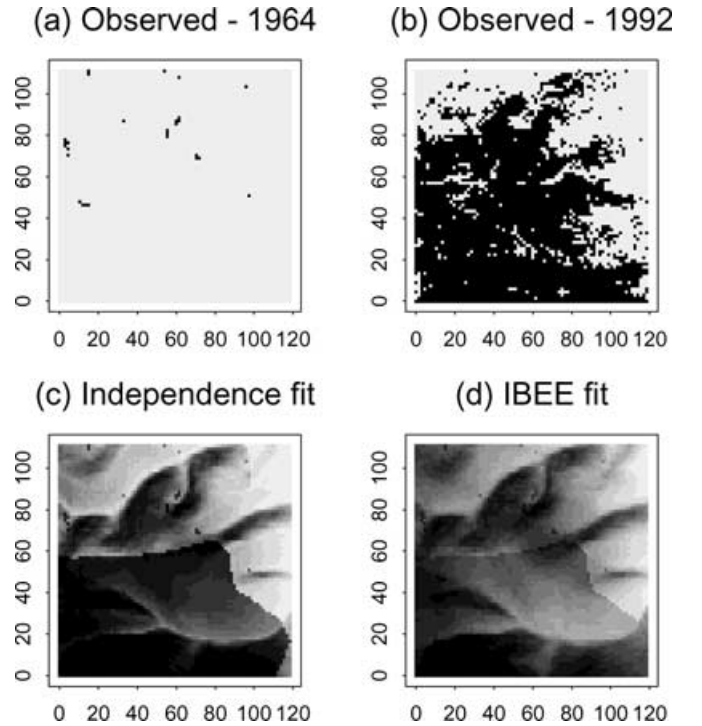
Figure 1. Observed values of tree dominance (black) and nondominance (white), together with fitted values using the independence and independent-block estimators. Scale is in units of 10 m; each pixel is 15 m $\times$ 15 m.

initial vegetation conditions and various cell-specific environmental and anthropogenic factors present during the 28-year period, which we model using the following variables: *tree.64* is an indicator variable for tree dominance in 1964; *alt (100 m)* is elevation above sea level; *str.dist (km)* is the distance to the nearest stream bed; *rd.dist (km)* is the distance to the nearest road; *slope* is the slope (on a scale from 0 to 1); *ns.aspct* gives the north–south component of the aspect ($N = 0/360$, $S = 180/360$); and *ew.aspct* gives the east–west component ($E = 90/360$, $W = 270/360$). As aspect and slope together affect the exposure of the cell to both precipitation and solar radiation, we also include the interaction terms *ns.slope* and *ew.slope*. The entire study region was fenced into separate areas for grazing (including one area in which grazing was not permitted), so we also include indicator variables for the type of grazing (goat or cattle) and the intensity (sporadic, moderate, or heavy; see Carmel and Kadmon [1999] for precise definitions).

Using model (2), we first computed $\hat{\boldsymbol{\beta}}_{\text{ind}}$, the estimate of $\boldsymbol{\beta}$ under independence, by using standard probit regression (all computations were done in S-plus). We obtained standard errors using subsampling of all overlapping windows of size 20×22 (see the Discussion). A directional variogram (not shown) of the Pearson residuals from the probit fit showed no evidence of anisotropy. To explore the form of the covariances on the latent scale, we divided the unit interval into eight equal-sized subintervals. For two such subintervals I and J, and for a given distance d, let $\hat{\eta}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{\text{ind}}$ and

$$c(I, J, d) = \text{sample covariance of } \{(Y_j, Y_k) : \Phi(\hat{\eta}_j) \in I, \Phi(\hat{\eta}_k) \in J, \|s_j - s_k\| = d\}.$$

If ϕ_L denotes the sample mean of $\{\phi(\hat{\eta}_l) : \Phi(\hat{\eta}_l) \in L\}$ and we define the empirical correlation on the latent scale to be $\sin\{c(I, J, d)/(\phi_I \phi_J)\}$, we see from equation (9) that this estimates $\rho(d)$. A lowess fit to the graph of these empirical correlations versus d (Figure 2) suggested the exponential working correlation function (10); using least squares we obtained

$$r_{\hat{\alpha}}(d) = 0.66(0.77)^d. \qquad (17)$$

For convenience of interpretation we shall write $\boldsymbol{\alpha}$ as (σ^2, ρ).

For the IBEE estimate $\hat{\boldsymbol{\beta}}_{\text{IBEE}}$ we first partitioned the 80×75 lattice into blocks B_i of size 16×15, the largest easily programmable block size for which inversion of the corresponding covariance matrices was feasible. As the estimated maximal correlation (using equation (17)) between a point in the center of any B_i and points not in B_i is at most $0.66 \, (0.77)^8 = 0.08$, this block size should capture much of the covariance structure. To estimate the working covariance parameters we transformed them to $\theta_i = \text{logit}(\alpha_i)$ and used (15) to construct an estimating equation analogous to equation (5). We formed a vector $\mathbf{w}$ comprising all W_{jk} corresponding to points s_j and s_k separated by at most $d_{\max} = 5$ units in both the horizontal and vertical directions (thereby including all pairs with preliminary estimated correlations as low as $0.66 \, (0.77)^5 = 0.18$), and used diag $\{\text{var}(w_i)\}$ in place of cov $(\mathbf{w})$ as a working covariance matrix. As the covariance parameters in equation (10) are not identified if either is zero, we used a small ridge parameter in the least-squares updates in order to keep the θ_i bounded. We then computed $\hat{\boldsymbol{\beta}}_{\text{IBEE}}$ and a final estimate of (σ^2, ρ) by alternating between the one-step iteration equation (14) and a one-step iteration for equation (15), with $\hat{\boldsymbol{\beta}}_{\text{ind}}$ and $(0.66, 0.77)$ as starting values. After eight iterations (taking
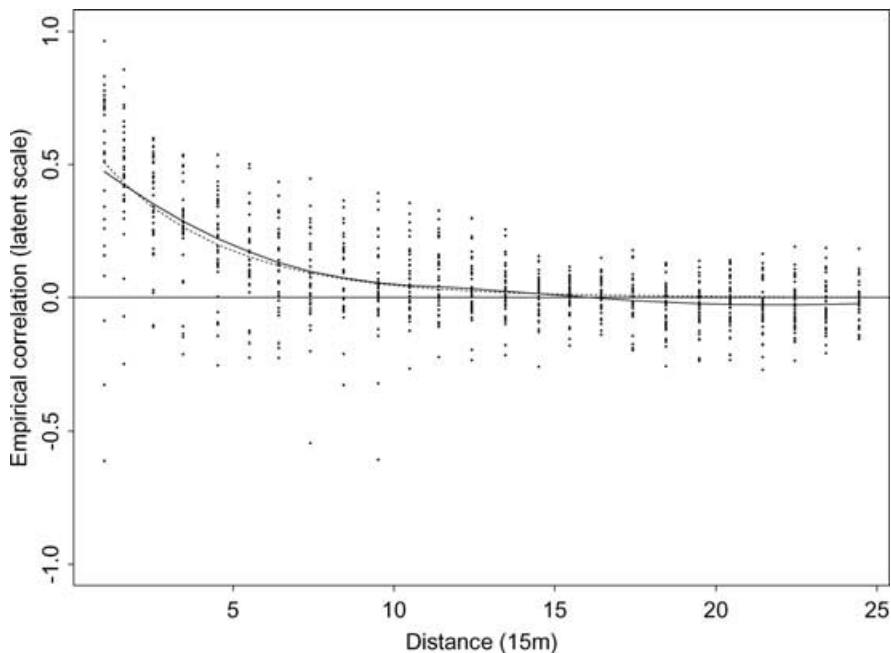
Figure 2. Empirical correlogram (on latent scale), together with fitted curves. Solid line is lowess fit; dashed line is least-squares fit cor $(d) = 0.66 \, (0.77)^d$ for $d > 0$.

Table 1
Results using independence, CL and IBEE

Variable	Coefficient estimate			Standard error			p-value		
	Ind	Cl	IBEE	Ind	Cl	IBEE	Ind	Cl	IBEE
const	−18.06	−18.38	−12.63	4.88	4.04	3.12	0.00	0.00	0.00
tree.64	1.16	1.35	0.55	0.27	0.27	0.27	0.00	0.00	0.05
alt	2.12	2.20	1.65	0.53	0.38	0.40	0.00	0.00	0.00
str.dist	−2.44	−3.16	−3.18	1.76	1.60	1.62	0.16	0.05	0.05
rd.dist	0.96	0.85	0.48	0.60	0.49	0.44	0.11	0.09	0.27
g.spor	−0.57	−0.61	−0.06	0.27	0.29	0.43	0.04	0.04	0.89
g.mod	−1.00	−0.96	−0.77	0.32	0.33	0.24	0.00	0.00	0.00
c.spor	0.53	0.54	−0.23	0.27	0.28	0.19	0.05	0.05	0.23
c.mod	−0.94	−0.92	−0.61	0.28	0.25	0.25	0.00	0.00	0.02
c.heavy	−1.66	−1.70	−0.79	0.44	0.39	0.23	0.00	0.00	0.00
ns.aspct	0.48	0.20	0.00	1.05	1.05	0.61	0.65	0.85	1.00
ew.aspct	0.98	0.99	−0.71	1.16	1.08	0.33	0.40	0.36	0.03
slope	4.27	2.91	0.94	4.17	4.16	1.85	0.31	0.48	0.61
ns.slope	−19.10	−17.65	−14.87	6.33	6.65	4.82	0.00	0.00	0.00
ew.slope	11.87	12.67	11.72	8.02	7.89	2.90	0.14	0.11	0.00

several minutes on a Sun Solaris Unix server, OS 5.9), we obtained

$$\hat{\sigma}^2 = 0.85, \hat{\rho} = 0.94, \qquad (18)$$

and $\hat{\boldsymbol{\beta}}_{\text{IBEE}}$. Standard errors were obtained as for $\hat{\boldsymbol{\beta}}_{\text{ind}}$, by subsampling windows of size 29×31.

Table 1 contains the IEE and IBEE results, together with estimates previously obtained by V. Landsman (2000) using CL. These were based on the sum of pairwise log-likelihoods U_{jk} for all pairs j, k up to a distance of 13 apart. As this involved on the order of 10^6 summands, each involving a bivariate normal integral, computations became feasible only after writing a number of C-plus routines which were called from within S-plus; the iterations then required several hours. The CL estimates and standard errors for σ^2 and ρ were 0.70(0.59) and 0.81(0.07).

In Table 1 we see that the CL and IEE point estimates and standard errors are quite similar. The IBEE standard errors, on the other hand, tend to be smaller: in fact, the ratio of IBEE to IEE standard errors, averaged over the coefficients, is 0.73 (as opposed to 0.95 for the corresponding CL to IEE ratio).

We focus therefore on comparing the IBEE and IEE results. Both methods give coefficient estimates with expected signs for *tree.64*, *rd.dist*, *str.dist*, and *alt* (we expect greater tree cover for sites at higher elevations because they generally get more precipitation, and are also more remote and thus less prone to human disturbance).

However, the p-values of the two methods in Table 1 lead to different conclusions regarding the effects of road construction and sporadic grazing. The IEE estimate finds road distance marginally significant, while it is not so using the IBEE results. Also, sporadic grazing by goats or cattle has no significant effect on tree cover, according to the IBEE estimate; whereas both types of grazing do have an effect, according to the IEE estimate. Moreover, the IEE results lead to the conclusion, difficult to interpret ecologically, that sporadic graz-

ing by cattle (as opposed to no grazing) actually *increases* tree cover.

Regarding slope and aspect, let p^*(aspect, slope) denote the probability that $Y = 1$ for a site with given slope and aspect, when the covariates not depending on slope or aspect are set to their mean values over the region. Figure 3 graphs the estimated p^* versus slope (the range of slopes in the research area was 0.0 to 0.35), together with (nonsimultaneous) 95% confidence limits, for the IEE and IBEE estimates and for northern, southern, western, and eastern aspects; observe that the IBEE envelopes are tighter than for IEE. To understand these graphs, we note that prevailing winds during rainfall in the research area come from the west and southwest (Arazi et al., 1997; Sharon and Arazi, 1997). Thus slopes with a western aspect tend to receive more rainfall than eastern slopes, leading to greater tree cover; moreover, as aspect together with slope affect the exposure to the slanting precipitation, this difference should increase as the slope increases. This is reflected in the curves in both Figures 3c and 3d. Both estimates also indicate that northern aspects are more favorable than southern to tree cover (Figures 3a and 3b). A likely explanation for this is that southern-facing slopes generally receive more sunlight, thereby leading to higher rates of evapotranspiration.

As a measure of overall fit, the estimates $\hat{p}_i$ using both methods are plotted in Figures 1c and 1d. The IBEE plot is much more graduated, because by accounting for correlations at the estimation step the estimator gives less weight to spatially contiguous groupings of identical Y values than does the IEE estimator; consequently their predictions are "shrunk" away from 0 and 1. A plot (not shown) of CL predictions was virtually identical to Figure 1c.

An appropriate measure of overall fit is $(\mathbf{Y} - \hat{\mathbf{p}})^T$ $\text{cov}^{-1}(\mathbf{Y})(\mathbf{Y} - \hat{\mathbf{p}})/(N - p)$, but computation is infeasible. Replacing cov $(\mathbf{Y})$ by $\text{diag}(\hat{p}_i/\{1 - \hat{p}_i\})$ gives the normalized Pearson's χ^2 (equal to 4.47, 6.87, and 0.87, respectively, for the IEE, CL, and IBEE estimates), but this is inappropriate as it ignores the spatial correlation in $\mathbf{Y}$. Choosing a rectangular
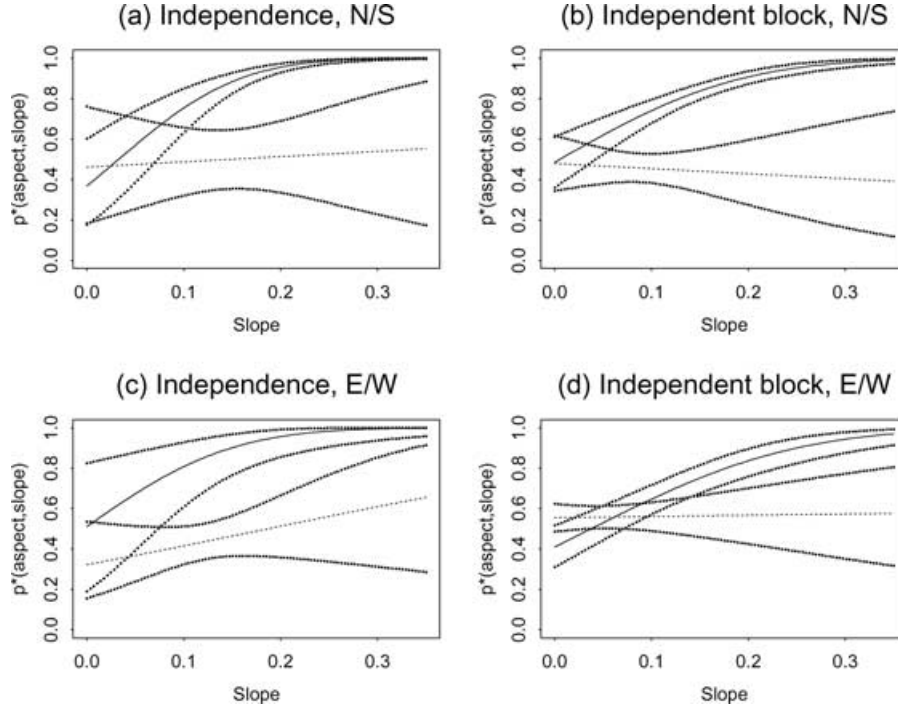
Figure 3. Estimated probability of tree dominance as a function of slope and aspect, when remaining covariates are set to their mean values, using independence and independent-block methods. Northern (solid line) and southern (dashed line) aspects are shown in (a) and (b); western (solid line) and eastern (dashed line) are shown in (c) and (d). Dotted lines give (nonsimultaneous) 95% confidence limits.

subset of 56 lattice points s_i that are 10 units apart in the horizontal and vertical directions (so that, using equations (9) and (18), the correlation among the corresponding Y_i should be at most 0.33) and computing the corresponding normalized Pearson's χ^2 gave respective values of 1.51, 1.73, and 1.18, indicating a better fit using independent blocks.

5. Efficiency

In our application the IBEE estimator appeared to be more efficient than both IEE and CL; we now present numerical results suggesting this may hold in a more general context as well. Assume that β is a scalar and cov ($\mathbf{Y}$) is known. If $\hat{\beta}$ is defined through the estimating equation

$$\mathbf{a}^T\{\mathbf{Y} - \mathbf{p}(\beta)\} = 0, \tag{19}$$

for $\mathbf{a} \in \Re^N$, then as with (16) a Taylor expansion gives its asymptotic variance as $\mathbf{a}^T$ cov ($\mathbf{Y}$) $\mathbf{a}/(\mathbf{a}^T \mathbf{Hx})^2$ with $\mathbf{H}$ defined in equation (7). Defining $\mathbf{c} = (c_i) = ([\Phi_i \{1 - \Phi_i\}]^{1/2} a_i)$ and $\mathbf{u} = (\phi_i x_i/[\Phi_i \{1 - \Phi_i\}]^{1/2})$, we may rewrite the variance as

$$\frac{\mathbf{c}^T\mathbf{Rc}}{(\mathbf{c}^T\mathbf{u})^2}, \tag{20}$$

where $\mathbf{R} =$ corr ($\mathbf{Y}$). A standard eigenvector argument shows that equation (20) achieves its minimum value, $\text{var}_{\text{opt}} = (\mathbf{u}^T \mathbf{R}^{-1} \mathbf{u})^{-1}$, when $\mathbf{c} = \mathbf{R}^{-1}\mathbf{u}$, corresponding to $\mathbf{a} = \text{cov}^{-1} (\mathbf{Y}) (\phi_i x_i)$. For the IEE estimator $\hat{\beta}_{\text{ind}}, \mathbf{a} = \text{diag}^{-1}(\Phi_i\{1-\Phi_i\})(\phi_i x_i)$, giving $\text{var}_{\text{ind}} = \mathbf{u}^T\mathbf{Ru}(\mathbf{u}^T \mathbf{u})^{-2}$; while the IBEE estimator that takes the subvectors $\mathbf{Y}_{(j)}$ of $\mathbf{Y}$ to be independent has $\text{var}_{\text{IBEE}} = \mathbf{u}^T\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{u}(\mathbf{u}^T\hat{\mathbf{R}}^{-1}\mathbf{u})^{-2}$ for $\hat{\mathbf{R}} = \text{diag}\{\text{corr}(\mathbf{Y}_{(j)})\}$. If the

CL estimator $\hat{\beta}_{\text{cl}}$ is defined as the solution to

$$\sum_{\|s_i - s_j\| \leq d} U_{ij}(\beta) = 0, \tag{21}$$

where each U_{ij} is the pairwise score function

$$\begin{pmatrix} \phi_i x_i \\ \phi_j x_j \end{pmatrix}^T \text{cov}^{-1} \begin{pmatrix} Y_i \\ Y_j \end{pmatrix} \begin{pmatrix} Y_i - p_i(\beta) \\ Y_j - p_j(\beta) \end{pmatrix},$$

then straightforward, albeit tedious computations give an explicit expression for $\mathbf{a}$ which enables us to write equation (21) as (19). We may then use (20) to compute var_{cl} numerically.

In our calculations we have taken cov (Y_i, Y_j) to be given (exactly) by equation (9), for an exponential correlation function $\rho(t) = \sigma^2 \rho^t$. It is of particular interest to compare the estimators when the efficiency of the IEE estimator,

$$\frac{\text{var}_{\text{opt}}}{\text{var}_{\text{ind}}} = \frac{(\mathbf{u}^T\mathbf{u})^2}{\mathbf{u}^T\mathbf{Ruu}^T\mathbf{R}^{-1}\mathbf{u}}, \tag{22}$$

is low. To determine such unfavorable configurations, observe that if the eigenvalues of $\mathbf{R}$ satisfy $\lambda_{\max}(\mathbf{R}) \gg \lambda_{\min}(\mathbf{R})$ and if $\mathbf{u}$ is the average, $\bar{\mathbf{v}}$, of the corresponding eigenvectors, then equation (22) is of the order $\lambda_{\min}(\mathbf{R})/\lambda_{\max}(\mathbf{R})$. For given σ^2, ρ, and β it is difficult to determine $\mathbf{x}$ to obtain such an eigenvector average, because $\mathbf{x}$ influences both $\mathbf{u}$ and $\mathbf{R}$ through the Φ_i; so for each combination of σ^2 and ρ we have approximated the unfavorable $\mathbf{x}$ by taking it to be $\bar{\mathbf{v}}$ for the matrix $\mathbf{R}_0$, which would obtain for $\beta = 0$.

Table 2

Asymptotic efficiencies (when covariance matrix is known) for 24×24 lattice when $p = 1$, $\beta = 1$ and $\mathrm{cov}(Y_i, Y_j) = \phi(x_i\beta)\phi(x_j\beta)\arcsin(\sigma^2\rho^{\|s_i - s_j\|})$. $C = $ condition number of the correlation matrix when $\beta = 0$. $\mathbf{1}$: vector of ones; $\bar{\mathbf{v}}$: unfavorable vector; $\mathbf{r}$: pseudorandom numbers. IBEE(k) uses $k \times k$ blocks; Cl(d) uses pairs of sites up to d units apart.

						Estimators				
Cov matrix					IBEE(k)			Cl(d)		
σ^2	ρ	C	$\mathbf{x}$	Ind	2	3	4	1	2	3
			$\mathbf{1}$	0.81	0.81	0.81	0.81	0.79	0.78	0.77
0.8	0.90	225	$\bar{\mathbf{v}}$	0.05	0.27	0.53	0.75	0.15	0.08	0.09
			$\mathbf{r}$	0.85	0.92	0.96	0.97	0.92	0.92	0.89
			$\mathbf{1}$	0.92	0.92	0.92	0.92	0.90	0.88	0.86
0.8	0.61	21	$\bar{\mathbf{v}}$	0.32	0.69	0.84	0.91	0.52	0.40	0.41
			$\mathbf{r}$	0.86	0.93	0.97	0.98	0.92	0.90	0.88
			$\mathbf{1}$	0.98	0.98	0.98	0.97	0.96	0.94	0.93
0.8	0.37	5	$\bar{\mathbf{v}}$	0.70	0.90	0.95	0.97	0.82	0.75	0.74
			$\mathbf{r}$	0.94	0.97	0.98	0.99	0.96	0.94	0.92
			$\mathbf{1}$	0.82	0.82	0.82	0.82	0.80	0.79	0.78
0.6	0.90	122	$\bar{\mathbf{v}}$	0.07	0.23	0.43	0.64	0.15	0.10	0.11
			$\mathbf{r}$	0.82	0.92	0.96	0.98	0.88	0.86	0.84
			$\mathbf{1}$	0.94	0.94	0.93	0.93	0.91	0.89	0.87
0.6	0.61	13	$\bar{\mathbf{v}}$	0.42	0.70	0.84	0.91	0.58	0.49	0.50
			$\mathbf{r}$	0.91	0.96	0.97	0.98	0.94	0.93	0.91
			$\mathbf{1}$	0.99	0.99	0.98	0.98	0.96	0.95	0.93
0.6	0.37	4	$\bar{\mathbf{v}}$	0.79	0.92	0.96	0.97	0.86	0.82	0.82
			$\mathbf{r}$	0.97	0.98	0.99	0.99	0.97	0.95	0.94

Table 2 shows the asymptotic efficiencies of the IEE, CL, and IBEE estimators for a 24×24 lattice, when $\beta = 1$. For each combination of σ^2 and ρ we considered three vectors $\mathbf{x}$: a vector of ones, a vector of uniformly distributed pseudo-random numbers with mean zero, and $\bar{\mathbf{v}}$ as described above. In the latter two cases $\mathbf{x}$ was normalized to have length $N^{1/2}$. We remark that the vector of ones is likely to be favorable to the IEE estimator, as it is in fact an eigenvector for the corresponding correlation matrix $(\sigma^2\rho^{\|s_i - s_j\|})$ in the one-dimensional (time-series) case.

From Table 2 we first observe that in virtually all cases the efficiency of the CL estimator *decreases* as a function of the maximum distance d in equation (21). Similar results, using simulations, were observed by Nott and Ryden (1999) in the context of covariance function estimation for image modeling; they gave evidence that differential weighting of the pairwise likelihoods improved performance.

We next note that all three estimators have approximately the same (relatively high) asymptotic efficiencies when $\mathbf{x}$ is a vector of ones or of pseudorandom numbers, but that the IEE estimator is relatively inefficient when $\mathbf{x} = \bar{\mathbf{v}}$, especially when the condition number of $\mathbf{R}_0$ is high. In these cases the CL estimator has approximately the same low efficiency, while the IBEE estimator performs considerably better.

6. Discussion

Albert and McShane (1995) have used GEE to analyze independent replications of a relatively small vector $\mathbf{Y}$ of spatially dependent components, using a working covariance matrix appropriate for continuous data. Our method differs in that we treat a large vector $\mathbf{Y}$, without replications. Also, although

assuming independence between blocks, by using Pearson's approximation within the blocks we explicitly take into account the binary nature of $\mathbf{Y}$; moreover, using the $c(I, J, d)$ as described preceding equation (17) allows us to explore the correlation structure on the underlying latent scale on which it is defined. This is important in view of recent research into problems with GEE that may arise when using misspecified working covariance matrices (Heagerty and Zeger, 1996; Qu, Lindsay, and Li, 2000; Oman and Zucker, 2001; Wang and Carey, 2003, 2004; Chaganty and Joe, 2004).

Our method has three "tuning parameters": the IBEE block size; the distance $d_{\max}$, used to define neighbors for estimating the working covariance parameters $\boldsymbol{\alpha} = (\sigma^2, \rho)$ as described following equation (17); and the subsampling-window size. We clearly wish to choose the IBEE blocks as large as is computationally feasible, because the major source of inaccuracy in our working correlation matrix is its being block-diagonal. Although our application and the asymptotic efficiencies in Table 2 suggest this approach should work well in practice, it would clearly be useful to study the performance of IBEE in a comprehensive simulation experiment.

Regarding $d_{\max}$, we also analyzed the vegetation data using $d_{\max} = 25$. The computations took only slightly longer than for $d_{\max} = 5$, and after eight iterations gave $\hat{\sigma}^2 = 0.82$ and $\hat{\rho} = 0.95$, essentially the same as equation (18). The regression coefficient estimates were also very close to those in Table 1, both in terms of values and significance, and led to identical subject-matter conclusions. On the other hand, for $d_{\max} = 1$ the algorithm did not converge, with $\hat{\boldsymbol{\alpha}}$ alternating between $(0.75, 0.9)$ and $(0.9, 0.75)$ (approximately). This probably reflects the fact that α_1 and α_2 are interchangeable if $d = 1$ in equation (10). Taking $d_{\max} = 1$ gives only $\|s_j - s_k\| = 1, 2^{\frac{1}{2}}$ in
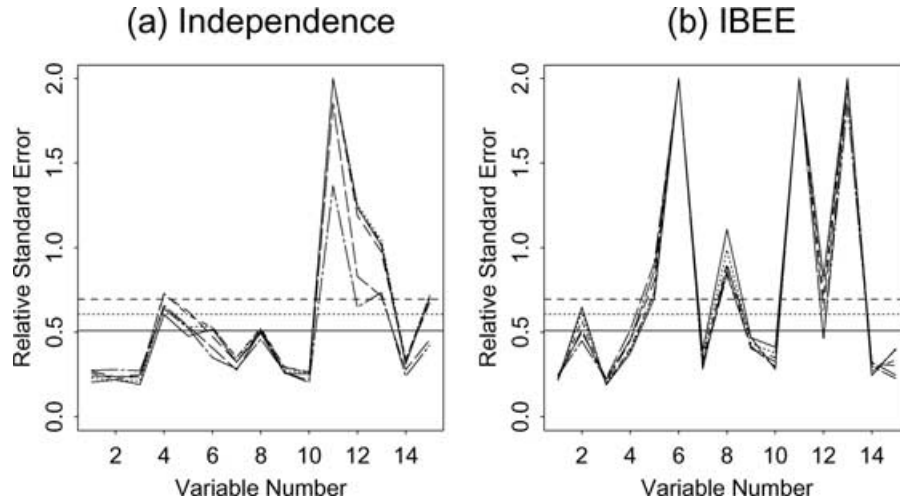
Figure 4. Relative standard errors for different subsampling-window sizes. Relative standard error for variable i (in Table 1) is $\min\{\hat{\sigma}_i/|\hat{\beta}_i|, 2\}$. Horizontal lines indicate significance at 5%, 10% and 15% levels. Window sizes: 8×8: ——— 20×22: $-----$ 10×11:. . . . . 29×31: — — — 15×16:- - - - - - 40×43:— - — -

equation (15), so that although $\boldsymbol{\alpha}$ is identified, the estimating equation is ill-conditioned. Increasing $d_{\max}$ to 2, the algorithm converged after 14 iterations to $\hat{\sigma}^2 = 0.89$ and $\hat{\rho} = 0.91$, somewhat further from equation (18); the regression coefficients were also less in agreement with Table 1. It thus appears that for the exponential correlation model, choosing $d_{\max}$ to include all correlations as low as 0.2 was necessary as well as adequate. The algorithm can, of course, be used with other working correlation functions, in which case different criteria might apply.

To choose the subsampling-window size in our application, as suggested by Heagerty and Lumley (2000) we estimated standard errors using a number of window sizes, and then chose the size giving the largest intercept standard error. Figure 4 shows the window sizes considered and the resulting relative standard errors $\hat{\sigma}_i/|\hat{\beta}_i|$ (truncated at 2) for the different variables i. For both IEE and IBEE, the standard errors using different window sizes are generally quite close to one another, and lead to similar levels of significance for the corresponding regression coefficients. An interesting exception occurs for variable 5 (*rd.dist*), whose significance level ranges from approximately 5% to 15% using IEE, as opposed to 15–27% using IBEE. This somewhat weakens the difference in conclusions concerning this variable, as discussed in Section 4. On the other hand, the different conclusions of IEE and IBEE regarding sporadic grazing by cattle (variable 8) still hold, regardless of the window size used.

Supplementary Materials

Web Appendices A (proofs of the results in Section 3) and B (software code for the computations in Section 4) are available under the Paper Information link at the Biometrics website http://www.tibs.org/biometrics.

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.

Arazi, A., Sharon, D., Khain, A., Huss, A., and Mahrer, Y. (1997). The windfield and rainfall distribution induced within a small valley: Field observations and 2-D numerical modeling. *Boundary-Layer Meteorology* **83,** 349–384.

Besag, J. E., York, J., and Mollie, A. (1991). Bayesian image restoration with applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43,** 21–59.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Carmel, Y. and Kadmon, R. (1999). Effects of grazing and topography on long term vegetation changes in a Mediterranean ecosystem. *Plant Ecology* **145,** 243–254.

Chaganty, N. R. and Joe, H. (2004). Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society, Series B* **66,** 851–860.

Christensen, O. F. and Waagepetersen, R. (2002). Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics* **58,** 280–286.

Christensen, O. F., Roberts, G. O., and Skold, M. (2006). Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics* **15,** 1–17.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43,** 671–681.

Crowder, M. J. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory* **2,** 305–330.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics (with discussion). *Applied Statistics* **47,** 299–350.

Doukhan, P. (1994). *Mixing: Properties and Examples.* New York: Springer-Verlag.

Guyon, X. (1995). *Random Fields on a Network: Modeling, Statistics and Applications.* New York: Springer-Verlag.

Heagerty, P. J. and Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association* **93,** 1099–1110.

Heagerty, P. J. and Lumley, T. (2000). Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association* **95,** 197–211.

Heagerty, P. J. and Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association* **91,** 1024–1036.

Landsman, V. (2000). An evaluation and application of a composite likelihood approach to analyzing spatial binary data. M. A. Thesis, Hebrew University of Jerusalem, Israel.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80,** 221–239.

Matheron, G. (1962). Traite de Geostatistique Appliquee, Tome I. *Memoires du Bureau de Recherches Geologiques et Minieres*, No. 14. Editions Technip, Paris.

Nott, D. J. and Ryden, T. (1999). Pairwise likelihood methods for inference in image models. *Biometrika* **86,** 661–676.

Oman, S. D. and Zucker, D. M. (2001). Modelling and generating correlated binary variables. *Biometrika* **88,** 287–290.

Pearson, K. (1901). Mathematical contributions to the theory of evolution—VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London, Series A* **195,** 1–47.

Pettitt, A. N., Weir, I. S., and Hart, A. G. (2002). A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing* **12,** 353–367.

Qu, A., Lindsay, B. G., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87,** 823–836.

Sharon, D. and Arazi, A. (1997). The distribution of wind-driven rainfall in a small valley: An empirical basis for numerical model verification. *Journal of Hydrology* **201,** 21–48.

Sherman, M. (1996). Variance estimation for statistics computed from spatial lattice data. *Journal of the Royal Statistical Society, Series B* **58,** 509–523.

Waller, L. A., Carlin, P. P., Xio, H., and Gelfand, A. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92,** 607–17.

Wang, Y.-G. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. *Biometrika* **90,** 29–41.

Wang, Y.-G. and Carey, V. (2004). Unbiased estimating equations from working correlation models for irregularly timed repeated measures. *Journal of the American Statistical Association* **99,** 845–853.

Weir, I. S. and Pettitt, A. N. (2000). Binary probability maps using a hidden conditional autoregressive Gaussian process with an application to Finnish common toad data. *Journal of the Royal Statistical Society, Series C* **49,** 473–484.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42,** 121–130.