REGULARIZED ESTIMATING EQUATIONS FOR MODEL SELECTION OF CLUSTERED SPATIAL POINT PROCESSES

Andrew L. Thurman[1], Rao Fu[2], Yongtao Guan[3] and Jun Zhu[2]

[1]*University of Iowa,* [2]*University of Wisconsin - Madison*
and [3]*University of Miami*

Abstract: Clustered spatial point processes are popular models for spatial point pattern data that contain clusters of events. For regression purposes, however, statistically rigorous methods for model selection appear to be lacking and are the focus of this paper. Here, an unbiased estimating equation is considered for parameter estimation to simplify computation and in addition, a weighted estimating equation is adopted to improve statistical efficiency. In particular, both regularized unweighted and regularized weighted estimating equations are developed for simultaneous variable selection and parameter estimation. Asymptotic properties of the proposed method are established and finite sample properties are assessed in a simulation study. For illustration, our method is applied to evaluate and quantify the relationship between the locations of a tropical tree species and over 200 covariates in a forest plot on the Barro Colorado Island.

Key words and phrases: Ecological application, spatial interaction, spatial point patterns, spatial statistics, variable selection, weighted estimating equations.

1. Introduction

Spatial point pattern data that contain clusters of events arise often in practice. For example, a particular tree species may grow in clusters within a forest stand due to various species characteristics and environmental factors. Another example is an infectious disease that may be contracted by susceptible individuals coming into contact with an infected individual, resulting in disease clusters in space. Clustered spatial point processes with regression account for spatial dependence and, in some situations, are more realistic than the inhomogeneous Poisson point processes that assume statistical independence among events. In many such studies, it is of interest to identify important factors underlying such spatial point patterns as species distributions or disease patterns. Identifying too few of the important factors can result in biased estimation of species or disease incidence maps, and false selection of extraneous factors can result in higher estimation variance. Statistically rigorous methods for selection of covariates appear to be lacking and are investigated in this paper.

A popular class of models for spatial point pattern data with clusters are the Cox processes (or, doubly stochastic Poisson processes) (Møller and Waagepetersen (2004)). These processes are Poisson point processes conditional on intensity functions that are nonnegative random fields. For parameter estimation and statistical inference, maximum likelihood estimation is possible but requires Markov chain Monte Carlo (MCMC) methods to approximate the likelihood function (Møller and Waagepetersen (2004)). Instead, an estimating equation based on the inhomogeneous Poisson point process has been developed by Waagepetersen (2007). Although this method is simpler and faster to implement than maximum likelihood estimation for clustered spatial point process models, the estimators are not as efficient. Thus, a weighted estimating equation has been proposed to regain some of the lost efficiency (Guan and Shen (2010)). Despite these methodological advances, it remains unclear how to perform model selection and in particular variable selection that identifies an appropriate subset of covariates in a clustered spatial point process with regression.

For spatial Poisson point processes, Thurman and Zhu (2014) developed penalized maximum likelihood estimation for variable selection. Here we extend this regularization method to simultaneous variable selection and parameter estimation for clustered spatial point processes in general. The main idea is to impose a penalty function on the objective functions associated with both unweighted and weighted estimating equations. While we focus our attention on an adaptive Lasso penalty function (Tibshirani (1996); Zou (2006)), comparison is made with the smoothly clipped absolute deviation (SCAD) (Fan and Li (2001)) and adaptive elastic net (Zou and Zhang (2009)) penalty functions. We establish asymptotic properties of our proposed method as to consistency, sparsity, and a central limit theorem.

The remainder of the paper is organized as follows. Section 2 defines spatial point processes and gives two examples. Section 3 describes methods based on estimating equations for the inference of these processes and develops a new method for variable selection and parameter estimation using regularized estimating equations. Section 4 gives the asymptotic properties, and Section 5 investigates the finite-sample properties of the proposed method in a simulation study. Section 6 illustrates the proposed method by a data example in ecology, followed by conclusions and discussion in Section 7.

2. Model

Let $(\Omega, \mathcal{A}, P)$ denote a probability space and $D \subset \mathbb{R}^d$ denote a spatial domain of interest. We consider the two-dimensional case $d = 2$. Let Y denote a mapping from $(\Omega, \mathcal{A}, P)$ to $\mathcal{N}_D$, where $\mathcal{N}_D$ denotes the set of locally finite configurations that are realizations $y \subset D$ such that $y \cap A$ is a finite set of spatial coordinates in

D for every bounded Borel set $A \subset D$. Let $N(A) = N_Y(A) = N(Y \cap A)$ denote the random number of events of Y in A. Then the mapping Y is said to be a spatial point process on D (Gaetan and Guyon (2010)).

First-order properties of a spatial point process indicate the spatial distribution of events in D. The moment measure of order one (or, intensity measure), μ_Y, is defined as $\mu_Y(B) = E\{N_Y(B)\} = E\{\sum_{s \in Y} I(s \in B)\}$, where B is a bounded Borel set in D and $I(\cdot)$ in an indicator function. Often it is assumed that there exists a first-order intensity function, $\lambda_Y(s)$, such that $\mu_Y(B)$ is the integral of λ_Y over the set B with respect to the Lebesgue measure $\mu_Y(B) = \int_B \lambda_Y(s)ds$. Essentially, $\mu_Y(ds) = \lambda_Y(s)ds$ can be interpreted as the probability that an event occurs in an infinitesimal Borel set ds.

2.1. Poisson process

Let $y_1, \ldots, y_n$ denote the observed spatial point pattern data comprising a set of n locations of events in D, and let $\mu(B) = \int_B \lambda(s)ds$ denote the intensity measure of the bounded Borel set $B \subset D$. A fundamental statistical model for spatial point pattern data is the spatial Poisson point process (henceforth, Poisson process), which is characterized by its intensity function $\lambda(\cdot)$. A formal definition of the Poisson process is given in supplementary materials.

Let $x(s) = (1, x_1(s), \ldots, x_p(s))^T$ denote a $(p+1) \times 1$ covariate vector at location $s \in D$. The intensity function $\lambda(\cdot)$ can be used to model the relationship between locations of events and covariates $x(\cdot)$. One commonly-used model has a log-linear specification $\lambda(s; \beta) = \exp\{x(s)^T\beta\}$, where $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is a $(p+1) \times 1$ regression coefficient vector.

2.2. Cox process

Let $\Lambda = \{\Lambda(s) : s \in D\}$ denote a nonnegative random field so that, with probability one, $\int_B \Lambda(s)ds < \infty$ for all bounded $B \subset D$. If the conditional distribution of X given Λ is a Poisson process on D with intensity function Λ, then X is said to be a Cox process driven by Λ (Møller and Waagepetersen (2004)). For an example of a type of Cox process, the Neyman-Scott process, see the supplementary materials.

3. Statistical Inference

3.1. Estimating equations

The log-likelihood function, up to a constant, for the Poisson process is in closed-form

$$\ell(\beta) = \sum_{i=1}^{n} \log \lambda(y_i; \beta) - \int_D \lambda(s; \beta)ds. \tag{3.1}$$

Let $u(\beta)$ denote the first-order derivative of $\ell(\beta)$ with respect to β. The maximum likelihood estimate (MLE) of β is the solution to the estimating equation

$$u(\beta) = \sum_{i=1}^{n} x(y_i) - \int_D x(s)\lambda(s;\beta)ds = 0. \tag{3.2}$$

For the Cox process described in Section 2.2, Monte Carlo methods can be applied for approximating the maximum likelihood estimates, but they are computationally prohibitive and rarely used in practice. In contrast, estimating equations provide a computationally simple and attractive alternative for estimating parameters of a clustered point process. It can be shown that (3.2) based on a Poisson process is an unbiased estimating equation for β in a Cox process (Waagepetersen (2007)). Let $\tilde{\beta}^{\mathrm{EE}}$ denote the solution to (3.2), referred to as the estimating equation (EE) estimate for a general clustered spatial point process.

The EE approach to estimating β produces a less efficient estimate than the MLE because information about the interaction of events is ignored. To regain some of the lost efficiency, a weighted estimating equation

$$u(\beta, w) = \sum_{i=1}^{n} w(y_i)x(y_i) - \int_D w(s)x(s)\lambda(s;\beta)ds = 0 \tag{3.3}$$

has been proposed (Guan and Shen (2010)), which corresponds to a weighted quasi-log-likelihood function

$$\ell_{\mathrm{W}}(\beta) = \sum_{i=1}^{n} w(y_i) \log \lambda(y_i;\beta) - \int_D w(s)\lambda(s;\beta)ds. \tag{3.4}$$

Let $\tilde{\beta}^{\mathrm{WEE}}$ denote the solution of (3.3), referred to as the weighted estimating equation (WEE) estimate.

To solve for β in (3.2), a quadrature approximation of (3.1) is employed,

$$\ell(\beta) \approx \sum_{i=1}^{n+M} v_i \Big\{ v_i^{-1}\Delta_i \log \lambda(s_i;\beta) - \lambda(s_i;\beta) \Big\}, \tag{3.5}$$

where v_i is a quadrature weight for point i, and Δ_i indicates if the point is an event or a dummy point. Because (3.5) is of the form of the log-likelihood for a weighted Poisson generalized linear model, the EE estimate $\tilde{\beta}^{\mathrm{EE}}$ can be computed using standard software packages for generalized linear models (Berman and Turner (1992)).

Similarly, to solve for β in (3.3), a quadrature approximation of (3.4) is

$$\ell_{\mathrm{W}}(\beta) \approx \sum_{i=1}^{n+M} w_i v_i \Big\{ v_i^{-1}\Delta_i \log \lambda(s_i;\beta) - \lambda(s_i;\beta) \Big\}, \tag{3.6}$$

where w_i is the value of the weight function at point i. The similarity between (3.5) and (3.6) allows us to compute $\tilde{\beta}^{\text{WEE}}$ using standard software packages for generalized linear models as well.

We divide the domain into a grid of rectangular pixels, and dummy points are centroids of these pixels. Then we set the quadrature weights $v_i = a/n_i$, where a is a common pixel area, and n_i is the number of events and dummy points in the same pixel as point i. Equations (3.5) and (3.6) become Riemann sum approximations of the integrals in (3.1) and (3.4), so the accuracy of the approximations depends on the pixel area a. For comparisons of the accuracy of different quadrature approximations see Baddeley and Turner (2000) and, for alternative approximations, see Rathbun (1996) and Waagepetersen (2008).

3.2. Regularized estimating equations

We define a regularized (or, penalized) quasi-log-likelihood function of β to be

$$\ell_{\text{P}}(\beta) = -\ell(\beta) + n \sum_{j=1}^{p} p_{\gamma_j}(|\beta_j|), \tag{3.7}$$

where $\ell(\beta)$ is the quasi-log-likelihood function given in (3.1) and γ_j is a tuning parameter corresponding to β_j for $j = 1, \ldots, p$. Similarly, we define a penalized weighted quasi-log-likelihood function of β to be

$$\ell_{\text{PW}}(\beta) = -\ell_{\text{W}}(\beta) + n \sum_{j=1}^{p} p_{\gamma_j}(|\beta_j|), \tag{3.8}$$

where $\ell_{\text{W}}(\beta)$ is given in (3.4). For example, the penalty term $p_{\gamma_j}(|\beta_j|) = \gamma_j|\beta_j|$ produces estimates that are possibly exactly zero, enabling variable selection and estimation simultaneously. It is also in the form of an adaptive Lasso, as the tuning parameters $\{\gamma_j\}$ vary for different regression coefficients $\{\beta_j\}$ (Zou (2006)). We call the values that minimize (3.7) and (3.8) penalized EE and penalized WEE estimates of β, and denote them by $\hat{\beta}^{\text{EE}}$ and $\hat{\beta}^{\text{WEE}}$, respectively.

To minimize $\ell_{\text{PW}}(\beta)$, we use a Laplace approximation of $\ell_{\text{W}}(\beta)$ in (3.4):

$$\ell_{\text{W}}^*(\beta) = (\beta - \hat{\beta}^{(m-1)})^T \frac{\partial \ell_{\text{W}}(\hat{\beta}^{(m-1)})}{\partial \beta} + \frac{1}{2}(\beta - \hat{\beta}^{(m-1)})^T \frac{\partial^2 \ell_{\text{W}}(\hat{\beta}^{(m-1)})}{\partial \beta \partial \beta^T}(\beta - \hat{\beta}^{(m-1)}),$$

where $\hat{\beta}^{(m-1)}$ is the minimizer of $\ell_{\text{PW}}(\beta)$ at the previous $(m-1)$th step, $m = 1, 2, \ldots$, and the initial value is set to $\hat{\beta}^{(0)} = \tilde{\beta}^{\text{WEE}}$. The intercept β_0 is left unpenalized, and we use the initial intercept estimate $\hat{\beta}_0^{(0)}$ as $\hat{\beta}_0^{(m-1)}$ for every m.

Next, the terms of $\ell_{\text{W}}^*(\beta)$ are rearranged. Let

$$y^* = (A^{-1})^T \left\{ \frac{\partial \ell_{\text{W}}(\hat{\beta}^{(m-1)})}{\partial \beta} - \frac{\partial^2 \ell_{\text{W}}(\hat{\beta}^{(m-1)})}{\partial \beta \partial \beta^T} \hat{\beta}^{(m-1)} \right\},$$

$X^* = A\text{diag}\{\gamma_j^{-1}\}_{j=1}^p$, and $\beta^* = \text{diag}\{\gamma_j\}_{j=1}^p\beta$, where A is the Cholesky factor of the negative Hessian of ℓ_W

$$-\frac{\partial^2 \ell_W(\hat{\beta}^{(m-1)})}{\partial\beta\partial\beta^T} = A^T A.$$

Then $\ell_W^*(\beta)$ can be rewritten as $\ell_W^*(\beta^*) = -(1/2)(y^* - X^*\beta^*)^T(y^* - X^*\beta^*)$ and thus, in the case of adaptive Lasso penalization, the estimate can be obtained via a least angle regression (LARS) algorithm (Efron et al. (2004)). Moreover, we let $\gamma_j = \gamma\log(n)(n|\tilde{\beta}_j^{\text{WEE}}|)^{-1}$ for $j = 1, \ldots, p$, where γ is a common tuning parameter (Zou (2006)). Alternative algorithms are available to obtain the estimates under SCAD (Fan and Li (2001)) and adaptive elastic net (Zou and Zhang (2009)) penalizations.

Define a weighted quasi-Bayesian information criterion (WQBIC) in this context by $\text{WQBIC}(\gamma) = -2\ell_W(\hat{\beta}(\gamma)) + e(\gamma)\log(n)$, where $e(\gamma) = \sum_{j=1}^p I\{\hat{\beta}_j(\gamma) \neq 0\}$ is the number of selected covariates with nonzero regression coefficient estimates. We fix a path of $\gamma \geq 0$ and select the tuning parameter γ and estimate $\hat{\beta}^{(m)}$ that minimize $\text{WQBIC}(\gamma)$. We replace $\hat{\beta}^{(m-1)}$ with $\hat{\beta}^{(m)}$ for $m = 1, 2, \ldots$ in $\ell_W^*(\beta)$ and iterate this procedure until some convergence criterion is met. After each iteration, $w(s)$ is re-estimated, since the selected covariates could be different from the previous iteration.

When data follow a Poisson process and the weights $w(s) \equiv 1$ for all s, WQBIC is exactly the Bayesian information criterion (BIC) corresponding to the log-likelihood function of a Poisson process. Zhang et al. (2010) gave conditions under which generalized information criteria, such as WQBIC, consistently selects the true model. These include conditions on the asymptotic behavior of the penalty function and goodness-of-fit measure, which in our case is $-2\ell_W(\hat{\beta})$. We have preliminary numerical evidence that WQBIC satisfies these conditions under an appropriately chosen penalty function. However, further investigation is needed that we discuss in Section 7.

The development and derivation above are for the WEE method. The EE method may be viewed as a special case of the WEE method and thus the estimation procedure can be applied with the weights $w(s) \equiv 1$. For further computational details including estimation of the weight function and standard errors, see the supplementary materials.

4. Asymptotic Properties

We adopt an increasing spatial domain framework (Guan and Loh (2007)). Let X denote a two-dimensional spatial point process observed over a spatial domain D_n with boundary ∂D_n, area $|D_n|$, and length of the boundary $|\partial D_n|$.

Let $\beta^0 = \left((\beta_1^0)^T, (\beta_2^0)^T\right)^T$ denote a p-dimensional vector of true coefficient values, where β_1^0 is the s-dimensional vector of nonzero coefficients and $\beta_2^0 = 0$ is $(p-s)$-dimensional, and x_1 and x_2 denote, respectively, the s and $(p-s)$ vectors of covariates. Let $\beta_{(j)}$ denote the jth component of the vector β. Write the penalized WEE estimate as $\hat{\beta} = \left(\hat{\beta}_1^T, \hat{\beta}_2^T\right)^T$. Let $a_n = \max\{\gamma_j : j = 1, \ldots, s\}$ and $b_n = \min\{\gamma_j : j = s+1, \ldots, p\}$.

Theorem 1. *Assume* (A.1)−(A.5), *given in the supplementary materials, hold and let* $a_n = O(|D_n|^{-1/2})$.

(a) *With probability tending to 1, there is a local maximizer $\hat{\beta}$ of $\ell_{PW}(\beta)$ such that* $\|\hat{\beta} - \beta^0\|_2 = O_p(|D_n|^{-1/2} + a_n)$.

(b) *If, in addition, $|D_n|^{1/2}b_n \to \infty$ as $n \to \infty$, then $P(\hat{\beta}_2 = 0) \to 1$.*

(c) *If, in addition, $a_n = o(|D_n|^{-1/2})$, then*

$$|D_n|^{1/2}\Sigma_n(w, \beta_1^0)^{-1/2}(\hat{\beta}_1 - \beta_1^0) \xrightarrow{d} N\left(0, I_{s \times s}\right),$$

where $\Sigma_n(w, \beta_1^0) = |D_n|A_n(w, \beta_1^0)^{-1}\{B_n(w, \beta_1^0) + C_n(w, \beta_1^0)\}A_n(w, \beta_1^0)^{-1}$ *and*

$$A_n(w, \beta_1^0) = \int_{D_n} w(s)x_1(s)x_1(s)^T \lambda(s; \beta_1^0)ds,$$

$$B_n(w, \beta_1^0) = \int_{D_n} w(s)^2 x_1(s)x_1(s)^T \lambda(s; \beta_1^0)ds,$$

$$C_n(w, \beta_1^0) =$$
$$\int_{D_n} w(s)x_1(s)\lambda(s; \beta_1^0)\left[\int_{D_n} w(u)x_1(u)^T \lambda(u; \beta_1^0)\{g(u - s) - 1\}du\right]ds.$$

In Theorem 1, part (a) establishes the existence of the penalized WEE estimate $\hat{\beta}$ and consistency at the rate $|D_n|^{1/2}$. Part (b) ensures the sparsity of $\hat{\beta}$, the estimate correctly sets β_2 to zero with probability tending to 1 as $|D_n| \to \infty$. Part (c) establishes the asymptotic normality of $\hat{\beta}_1$ at rate $|D_n|^{1/2}$. The proof of Theorem 1 is given in the supplementary materials.

5. Simulation Study

5.1. Simulation set-up

We conducted a simulation study with a setup similar to a previous study by Waagepetersen (2007). The spatial domain was $D = [0, 1{,}000] \times [0, 500]$. We centered and scaled the 201×101 pixel images of elevation (z_1) and slope (z_2), which are contained in the **bei** data set of the **spatstat** library in R (R Development Core Team (2011)), and used them as two covariates to generate point pattern

data from a Thomas process, a type of Neyman-Scott process. See the supplementary materials for more on the Thomas process. In addition, twenty 201×101 pixel images of covariates were generated as noise covariates with regression coefficients zero. Each noise covariate (z_j) was first generated as standard Gaussian white noise but transformed, together with z_1 and z_2, to have multicollinearity. We took $x(s) = V^T z(s)$, where $z(s) = (z_1(s), z_2(s), z_3(s), \ldots, z_{22}(s))^T$, $\Sigma = V^T V$, and $(\Sigma)_{ij} = (\Sigma)_{ji} = 0.5^{|i-j|}$ for $i = 1, 2, \ldots, 22$, $j = 1, \ldots, 22$, except $(\Sigma)_{12} = (\Sigma)_{21} = 0$, to preserve the correlation between z_1 and z_2 from the data set. Sample variograms indicated that x_1 and x_2 were spatially dependent which appears to induce some spatial dependence in x_3 but little in $x_4, \ldots, x_{22}$. Two values of the expected number of children events, $\mu = 400$ and $1,600$, were used to represent relatively small and large spatial point patterns. The true intensity function of the Thomas process was set to $\lambda(s) = \exp\{\beta_0 + \beta_1 x_1(s) + \beta_2 x_2(s)\}$, where $\beta_1 = 2$ represents a relatively large effect of elevation, $\beta_2 = 0.5$ reflects a relatively small effect of slope, and β_0 was chosen so that μ was 400 or 1,600. Three values of κ ($\kappa = 5 \times 10^{-5}$, 1×10^{-4}, and 5×10^{-4}) were used for different levels of spatial interactions while we let $\omega = 10$. The three values of κ correspond to 25, 50, and 250 expected number of parent events in D, respectively. For each of the six combinations of κ and μ, 500 spatial point patterns were generated from the corresponding Thomas process.

In each simulation scenario, the EE and WEE methods under the adaptive Lasso (AL) penalty were applied, as described in Section 3.2. We focused on one iteration and used $\hat{\beta}^{(1)}$ as the final estimate. For comparison, we considered two alternative penalty functions. For the SCAD penalty, we replaced the penalty term in (3.7) and (3.8) with $n \sum_{j=1}^{p} p_\gamma^S(|\beta_j|)$, where

$$p_\gamma^S(\beta) = \gamma \left\{ I(\beta \leq \gamma) + \frac{(a\gamma - \beta)_+}{(a-1)\gamma} I(\beta > \gamma) \right\},$$

for $\beta > 0$, γ and a are the tuning parameters. We applied a one-step locally linear approximation (LLA) of the SCAD penalty (Zou and Li (2008)) and set $a = 3.7$ (Fan and Li (2001)). For the adaptive elastic net (AENET) penalty, the penalty term in (3.7) and (3.8) was replaced with $n \sum_{j=1}^{p} p_{\gamma_j}^E(\beta_j)$ where

$$p_{\gamma_j}^E(\beta_j) = \gamma \left\{ \frac{1}{2}(1-\alpha)\beta_j^2 + \alpha \hat{w}_j |\beta_j| \right\},$$

γ is a tuning parameter, $\hat{w}_j$ is a data adaptive weight similar to adaptive Lasso, and $0 \leq \alpha \leq 1$. For solving (3.7) and (3.8), the `glmnet` library in R was used. Specifically, a decreasing sequence of γ was identified first, in which the starting value $\gamma_{\max}$ is the γ such that the entire vector $(\hat{\beta}_1, \ldots, \hat{\beta}_p)^T = 0$. For each value of γ, a quadratic approximation was formed to the negative log-likelihood

evaluated at the current estimates. Then, a coordinate descent method was applied to solve a penalized weighted least squares problem (Friedman, Hastie, and Tibshirani (2010)). Finally, the WQBIC(γ) was minimized to obtain γ.

5.2. Simulation results

Table 1 presents the selection properties of the penalized EE and penalized WEE methods under different penalty functions. For different values of μ and κ, we report the proportion of times when the individual covariates, elevation x_1, and slope x_2 were correctly kept in the selected model, and the average proportion of times when the noise covariates x_3 to x_{22} were correctly selected. Also given is the proportion of times when the entire correct model (comprising only x_1 and x_2) was correctly selected.

For the EE and WEE methods, under larger values of κ, which indicate weaker spatial dependence, x_1 and x_2 are selected and x_3 to x_{22} are eliminated more frequently from the model. However, WEE tends to underfit models, leaving out x_2, which has a smaller regression coefficient, while EE tends to overfit models, incorrectly selecting covariates among x_3 to x_{22}. As μ and thus the sample size n increases, WEE improves in selecting x_2, and hence the entire model, while EE continues to select covariates among x_3 to x_{22}, resulting in poorer selection of the entire model. While the solution paths under the penalized EE and WEE methods are comparable, the selected tuning parameter ($\hat{\gamma}$) tends to be smaller with EE, leading to an overselection by EE. We conjecture that overselection by EE is a way to compensate for initially ignoring the additional variation due to spatial clustering of events. Among different penalties, for EE, the selection properties of AL and AENET are similar, but the SCAD penalty can have quite poor selection properties when $\mu = 1,600$ due to selecting among x_3 to x_{22}. For WEE, all three penalties perform similarly, with the AL correctly selecting the entire model with somewhat higher frequencies than AENET or SCAD.

Tables 2 and 3 give properties of the estimates for β_1 and β_2, respectively, in terms of bias, standard deviation (SD), and root mean squared error (RMSE), based on the first 100 nonzero estimates. Under the oracle where the correct model containing x_1 and x_2 is assumed to be known, WEE tends to have somewhat larger bias but smaller SD and RMSE than EE. When penalties are applied, similar results are observed except for β_1 under AL. These results reflect a bias-variance trade-off and, for the most part, an overall improvement of WEE over EE in terms of smaller RMSE. Further, as expected, with larger values of μ or κ, the bias, SD, and RMSE values tend to be smaller due to larger sample size or weaker spatial dependence. Among the different penalties, EE has similar RMSE values. In contrast, for estimates of β_1, WEE with AL penalization tends to have the largest RMSE values, followed by SCAD, with AENET having the

Table 1. For different values of the model parameters μ ($\times 100$) and κ ($\times 10^{-5}$), the proportion of times when the covariates x_1 and x_2 were selected, the average proportion of times when the noise covariates x_3 to x_{22} ($x_3 \sim x_{22}$) were not selected, and the proportion of times when the entire correct model (x_1, x_2) was correctly selected, under adaptive Lasso (AL), smoothly clipped absolute deviation (SCAD), and adaptive elastic net (AENET) penalties, using either estimating equations (EE) or weighted estimating equations (WEE).

		AL				SCAD				AENET			
μ	κ	x_1	x_2	$x_3 \sim x_{22}$	x_1, x_2	x_1	x_2	$x_3 \sim x_{22}$	x_1, x_2	x_1	x_2	$x_3 \sim x_{22}$	x_1, x_2
							EE						
4	5	1.00	0.78	0.91	0.24	1.00	0.76	0.83	0.05	1.00	0.78	0.91	0.15
	10	1.00	0.87	0.95	0.41	1.00	0.83	0.88	0.08	1.00	0.86	0.95	0.38
	50	1.00	1.00	0.98	0.71	1.00	0.98	0.94	0.33	1.00	0.87	0.98	0.65
16	5	1.00	0.90	0.73	0.05	1.00	0.92	0.45	0.00	1.00	0.89	0.81	0.08
	10	1.00	0.92	0.87	0.18	1.00	0.95	0.54	0.00	1.00	0.91	0.89	0.18
	50	1.00	1.00	0.97	0.53	1.00	1.00	0.69	0.00	1.00	0.99	0.98	0.75
							WEE						
4	5	0.95	0.33	1.00	0.32	0.95	0.21	1.00	0.20	0.96	0.23	1.00	0.22
	10	0.98	0.53	1.00	0.51	0.98	0.39	1.00	0.37	0.98	0.47	1.00	0.46
	50	1.00	0.87	0.99	0.76	1.00	0.78	0.97	0.57	1.00	0.50	1.00	0.48
16	5	0.97	0.42	1.00	0.42	0.98	0.21	1.00	0.21	0.97	0.25	1.00	0.25
	10	0.99	0.62	1.00	0.62	0.99	0.46	1.00	0.45	0.99	0.45	1.00	0.45
	50	1.00	0.96	1.00	0.94	1.00	0.92	0.99	0.83	1.00	0.92	1.00	0.91

smallest RMSE values. For estimates of β_2, the RMSE values are similar among the different penalties.

The standard error estimation properties for estimates of β_1 and β_2 are shown in Tables S1 and S2 in the supplementary materials, respectively. For the penalized methods, in Table S1, x_1 was always included in the model but x_2 to x_{22} were subject to penalty and, in Table S2, x_2 was always included in the model but x_1 and x_3 to x_{22} were penalized. The medians of the estimated standard errors were compared with the nominal standard errors. The results suggest that the standard error estimates under different penalties perform reasonably well for each scenario.

The computation was carried out using `R` on a 64 bit Linux operating system with 8 cores and 48-64 GB of RAM. The average computing time for different penalties and different parameter values ranged from 0.4 to 6.7 seconds.

6. Data Example

In a 50 hectare region ($D = 1{,}000\text{m} \times 500\text{m}$) of the Barro Colorado Island (BCI) in central Panama, all free-standing woody stems at least 1 cm diameter at

Table 2. For different values of the model parameters μ ($\times 100$) and κ ($\times 10^{-5}$), the bias, standard deviation (SD), and root mean squared error (RMSE) values for estimates of β_1, under adaptive Lasso (AL), smoothly clipped absolute deviation (SCAD), adaptive elastic net (AENET) penalties and the true model (Oracle), using either estimating equations (EE) or weighted estimating equations (WEE).

		AL			SCAD			AENET			Oracle		
μ	κ	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
						EE							
4	5	-0.04	0.46	0.46	-0.09	0.48	0.48	-0.11	0.48	0.49	-0.09	0.51	0.52
	10	0.01	0.43	0.43	-0.03	0.45	0.44	-0.06	0.39	0.40	0.00	0.44	0.43
	50	0.00	0.18	0.18	-0.03	0.19	0.19	-0.10	0.37	0.38	-0.01	0.18	0.18
16	5	-0.03	0.45	0.45	-0.05	0.46	0.46	-0.08	0.44	0.45	-0.04	0.47	0.47
	10	-0.06	0.36	0.37	-0.08	0.37	0.37	-0.10	0.36	0.37	-0.07	0.37	0.37
	50	0.01	0.16	0.16	-0.01	0.16	0.16	-0.01	0.16	0.16	0.00	0.16	0.16
						WEE							
4	5	-0.36	0.50	0.61	-0.21	0.41	0.46	-0.21	0.38	0.43	-0.15	0.47	0.49
	10	-0.24	0.37	0.44	-0.16	0.33	0.36	-0.15	0.30	0.33	-0.05	0.30	0.30
	50	-0.09	0.27	0.28	-0.09	0.22	0.24	-0.13	0.30	0.32	-0.03	0.20	0.20
16	5	-0.32	0.38	0.50	-0.23	0.34	0.41	-0.20	0.30	0.36	-0.07	0.30	0.31
	10	-0.21	0.29	0.36	-0.15	0.27	0.31	-0.12	0.24	0.27	-0.05	0.25	0.26
	50	-0.10	0.19	0.21	-0.11	0.16	0.19	-0.06	0.14	0.15	-0.04	0.12	0.12

breast height were tagged, measured, mapped, and identified to the species level, resulting in maps of over 300 different species of trees (Condit (1998); Hubbell et al. (1999, 2005)). It is of interest to determine how the presence of a given tree species is related to the presence of other tree species and various environmental factors. Because the number of tree species and environmental factors is large, the methodology developed here is well-suited to selecting covariates and estimating regression coefficients in a computationally efficient manner. The selected covariates could produce more precise mapped estimates of species occurrence and suggest important ecological relationships.

Here we focus the analysis on the locations of 4,026 *B. pendula* tree stems. We model the intensity of *B. pendula* trees as a log-linear function of elevation, slope, 13 soil characteristics, and the intensities of 214 other tree species. Figure 1 contains maps of the locations of the *B. pendula* tree stems, elevation, slope, and concentration of nitrogen. *B. pendula* trees seem to appear in greater abundance in areas of high elevation and steep slope, but do not appear to be closely related to the level of soil nitrogen concentration.

We applied six approaches to variable selection and estimation to the *B. pendula* data set. Each of the six approaches was a combination of either the EE or WEE method under AL, SCAD, or AENET penalty. Out of the 229 covariates,

Table 3. For different values of the model parameters μ ($\times 100$) and κ ($\times 10^{-5}$), the bias, standard deviation (SD), and root mean squared error (RMSE) values for estimates of β_2, under adaptive Lasso (AL), smoothly clipped absolute deviation (SCAD), adaptive elastic net (AENET) penalties and the true model (Oracle), using either estimating equations (EE) or weighted estimating equations (WEE).

		AL			SCAD			AENET			Oracle		
μ	κ	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
							EE						
4	5	0.01	0.31	0.31	0.03	0.32	0.32	0.05	0.32	0.32	-0.11	0.40	0.41
	10	0.02	0.19	0.19	-0.01	0.19	0.19	0.04	0.16	0.16	0.00	0.23	0.23
	50	-0.08	0.13	0.15	-0.11	0.13	0.17	0.03	0.14	0.14	-0.03	0.13	0.13
16	5	0.02	0.30	0.30	0.02	0.33	0.32	0.04	0.29	0.29	-0.05	0.36	0.36
	10	-0.01	0.23	0.23	-0.03	0.23	0.23	0.02	0.21	0.21	-0.12	0.28	0.30
	50	-0.02	0.12	0.12	-0.04	0.12	0.12	-0.01	0.11	0.11	0.00	0.11	0.11
							WEE						
4	5	0.01	0.28	0.28	0.09	0.31	0.32	0.15	0.31	0.34	-0.07	0.30	0.31
	10	-0.04	0.15	0.15	-0.01	0.15	0.15	0.10	0.12	0.15	0.00	0.18	0.18
	50	-0.10	0.12	0.15	-0.13	0.12	0.18	0.07	0.12	0.14	-0.04	0.15	0.15
16	5	0.03	0.19	0.19	0.14	0.25	0.28	0.18	0.25	0.30	-0.05	0.28	0.28
	10	-0.03	0.13	0.13	-0.01	0.14	0.14	0.10	0.11	0.15	-0.04	0.20	0.21
	50	-0.09	0.09	0.13	-0.13	0.10	0.16	-0.01	0.08	0.08	-0.02	0.08	0.08

61 covariates were selected using the penalized EE method against only 6 using the penalized WEE method under the AL penalty. In comparison, 86 and 97 out of 229 covariates were selected using penalized EE and 0 and 3 covariates using penalized WEE under SCAD and AENET penalties. A total of 56 covariates selected using penalized EE were common among the three penalties. The penalized WEE method chose far fewer covariates than the EE method under each penalty. To save space, Table 4 reports the parameter estimates and their standard errors for the 7 covariates selected by penalized WEE method under the AL, SCAD, and AENET penalties. Although the magnitudes of the penalized EE and penalized WEE estimates can be quite different, the signs all agree with each other. Also, the standard errors of the penalized WEE estimates are consistently smaller than those of the penalized EE estimates, demonstrating the gain of statistical efficiency achieved by WEE.

These results suggest that *B. pendula* trees are more likely to occur in areas of higher elevation and slope. Further, higher levels of manganese (Mn) and lower levels of phosphorus (P) and zinc (Zn) concentrations in soil are associated with higher incidence of *B. pendula* trees. Two of the 214 tree species, *O. whitei* and *P. panamense*, were selected and appear to be positively associated with the occurrences of *B. pendula* trees. Figure 2 contains plots of these two covariates.

Table 4. Estimates and standard errors (SE) of regression coefficients under adaptive Lasso (AL), smoothly clipped absolute deviation (SCAD), and adaptive elastic net (AENET), using either estimating equations (EE) or weighted estimating equations (WEE).

		AL		SCAD		AENET	
	Coefficient	Estimate	SE	Estimate	SE	Estimate	SE
EE	Intercept	-11.60	3.69	-11.06	3.63	-10.51	3.48
	Elevation	0.05	0.02	0.04	0.02	0.04	0.02
	Slope	4.56	2.89	4.24	2.81	3.69	2.68
	Mn	0.0011	0.0012	0.0012	0.0012	0.0011	0.0011
	P	-0.34	0.13	-0.31	0.13	-0.28	0.12
	Zn	-0.11	0.11	-0.07	0.11	-0.07	0.10
	O. whitei	4.25	7.21	4.66	6.86	4.00	6.68
	P. panamense	5.91	5.66	5.59	5.55	4.87	5.27
WEE	Intercept	-12.94	2.10	-5.51	0.43	-5.32	0.30
	Elevation	0.05	0.01	0.00	–	0.00	–
	Slope	8.83	1.96	0.00	–	2.40	1.83
	Mn	0.0009	0.0007	0.00	–	0.00	–
	P	-0.33	0.08	0.00	–	-0.14	0.07
	Zn	-0.07	0.04	0.00	–	0.00	–
	O. whitei	0.00	–	0.00	–	8.24	6.07
	P. panamense	1.31	4.28	0.00	–	0.00	–

Notice that the locations of *O. whitei* and *P. panamense* trees coincide with the locations of *B. pendula* trees, especially in the far right and far left regions of the domain, respectively. Figure S1 in the supplementary materials contains plots of the log intensity functions for each of the six approaches. The intensity function using the penalized WEE method is smoother than that using EE, possibly because many of the 214 tree species were selected using the penalized EE method, and these covariates can be highly spatially variable.

7. Conclusions and Discussion

We have developed a regularized method to perform variable selection and parameter estimation for spatial point process models with clusters using penalized unweighted and weighted estimating equations. The simulation results indicated that, as the number of events increased, the selection and estimation properties of the penalized WEE method tended to improve; the estimation properties of the penalized EE method tended to improve as well, but the selection properties weakened for larger spatial point patterns. In addition, a theoretical result showed that the proposed methods produce estimates that are consistent, sparse, and asymptotically normal. An application in forest ecology demonstrated that, for over 200 covariates, our method can be applied to perform vari-
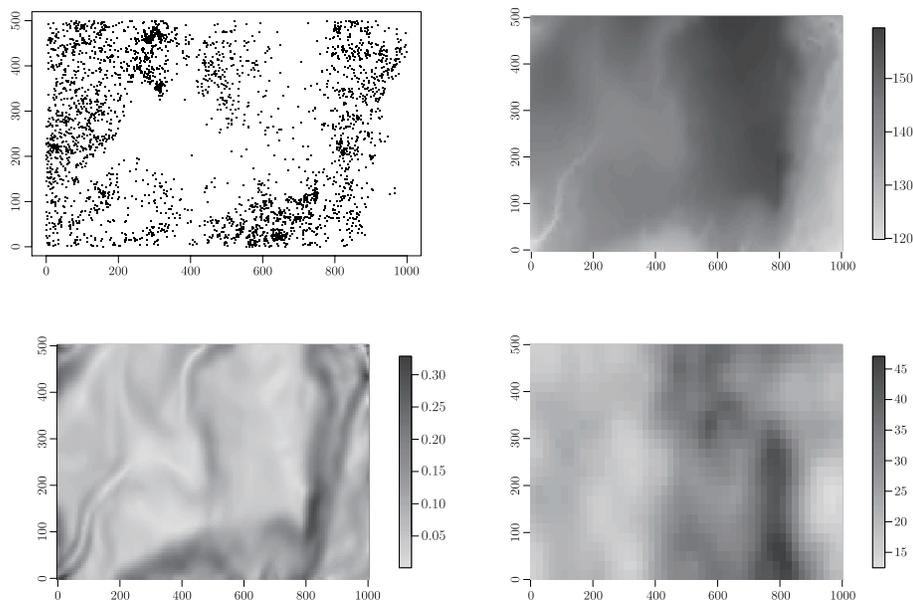
Figure 1. Map of locations of *B. pendula* stems (upper left), elevation (upper right), slope (lower left), and soil nitrogen concentration (lower right). Darker colors indicate larger values.
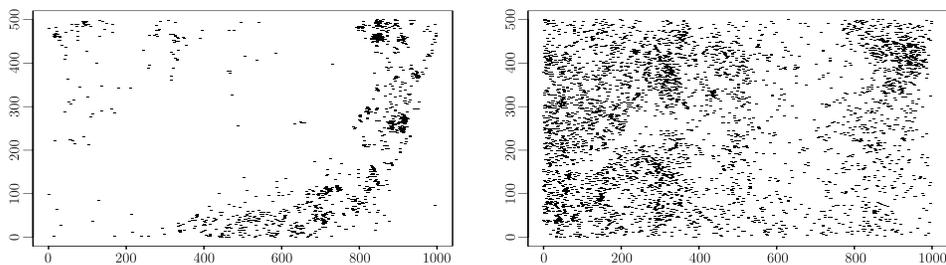
Figure 2. Maps of locations of *O. whitei* (left) and *P. panamense* (right) stems.

able selection and estimate regression coefficients in a computationally efficient manner.

The simulation study in Section 5 provided evidence that the penalized WEE method, compared to the penalized EE method, has superior selection properties and among the nonzero estimates, has slightly larger biases but smaller variances with an overall improved estimation. Further, when applying the penalized EE method to large spatial point patterns, we suspect that covariates were overselected in the forest ecology example. In contrast, the penalized WEE method tends to downweigh areas with many events when interactions are strong, so covariates unrelated to the response tended to be correctly ignored.

The theoretical results in Section 4 have been established under increasing domain asymptotics, building upon results established for unpenalized EE (Guan and Loh (2007)) and unpenalized WEE estimation (Guan and Shen (2010)). While Waagepetersen (2007) considered infill asymptotics for unpenalized EE estimation, the class of processes is more restrictive, but it would be interesting to investigate the theoretical properties of penalized EE and WEE under infill asymptotics. Further, the WQBIC measure used to select the tuning parameter is a special case of the generalized information criterion (Zhang et al. (2010)). We have preliminary numerical evidence that the regularity conditions necessary for selection consistency are satisfied, but further investigation is needed. We leave this for future work.

Acknowledgement

References

Baddeley, A. and Turner, R. (2000). Practical maximum pseudo-likelihood for spatial point patterns (with discussion). *Austral. N. Z. J. Statist.* **42**, 283-322.

Berman, M. and Turner, T.R. (1992). Approximating point process likelihoods with GLIM. *J. Roy. Statist. Soc. Ser. C* **41**, 31-38.

Condit, R. (1998). *Tropical Forest Census Plots.* Springer-Verlag and R. G. Landes Company, Berlin, Germany and Georgetown, Texas.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.* **32**, 407-451.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96**, 1348-1360.

Friedman, J., Hastie, T and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Statist. Software* **33**, Issue 1.

Gaetan, C. and Guyon, X. (2010). *Spatial Statistics and Modeling*. Springer, New York.

Guan, Y. and Loh, J. M. (2007). A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *J. Amer. Statist. Assoc.* **102**, 1377-1386.

Guan, Y. and Shen, Y. (2010). A weighted estimating equation approach to inhomogeneous spatial point processes. *Biometrika* **97**, 867-880.

Hubbell, S. P., Condit, R. and Foster, R. B. (2005). Barro colorado forest census plot data. URL http://ctfs.arnarb.harvard.edu/webatlas/datasets/bci.

Hubbell, S. P., Foster, R. B., O'Brien, S. T., Harms, K. E., Condit, R., Wechsler, B., Wright, S. J. and Loo de Lao, S. (1999). Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* **283**, 554-557.

Møller, J. and Waagepetersen, R. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall, Boca Raton.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna, Austria ISBN 3-900051-07-0 http://www.R-project.org/.

Rathbun, S. L. (1996). Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics* **52**, 226-242.

Thurman, A. L. and Zhu, J. (2014). Variable selection for spatial Poisson point processes via a regularization method. *Statist. Methodology* **17**, 113-125.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58**, 267-288.

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63**, 252-258.

Waagepetersen, R. (2008). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika* **95**, 351-363.

Zhang, Y., Li, R. and Tsai, C.-L. (2010). Regularization parameter selections via generalized information criterion. *J. Amer. Statist. Assoc.* **105**, 312-323.

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101**, 1418-1429.

Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* (with discussion) **36**, 1509-1533.

Zou, H. and Zhang, H.H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Ann. Statist.* **37**, 1733-1751.

Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, USA.

E-mail: andrew-thurman@uiowa.edu

Department of Statistics, University of Wisconsin, Madison, WI 53706, USA.

E-mail: rfu7@wisc.edu

Department of Management Science, University of Miami, Miami, FL 33124, USA.

E-mail: yguan@bus.miami.edu

Department of Statistics and Department of Entomology, University of Wisconsin, Madison, WI 53706, USA.

E-mail: jzhu@stat.wisc.edu