Statistical Methodology 17 (2014) 113–125

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Variable selection for spatial Poisson point processes via a regularization method

Andrew L. Thurman [a], Jun Zhu [b,c,*]

[a] Department of Statistics, University of Iowa, 241 Schaeffer Hall, Iowa City, IA 52242, USA

[b] Department of Statistics, University of Wisconsin–Madison, 1300 University Avenue, Madison, WI 53706, USA

[c] Department of Entomology, University of Wisconsin–Madison, 1300 University Avenue, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 2 June 2012
Received in revised form
14 June 2013
Accepted 1 August 2013

Keywords:
Adaptive Lasso
Intensity function
Maximum likelihood estimation
Model selection
Spatial statistics

ABSTRACT

It is often of interest to use regression analysis to study the relationship between occurrence of events in space and spatially-indexed covariates. One model for such regression analysis is the Poisson point process. Here, we develop a method to perform the selection of covariates and the estimation of model parameters simultaneously for this model via a regularization method. We assess the finite-sample properties of our method with a simulation study. In addition, we propose a variant of our method that allows the selection of covariates at multiple pixel resolutions. For illustration, we consider the locations of a tree species, *Beilschmiedia pendula*, in a study plot at Barro Colorado Island in central Panama. We find that *Beilschmiedia pendula* occurs in greater abundance at locations with higher elevation and steeper slope. Also, we identify three species to which *Beilschmiedia pendula* tends to be attracted, two species by which it appears to be repelled, and a species with no apparent relationship.

1. Introduction

Spatial point processes model the random locations of events in space and have been studied as statistical models for the analysis of spatial point pattern data (e.g., [14]). Applications of spatial point process models and methods appear in such varied fields as ecology, geography, forestry, and

* Corresponding author at: Department of Statistics, University of Wisconsin–Madison, 1300 University Avenue, Madison, WI 53706, USA. Tel.: +1 6088907619.

E-mail addresses: thurman@stat.wisc.edu (A.L. Thurman), jzhu@stat.wisc.edu, jun.zhu.e@gmail.com (J. Zhu).

1572-3127/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.stamet.2013.08.001

epidemiology. A common objective of these studies is to connect auxiliary spatial information to the occurrence of events of interest in space. Spatial point processes with regression are well-suited for this objective and are the focus of this paper.

For example, in a long-term ecological monitoring program, nearly 400,000 individual trees have been censused since the 1980s at the Barro Colorado Island (BCI) in central Panama [4,12,11]. In particular, free-standing woody stems with at least 10 mm diameter at breast height were identified, tagged, and mapped, resulting in maps of tree stems for over 300 species. Using spatial point processes with regression to analyze these data allows for the investigation of the relationships between the tree species and in relation to environmental factors. Further, with a regularization method to be developed here, we may investigate the varying spatial scales simultaneously for these relationships. We will illustrate this in Sections 5 and 6.

Poisson point processes are popular for the purpose of regression analysis, because they are theoretically tractable and straightforward to implement computationally. The moments of the point process and the likelihood function can be derived analytically. Likelihood-based methods have been developed for fitting these models to the data and obtaining parameter estimates. An approximation of the likelihood can be readily implemented to fit models using the existing software for weighted generalized linear models [2]. In addition, the asymptotic properties of maximum-likelihood estimators for the parameters of these models, namely consistency and asymptotic normality, have been established. For example, Rathbun and Cressie [16] adopted an increasing spatial domain asymptotic framework, where the spatial domain of the Poisson point process is allowed to grow, and established asymptotic results using the theory of local asymptotic normality [13].

A challenging component of any regression procedure is deciding which set of covariates should be chosen for the model. Selection of covariates (or, variable selection) has been an active area of research in statistics, because valid inference relies heavily on appropriate model specification that mitigates overfitting which results in variance inflation and underfitting which leads to bias. Traditional methods include the stepwise procedures (see, e.g., [5]). Starting from an initial set of covariates, the stepwise procedures consider adding or eliminating a covariate from the current set at each step using a criterion such as an F-statistic. In certain cases, such a procedure can be biased. For example, the F-test corresponding to the F-statistic was developed under the assumption that the two candidate models are fixed in advance, so a stepwise procedure with this criterion tends to be biased because the candidate models are chosen adaptively and thus not fixed [9].

More recently, Tibshirani [17] introduced variable selection via a penalized (or, regularization) approach. In particular, the least absolute shrinkage and selection operator (Lasso) produced both zero and nonzero estimates of the regression coefficients, performing selection and estimation simultaneously. This regularization approach has been modified to achieve improved selection properties (e.g., [7,22,24]). Most research in variable selection has focused on independently distributed data, but methods are emerging for time-series data [18], spatial lattice data [21,10,20], and geostatistical data [19,3].

The purpose of this paper is to develop a regularization approach to variable selection for Poisson point processes with regression. Although there are different possible approaches to variable selection, we find a regularization approach to be desirable. The *oracle* properties provide that asymptotically the correct subset of covariates is chosen and that the penalized estimators have the same asymptotic variances as if the correct subset of covariates was known *a priori*. The adaptive Lasso is one such method that has been shown to contain these oracle properties in linear regression for independent data [22] and dependent data [18,20,3]. Here, we provide empirical evidence via simulation that the oracle properties hold for the Poisson point process with regression as well.

The remainder of the paper is organized as follows. Section 2 gives an introduction to the Poisson point process model. Section 3 describes our method for variable selection and parameter estimation in the Poisson point process model. Section 4 presents the results of a simulation study for our method, followed by an application to the BCI forest census data set in Section 5. Section 6 proposes a variant of our method for variable selection and parameter estimation that allows for multiple pixel resolutions of covariates. Section 7 gives some concluding remarks.

2. Poisson point process

2.1. Poisson point process model

Let $(\Omega, \mathcal{A}, P)$ denote a probability space and $D \subset \mathbb{R}^d$ denote a spatial domain of interest in the d-dimensional space. Here, 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*. Locally finite configurations are those realizations $y \subset D$ such that $y \cap A$ is finite for every bounded Borel set $A \subset D$. In other words, y is a finite set of spatial coordinates in 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 *point process on* D [8].

Let $s_1, \ldots, s_n$ denote the observed spatial point pattern data comprising a set of n locations of events in D. Statistical models for spatial point pattern data treat the locations of events as a realization of a spatial point process Y on D. The first-order moment measure, or *intensity measure*, μ, is defined as

$$\mu(B) = E\{N(B)\} = E\left\{\sum_{\xi \in Y} 1(\xi \in B)\right\},$$

where B is a bounded Borel set in D. Often it is assumed that there exists an *intensity function* $\lambda(s)$ such that $\mu(B)$ is the integral of λ over the set B with respect to the Lebesgue measure

$$\mu(B) = \int_B \lambda(s)ds.$$

A fundamental statistical model for spatial point pattern data is the *Poisson point process* with intensity function λ defined by the following two conditions:

1. any bounded Borel set B in D has $\mu(B) \in (0, \infty)$ and $N(B) \sim \text{Poisson}(\mu(B))$.
2. the joint density f of the event locations $s_1, \ldots, s_n$ is proportional to the product of the intensity functions $\lambda(s_i)$, conditional on $N(B)$: $f(s_1, \ldots, s_n | N(B) = n) \propto \prod_{i=1}^n \lambda(s_i)$.

Condition 1 says that the number of events in B follows a Poisson distribution with an average number of events $\mu(B)$. Condition 2 says that, conditional on the number of events, the locations of events are independent, and the probability that an event occurs in a small region around a point s with area ds is $\lambda(s)ds$. Thus, events are more likely to occur in the regions of D with higher intensity values λ. See Fig. 1 for an example. For this model, a primary interest lies in estimating λ.

Let $x(s)$ denote a $p \times 1$ vector of covariates at location $s \in D$. The intensity function $\lambda(s)$ can be used to model the relationship between locations of events and covariates. One commonly-used model is the log–linear specification

$$\lambda(s; \beta) = \exp\{\beta_0 + x(s)^T \tilde{\beta}\},$$

where β_0 is the intercept, $\tilde{\beta}$ is a $p \times 1$ vector of regression coefficients, and $\beta = (\beta_0, \tilde{\beta}^T)^T$.

2.2. Maximum likelihood estimation

From the definition of the Poisson point process above, the log-likelihood function of a Poisson point process can be derived as follows. Let $g(s_1, \ldots, s_n, n)$ denote the joint density function of the event locations and the number of events. Let $N = N(D)$ and $P(N = n)$ denote the probability mass function of the number of events. Also, define $\theta = \int_D \lambda(s; \beta)ds$. It follows that

$$g(s_1, \ldots, s_n, n) = f(s_1, \ldots, s_n | N = n)P(N = n) = \left\{\prod_{i=1}^n \theta^{-1}\lambda(s_i; \beta)\right\} \frac{\exp(-\theta)\theta^n}{n!}.$$

After the simplification of this expression, the log-likelihood function of β, up to an additive constant, is defined to be

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

[8]. Maximizing $\ell(\beta)$ gives the maximum likelihood estimator (MLE) of β, which is denoted by $\hat{\beta}_{\text{MLE}}$.

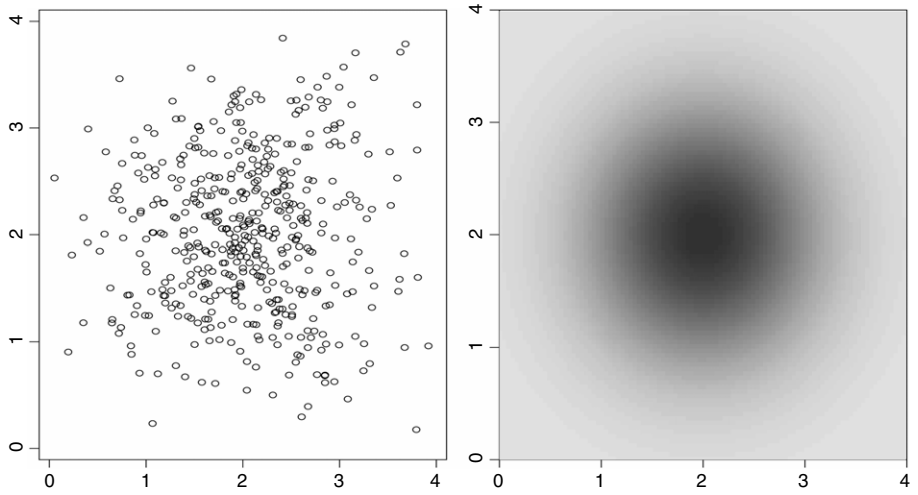

Fig. 1. Example of a Poisson point process—spatial point pattern data generated from a Poisson point process (left) and the intensity function of the process (right). Darker colors indicate larger values. Note that events occur more frequently where the intensity function is larger.

To maximize (1), we implement a quadrature scheme developed by Baddeley and Turner [1]. First, the spatial domain is partitioned into M rectangular pixels of equal area, denoted by a. Then one dummy point is placed in the center of each pixel. Let $(s_i, \Delta_i),\ i = 1, \ldots, n + M$ denote the pair such that s_i is a spatial location and Δ_i indicates whether the point is an event of the spatial point process ($\Delta_i = 1$) or a dummy point ($\Delta_i = 0$). Without loss of generality, assume that $s_1, \ldots, s_n$ are the observed events as before, and $s_{n+1}, \ldots, s_{n+M}$ are the dummy points. The weights ascribed to each point are $w_i = n_i^{-1}a$, where n_i is the number of events and dummy points in the same pixel as point $i, i = 1, \ldots, n + M$. A Riemann sum approximation $\int_D \lambda(s; \beta)ds \approx \sum_{i=1}^{n+M} w_i\lambda(s_i; \beta)$ is substituted into the log-likelihood function. Thus, the log-likelihood of β in (1) can be approximated by

$$\ell(\beta) \approx \sum_{i=1}^{n+M} w_i\Big\{w_i^{-1}\Delta_i \log\lambda(s_i; \beta) - \lambda(s_i; \beta)\Big\}. \tag{2}$$

Berman and Turner [2] showed that maximizing (2) is equivalent to fitting a weighted Poisson generalized linear model, which can be performed using standard statistical software.

3. Regularization method for variable selection and parameter estimation

3.1. Penalized maximum likelihood

In general, a regularization method attempts to minimize an objective function of the form $-\ell(\beta)+np_\gamma(\beta)$, where $\ell(\beta)$ is the log-likelihood function for some model of interest, n is the number of observations, and $p_\gamma(\beta)$ is a nonnegative penalty function parametrized by γ. A Lasso penalty is $p_\gamma(\beta) = \gamma\sum_{j=1}^{p}|\beta_j|$ [17]. Here, this penalty function is not differentiable at $\beta = 0$ and produces solutions that contain some estimates that are exactly 0. Thus, variable selection and parameter estimation can be obtained simultaneously. The adaptive Lasso penalty function $p_\gamma(\beta) = \sum_{j=1}^{p}\gamma_j|\beta_j|$ improves the selection properties of the Lasso by allowing a different tuning parameter for each regression coefficient and thus providing more flexibility.

Let $\ell(\beta)$ be the log-likelihood function in (1) and γ_j be a nonnegative tuning parameter in the adaptive Lasso penalty, $j = 1, \ldots, p$. For the Poisson point process model, we define the penalized log-likelihood function of β as

$$\ell_{\mathrm{p}}(\beta) = -\ell(\beta) + n\sum_{j=1}^{p} \gamma_j |\beta_j|. \tag{3}$$

The value that minimizes $\ell_{\mathrm{p}}(\beta)$ is a penalized maximum likelihood estimate of β.

3.2. Computational algorithm

To minimize $\ell_{\mathrm{p}}(\beta)$ in (3), we adopt an iterative algorithm. At step 0, the MLE of β is used as an initial value, $\hat{\beta}^{(0)} = \hat{\beta}_{\mathrm{MLE}}$. At step m, we use a Laplace approximation of $\ell(\beta)$,

$$\ell^*(\beta) = (\beta - \hat{\beta}^{(m-1)})^T \frac{\partial \ell(\hat{\beta}^{(m-1)})}{\partial \beta} + (1/2)(\beta - \hat{\beta}^{(m-1)})^T \frac{\partial^2 \ell(\hat{\beta}^{(m-1)})}{\partial \beta \partial \beta^T} (\beta - \hat{\beta}^{(m-1)}), \tag{4}$$

where $\hat{\beta}^{(m-1)}$ is attained at step $m-1$, $m = 1, 2, \ldots$.

Next, we rearrange the terms of $\ell^*(\beta)$. Define $y^* = (A^{-1})^T \left\{ \frac{\partial \ell(\hat{\beta}^{(m-1)})}{\partial \beta} - \frac{\partial^2 \ell(\hat{\beta}^{(m-1)})}{\partial \beta \partial \beta^T} \hat{\beta}^{(m-1)} \right\}$, $X^* = A\mathrm{diag}\{\gamma_j^{-1}\}_{j=1}^p$, and $\beta^* = \mathrm{diag}\{\gamma_j\}_{j=1}^p \beta$, where A is the Cholesky factor of $\frac{\partial^2 \ell(\hat{\beta}^{(m-1)})}{\partial \beta \partial \beta^T}$; that is, $-\frac{\partial^2 \ell(\hat{\beta}^{(m-1)})}{\partial \beta \partial \beta^T} = A^T A$. Then the approximation $\ell^*(\beta)$ in (4) can be rewritten in a quadratic form

$$\ell^*(\beta) = -(1/2)(y^* - X^*\beta)^T (y^* - X^*\beta). \tag{5}$$

The intercept β_0 is unpenalized, and we use the profile intercept estimate $\hat{\beta}_0^{(0)}$ as $\hat{\beta}_0^{(m-1)}$ for every m. Moreover, we set $\gamma_j = \gamma \log(n)(n|\hat{\beta}_j^{(0)}|)^{-1}$, $j = 1, \ldots, p$, such that γ is a common tuning parameter, where recall that $\hat{\beta}^{(0)} = \hat{\beta}_{\mathrm{MLE}}$ [22]. Let $\hat{\beta}(\gamma)$ denote the minimizer of $\ell_{\mathrm{p}}^*(\beta) = -\ell^*(\beta) + n\sum_{j=1}^p \gamma_j |\beta_j|$, which can be obtained via the well-known least angle regression (LARS) algorithm due to (5) [6].

To select the tuning parameter γ, the Bayesian information criterion (BIC) in this context is defined to be

$$\mathrm{BIC}(\gamma) = -2\ell(\hat{\beta}(\gamma); \gamma) + e(\gamma)\log(n),$$

where $e(\gamma) = \sum_{j=1}^p I\{\hat{\beta}_j(\gamma) \neq 0\}$ is the number of nonzero regression coefficient estimates. We fix a path of $\gamma \geq 0$ and select the tuning parameter γ and the estimate $\hat{\beta}^{(m)}$ that minimize $\mathrm{BIC}(\gamma)$. Then replace $\hat{\beta}^{(m-1)}$ by $\hat{\beta}^{(m)}$, $m = 1, 2, \ldots$, in the Laplace approximation of $\ell(\beta)$ in (4), and iterate this procedure until some convergence criterion is met. Denote this final estimate by $\hat{\beta}$. The Laplace approximation of $\ell(\beta)$ and iterative updating by the LARS algorithm were previously studied for spatial lattice data in [20].

3.3. Standard error

We approximate the standard errors of the regression coefficient estimates using a plug-in estimate of the asymptotic covariance matrix given in [16]. Without loss of generality, suppose that the covariates are rearranged, so that we can write $\hat{\beta} = \left((\hat{\beta}^1)^T, (\hat{\beta}^2)^T\right)^T$, where $\hat{\beta}^1$ is the vector of nonzero regression coefficient estimates and $\hat{\beta}^2 = 0$. Let $x^1(s_i)$ denote the column vector of covariates corresponding to $\hat{\beta}^1$ at location s_i. Let W be an $(n+M)\times(n+M)$ diagonal matrix with ith diagonal entry $w_i \exp\{x^1(s_i)^T \hat{\beta}^1\}$, where w_i is the quadrature weight. Let $X = \left[x^1(s_1), \ldots, x^1(s_{n+M})\right]^T$. Assuming that the true model only depends on $x^1(\cdot)$, the observed Fisher information is $-\frac{\partial^2 \ell(\beta^1)}{\partial \beta^1 (\partial \beta^1)^T}$. We substitute $\hat{\beta}^1$ for β^1 and compute $-\frac{\partial^2 \ell(\hat{\beta}^1)}{\partial \beta^1 (\partial \beta^1)^T} = \left(X^T W X\right)^{-1}$ to obtain an estimate of the covariance matrix of $\hat{\beta}^1$.

Table 1
Simulation results—average number of events in the spatial point pattern and proportion of samples where $\hat{\beta}_j$ was nonzero, $j = 1, 2, \ldots, 7$, using LARS_1, LARS_2, LARS_m, and Oracle, for spatial domains $[0, 10]^2$, $[0, 20]^2$, $[0, 30]^2$.

D	No. events	Procedure	β_1	β_2	β_3	β_4	β_5	β_6	β_7
$[0, 10]^2$	132.22	LARS_1	0.96	0.69	0.60	0.38	0.15	0.14	0.45
		LARS_2	0.93	0.68	0.60	0.35	0.10	0.12	0.38
		LARS_m	0.93	0.68	0.60	0.35	0.08	0.11	0.38
		Oracle	1.00	1.00	1.00	1.00	0.00	0.00	0.00
$[0, 20]^2$	1119.78	LARS_1	1.00	1.00	1.00	0.86	0.10	0.06	0.02
		LARS_2	1.00	1.00	1.00	0.85	0.07	0.04	0.02
		LARS_m	1.00	1.00	1.00	0.84	0.07	0.04	0.02
		Oracle	1.00	1.00	1.00	1.00	0.00	0.00	0.00
$[0, 30]^2$	1883.57	LARS_1	1.00	1.00	1.00	1.00	0.05	0.04	0.01
		LARS_2	1.00	1.00	1.00	1.00	0.05	0.03	0.01
		LARS_m	1.00	1.00	1.00	1.00	0.05	0.03	0.01
		Oracle	1.00	1.00	1.00	1.00	0.00	0.00	0.00

4. Simulation study

An increasing spatial domain asymptotic framework was used to assess the performance of our method. The spatial domains were $[0, 10]^2$, $[0, 20]^2$, and $[0, 30]^2$. Seven independent Gaussian random fields with exponential covariance functions $\mathcal{C}(h) = \sigma^2 \exp(-\phi^{-1}h)$ were generated as covariates $z(\cdot)$ with a 1×1 unit pixel resolution, where h is the Euclidean distance between two locations in D and σ^2 and ϕ are parameters controlling the variance and range of spatial correlation between locations. The parameters were set to $\sigma^2 = 0.05$ and $\phi = 0.5$. We used Gaussian random fields to build spatial correlation into the covariates, modeling likely ecological factors such as temperature or soil nutrient content. To induce correlation between the covariates, we defined Σ such that $(\Sigma)_{jj'} = 0.5^{|j-j'|}$ and transformed the vector of independent covariates $z(s)$ to the vector of correlated covariates $x(s) = B^T z(s)$, where $\Sigma = B^T B$. We simulated spatial point patterns from a Poisson point process with the log–linear intensity function specification $\lambda(s; \beta) = \exp\{\beta_0 + x(s)^T \tilde{\beta}\}$, with $\beta_0 = 0$ and $\tilde{\beta} = (4, 3, 2, 1, 0, 0, 0)^T$.

Four fitting procedures were compared. LARS_m updates the estimates until convergence. In the case of a large data set, LARS_m may be computationally prohibitive. We also investigate LARS_1, which takes $\hat{\beta}^{(1)}$ after one iteration as the parameter estimate, and LARS_2, which takes $\hat{\beta}^{(2)}$ after two iterations. Also, as a benchmark for comparing these LARS procedures, the oracle fits the true model with the first four covariates. Each of the procedures was run on each of 100 simulated spatial point patterns.

Table 1 shows the selection properties of our proposed method. For each spatial domain of the simulated data, the average sample size of the 100 simulated spatial point patterns is given. Then, we computed the relative frequency of nonzero estimates for each regression coefficient and did so for each fitting procedure. In the oracle case of no errors in selecting the covariates, the first four columns should be ones and the last three columns should be zeros.

On the smallest $[0, 10]^2$ spatial domain, the average number of events per pattern is about 132. The LARS_1 procedure selects the most nonzero regression coefficients, but the LARS_2 and LARS_m procedures correctly identify more of the zero regression coefficients. On the $[0, 20]^2$ spatial domain, there are on average about 1120 events per pattern. All three LARS procedures seem to select the correct covariates on average as their values are close to one for the first four covariates and zero for the last three covariates. On the largest $[0, 30]^2$ spatial domain, with about 1884 points per pattern, the selection is the most accurate. All three LARS procedures select the first four nonzero covariates correctly. The last three covariates are incorrectly selected in no more than five of the 100 simulated patterns. Table 1 also suggests that more iterations in the LARS procedure tend to select more zero regression coefficients. Furthermore, the LARS procedures tend to result in less error for nonzero regression coefficients with larger magnitudes. For example, on the $[0, 10]^2$ spatial domain, the first covariate is correctly selected more often than the second covariate, the second is correctly

Table 2
Simulation results—bias, variance (Var), and mean squared error (MSE) of $\hat{\beta}_j$ for $j = 1, 2, 3, 4$, using LARS_1, LARS_2, LARS_m, and Oracle, for spatial domains $[0, 10]^2$, $[0, 20]^2$, $[0, 30]^2$.

D		$[0, 10]^2$			$[0, 20]^2$			$[0, 30]^2$		
Coefficient	Procedure	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
$\beta_1 = 4$	LARS_1	−0.367	3.515	3.649	−0.116	0.107	0.120	−0.030	0.037	0.038
	LARS_2	−0.194	3.821	3.859	−0.087	0.101	0.109	−0.009	0.038	0.038
	LARS_m	−0.158	3.768	3.793	−0.086	0.098	0.105	−0.009	0.038	0.038
	Oracle	−0.007	1.522	1.522	−0.055	0.089	0.092	0.002	0.029	0.029
$\beta_2 = 3$	LARS_1	−0.357	4.565	4.693	0.120	0.218	0.233	0.039	0.043	0.045
	LARS_2	−0.115	5.146	5.159	0.082	0.188	0.195	0.016	0.042	0.042
	LARS_m	−0.116	5.192	5.205	0.080	0.184	0.191	0.015	0.042	0.042
	Oracle	−0.010	2.335	2.336	0.051	0.165	0.167	−0.003	0.039	0.039
$\beta_3 = 2$	LARS_1	−0.278	2.770	2.848	0.012	0.109	0.109	0.000	0.032	0.032
	LARS_2	−0.084	3.438	3.445	0.019	0.114	0.114	0.005	0.032	0.032
	LARS_m	−0.034	3.553	3.554	0.020	0.113	0.114	0.006	0.032	0.032
	Oracle	−0.008	1.665	1.665	−0.011	0.059	0.060	0.001	0.022	0.022
$\beta_4 = 1$	LARS_1	−0.264	1.530	1.600	−0.231	0.168	0.222	−0.152	0.052	0.075
	LARS_2	−0.187	1.899	1.934	−0.094	0.209	0.218	−0.044	0.045	0.047
	LARS_m	−0.180	1.922	1.954	−0.084	0.221	0.228	−0.041	0.045	0.047
	Oracle	0.066	1.290	1.294	0.028	0.068	0.069	−0.022	0.036	0.037

selected more often than the third, and the third is correctly selected more often than the fourth. These comparisons correspond to the ordering of the magnitudes of the true regression coefficients of the covariates.

Table 2 shows the properties of the estimates of nonzero regression coefficients by providing the biases, variances, and mean squared errors (MSE) for the regression coefficients β_1, β_2, β_3, and β_4 over the three spatial domains $[0, 10]^2$, $[0, 20]^2$, and $[0, 30]^2$ under the four fitting procedures.

On the smallest $[0, 10]^2$ spatial domain, the estimators by the LARS procedures can be quite negatively biased compared against the oracle estimator. Also, the LARS_1 estimates seem to be more biased than the LARS_2 and LARS_m estimates, whereas the LARS_2 and LARS_m estimates seem to have similar bias values. The variance and MSE of the LARS procedures are larger than those for the oracle estimator, and among the LARS procedures, the LARS_1 estimates seem to contain lower variance and MSE values than the LARS_2 and LARS_m procedures. On the $[0, 20]^2$ spatial domain, although the LARS estimators still have more bias than the oracle estimator, the difference in bias between LARS and oracle is much smaller than in the $[0, 10]^2$ spatial domain. Again, the LARS_1 procedure seems to produce more biased estimates than the LARS_2 or LARS_m procedures. The variance and MSE values for the LARS procedures are larger than the oracle procedures, but they are more similar to the oracle procedure than in the $[0, 10]^2$ spatial domain. Among the LARS procedures, it seems that the variance and MSE values are similar. On the largest $[0, 30]^2$ spatial domain, the bias, variance, and MSE values are similar and are small for all procedures. Generally, it seems that our method produces estimators that are relatively stable on the spatial domains with over 1000 events, but there can be substantial bias in smaller samples, especially for the one-step procedure, LARS_1.

Finally, the simulation results provide evidence that the oracle properties of the regularization estimators are likely to hold. In Table 1, as the spatial domain grows, the variable selection improves, and in Table 2, the variance and MSE values for the LARS estimates tend to the values given by the oracle procedure.

Overall, the results of the simulation study suggest that this regularization method is adequate for variable selection and estimation under reasonable conditions. On the $[0, 10]^2$ spatial domain, where the average number of events in each spatial point pattern was small, the method underselected the first four covariates, overselected the last three covariates, and its estimates had larger bias and variance values than the oracle procedure. On both of the larger spatial domains, where the average number of events was greater than 1000, the correct covariates were selected more frequently, and the

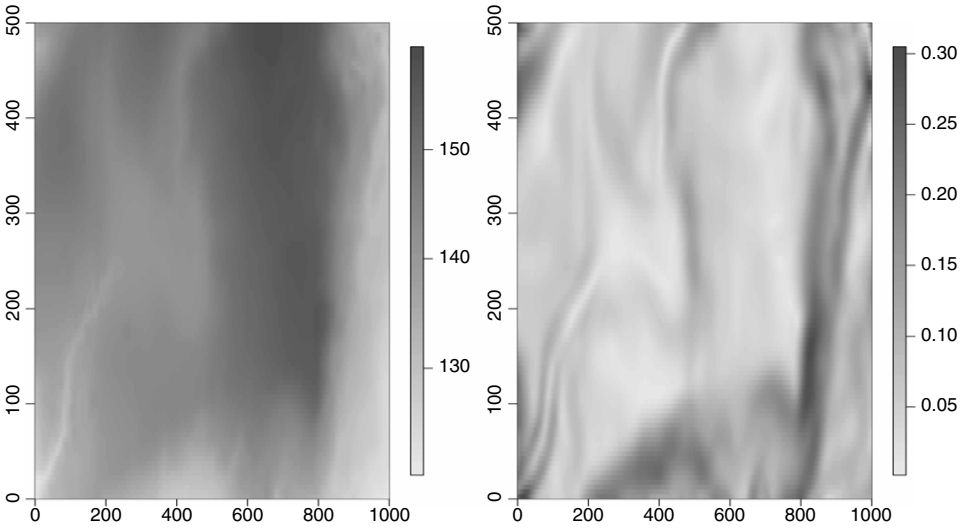

Fig. 2. Barro Colorado Island data—elevation (left) and slope (right). Darker colors indicate larger values.

bias and variance values were comparable to the oracle procedure. Also, the selection and estimation performance improved as the magnitude of the regression coefficients increased.

5. Application

Here we focus our analysis on the locations of 4026 *Beilschmiedia pendula* (*B. pendula*) tree stems in a 50 hectare (ha) (500 m × 1000 m) study area. In addition to the locations of various tree species, information is available on elevation and slope at a 5 m × 5 m pixel resolution. We model the intensity of the *B. pendula* trees, $\lambda(s;\beta)$, as a log–linear function of elevation, slope, and the incidences of six other tree species

$$\lambda(s;\beta) = \exp\left\{\beta_0 + \sum_{j=1}^{6} \beta_j x_j(s) + \beta_7 \text{elevation}(s) + \beta_8 \text{slope}(s)\right\},$$

where $x_j(s) = I$ (tree species j at location s) for $j = 1, \ldots, 6$. The six tree species are *Eugenia nesiotica* (*E. nesiotica*), *Eugenia oerstediana* (*E. oerstediana*), *Piper cordulatum* (*P. cordulatum*), *Protium panamense* (*P. panamense*), *Sorocea affinis* (*S. affinis*), and *Talisia croatii* (*T. croatii*).

Fig. 2 contains plots of elevation and slope, and Fig. 3 contains plots of the locations of *B. pendula* trees. As an example, we show in Fig. 3 the number of *P. panamense* in 5 m × 5 m, 10 m × 10 m, and 20 m × 20 m pixels.

Comparing the covariates in Fig. 2 to the *B. pendula* trees plotted in Fig. 3, these trees appear in greater abundance in areas with more moderate elevation and steeper slope. Fig. 3 seems to show that *B. pendula* and *P. panamense* trees tend to grow in similar regions across the study area.

We implemented our regularization method in Section 3 to select covariates computed at 5 m × 5 m, 10 m × 10 m, and 20 m × 20 m pixel resolutions. The values of elevation and slope on the 10 m × 10 m and 20 m × 20 m pixel resolutions are pixel averages. The estimates of the regression coefficients and their standard errors for these implementations are given in Table 3.

The results show that elevation, slope, *P. cordulatum*, and *P. panamense* are consistently selected regardless of pixel resolution, while *E. nesiotica* is not selected at any pixel resolution. The *B. pendula* trees seem to be attracted to *P. cordulatum*, *P. panamense*, *S. affinis*, and areas of increasing elevation and slope but are repelled by *T. croatii* and *E. oerstediana*.

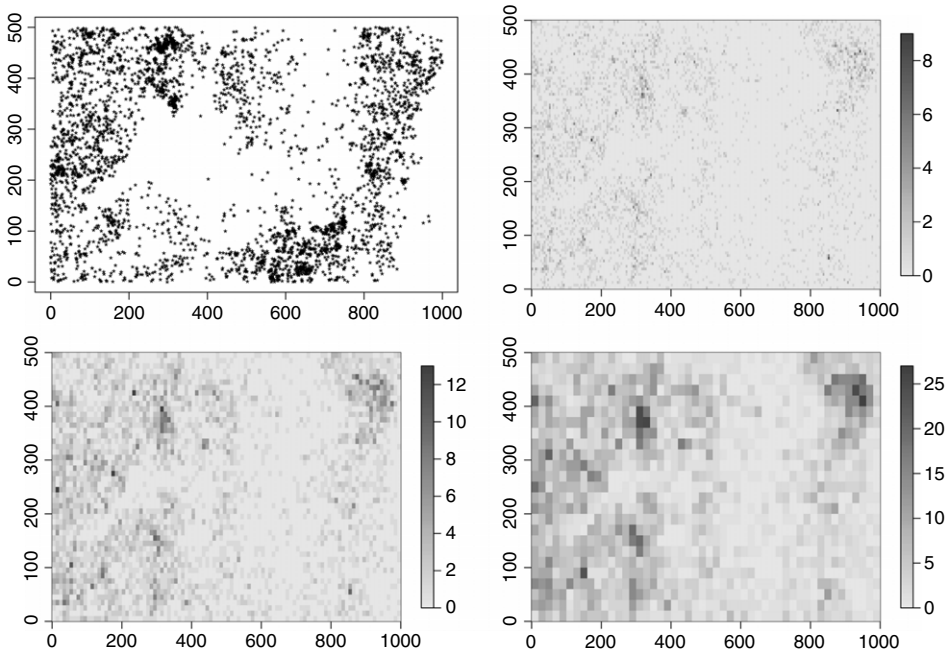

Fig. 3. Barro Colorado Island data—locations of Beilschmiedia pendula trees (upper left) and number of Protium panamense trees at 5 m × 5 m (upper right), 10 m × 10 m (lower left), and 20 m × 20 m (lower right) pixel resolutions. Darker colors indicate larger values.

Table 3
Barro Colorado Island data analysis—estimates of regression coefficients and standard errors (SE) for the 5 m × 5 m, 10 m × 10 m, and 20 m × 20 m pixel resolutions, *one pixel resolution at a time*.

Coefficient	Pixel resolution		
	5 m × 5 m	10 m × 10 m	20 m × 20 m
Intercept	−9.24	−10.12	−11.05
SE	0.33	0.35	0.37
Eugenia nesiotica	–	–	–
SE	–	–	–
Eugenia oerstediana	–	−6.57	−9.26
SE	–	1.41	2.10
Piper cordulatum	3.28	8.04	15.94
SE	0.31	0.47	0.85
Protium panamense	8.15	14.40	24.24
SE	0.42	0.78	1.22
Sorocea affinis	–	5.48	9.18
SE	–	1.32	2.19
Talisia croatii	–	−19.55	−39.67
SE	–	3.56	6.51
Elevation	0.026	0.031	0.036
SE	0.002	0.002	0.003
Slope	6.29	6.55	6.87
SE	0.24	0.24	0.26

6. Selection of pixel resolution

In the method above, the choice of pixel resolution for computing covariates is pre-selected, as a single pixel resolution is considered each time. The set of selected covariates may not be consistent

Table 4
Barro Colorado Island data analysis—estimates of regression coefficients and standard errors (SE) for the 5 m × 5 m, 10 m × 10 m, and 20 m × 20 m pixel resolutions, *three pixel resolutions simultaneously*.

Coefficient	Pixel resolution		
	5 m × 5 m	10 m × 10 m	20 m × 20 m
Intercept		−10.99	
SE		0.37	
Eugenia nesiotica	–	–	–
SE	–	–	–
Eugenia oerstediana	–	−6.55	–
SE	–	1.42	–
Piper cordulatum	–	1.78	14.14
SE	–	0.71	1.18
Protium panamense	3.30	–	20.49
SE	0.53	–	1.41
Sorocea affinis	–	–	9.31
SE	–	–	2.20
Talisia croatii	–	–	−41.08
SE	–	–	6.51
Elevation	0.035	–	–
SE	0.002	–	–
Slope	–	6.86	–
SE	–	0.25	–

across pixel resolutions. For example, three of the tree species selected at the 10 m × 10 m and 20 m × 20 m pixel resolutions were not selected at the 5 m × 5 m pixel resolution in Table 3. Here, we propose a variant of our method to compute covariate values at multiple pixel resolutions and simultaneously select covariates at different pixel resolutions.

Suppose that the spatial domain D is divided into pixels on K different grids $G_1, G_2, \ldots, G_K$. Let M_k denote the number of pixels in pixel resolution k, $k = 1, 2, \ldots, K$. We set $M_1 = d_1 d_2$, where d_1 and d_2 are integers, so that $G_1 = d_1 \times d_2$. Then we set $M_k = 2^{2(k-1)} d_1 d_2$, $k = 1, 2, \ldots, K$, so that $G_k = 2^{k-1} d_1 \times 2^{k-1} d_2$. To obtain covariate values at each pixel resolution, we start at the finest grid, G_K, and determine covariates as in Section 2. Then each pixel on a coarser grid G_k corresponds to a $2^{K-k} \times 2^{K-k}$ grid of pixels on G_K. We compute covariates on G_k by averaging the covariate values over the appropriate $2^{K-k} \times 2^{K-k}$ subset of pixels on G_K. Recall that in order to compute the Poisson point process log-likelihood function, all covariates are computed on the same number of pixels M. Here in order to impose this condition, we create new grids G_k^* of dimension $2^{K-1} d_1 \times 2^{K-1} d_2$ corresponding to grid G_k that contain blocks of $2^{K-k} \times 2^{K-k}$ constant values corresponding to those values obtained by averaging the values in G_K, $k = 1, 2, \ldots, K$. Note that $G_K^* = G_K$. In our variant method, the grids define K pixel resolutions for the p covariates, and the pK covariate grids are used for obtaining penalized estimates as described in Section 3. Consequently, this variant method provides a way to select each covariate on the pixel resolutions that have higher association with the incidence of events. Henceforth, we will refer to this variant as the multiple-resolution procedure.

In the Barro Colorado Island data, we set $d_1 = 25$, $d_2 = 50$, and $K = 3$. Thus, $G_1 = 25 \times 50$, $G_2 = 50 \times 100$, and $G_3 = 100 \times 200$. Each of the covariates is a 100×200 grid of values at all pixel resolutions. The covariate values on the 100 × 200 grid with the finest pixel resolution are computed using the quadrature scheme described in Section 2.2. At each coarser pixel resolution G_k, the covariate values from the 100 × 200 grid are averaged over $2^{3-k} \times 2^{3-k}$ subregions, $k = 1, 2$.

Table 4 contains estimates for the regression coefficients of the same covariates as before, except with covariates computed at three pixel resolutions simultaneously. The results show that all of the eight covariates are selected on exactly one pixel resolution except *P. cordulatum* and *P. panamense*, and again, *E. nesiotica* is not selected at any pixel resolution. It appears that *B. pendula* is attracted to *P. cordulatum* at the two coarsest pixel resolutions and *P. panamense* at the finest and coarsest pixel resolutions. Selecting covariates at different pixel resolutions appears to highlight the interesting differences in the varying range of attraction between *B. pendula* and the other tree species. Furthermore,

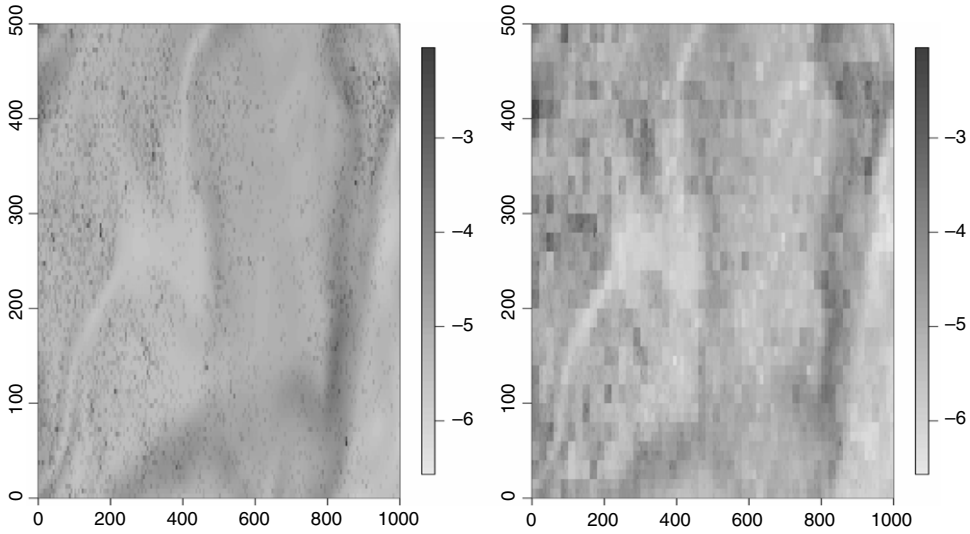

Fig. 4. Barro Colorado Island data analysis—estimated log-intensity functions at 5 m × 5 m pixel resolution using a single-resolution (left) and a multiple-resolution procedure (right). Darker colors indicate larger values.

the relationships inferred in Table 3 are retained in Table 4, with the advantage of not having to choose a particular pixel resolution *a priori*. In contrast, the fit of one pixel resolution at a time could potentially fail to select the important covariates that operate at a different pixel resolution.

Fig. 4 contains images of the fitted log-intensities as functions of the covariates selected in Tables 3 and 4, obtained by the single-resolution and multiple-resolution estimation procedures. The single-resolution image is from the estimation performed at the 5 m × 5 m pixel resolution. It appears that the multiple-resolution procedure allows more flexibility in how the covariates influence the intensity function of the event process. Each 5 m × 5 m pixel in the image corresponding to the single-resolution procedure contains a distinct value, whereas the image corresponding to the multiple-resolution procedure allows 10 m × 10 m and 20 m × 20 m pixels to have homogeneous intensity. Essentially, the single-resolution procedure is a special case of the multiple-resolution procedure: selection and estimation are performed at one fixed pixel resolution, and regression coefficients at the other pixel resolutions are set to zero.

7. Conclusion and discussion

In this paper, we have developed a regularization method for performing variable selection and parameter estimation in Poisson point process models with regression. Our simulation results indicate that, as the number of events increases, our method is more likely to select the correct set of covariates and obtain estimators that are closer to being as efficient as if we had fit the true model. We have also proposed a method to select covariates at different pixel resolutions. The Barro Colorado Island (BCI) data example has demonstrated that there can be less ambiguity inferring the effects of covariates by allowing different pixel resolutions simultaneously rather than fixing the pixel resolution in the covariates *a priori*.

Our case study is a data set containing a complete census of trees in the study area. More generally, other ecological events of interest could be considered. For example, a survey of statistical methods for point processes discusses applications where events of interest include the locations of minke whales and ants' nests [15]. Our method can also be applied in studies where only a random sample of events is collected. To estimate the intercept and hence obtain an estimate of abundance, an entire census is necessary, but estimates of the other regression coefficients remain valid.

A few practical issues should be considered when applying this methodology. Our method was developed specifically for the data that follow a Poisson point process, and the simulation study

above did not address what would happen if the independence assumption is violated. Extending regularization methods to point processes with dependent events will help to broaden the types of data for which the developed methodology applies.

In the BCI data example, the 5 m $\times$ 5 m pixel resolution images of covariates contain 20,301 pixels, and the point pattern data set contains 4026 events. The application of this method to data on this scale took three seconds on an Apple desktop computer with a 2.4 GHz Intel Core 2 Duo processor and 3 GB of RAM. Also, we simulated spatial point patterns with 10,000 and 20,000 random events to investigate how computational time scales with the size of the spatial point pattern data set. On the same computer, these two spatial point patterns took six and thirteen seconds, respectively.

As with other regression methods, the adequacy of inference depends on multiple factors. As we have shown in the simulation study, both selection and estimation properties are expected to improve with increasing area of the spatial domain, and the selection is expected to improve with increasing size of the effects of covariates. Additional factors, such as the spatial dependence range of each covariate, the cross-correlation among covariates, and the number of covariates, p, could potentially affect both the selection and estimation properties. While general sample size recommendations are challenging, we recommend exhibiting caution as when using other regression methods.

Here, we have considered the case where the number of events is at least as large as the number of covariates. A new algorithm will need to be developed in order to apply this method to data where the number of events is smaller than the number of covariates. Likewise, multicollinearity is another issue that may need to be addressed as in many regression analyses, by using methods like principle component analysis or elastic net [23]. Finally, providing the theoretical properties of the penalized maximum likelihood estimator given in Section 3 can help improve our understanding of the method and its applicability. We leave these for future research.

Acknowledgments

Funding has been provided for this research from USDA Cooperative State Research, Education and Extension Service (CSREES) Hatch and McIntire-Stennis projects. We thank the editor, reviewers, and Professor Yongtao Guan for constructive comments that helped improve the paper. The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.

References

[1] A. Baddeley, R. Turner, Practical maximum pseudo-likelihood for spatial point patterns (with discussion), Australian and New Zealand Journal of Statistics 42 (2000) 283–322.
[2] M. Berman, T.R. Turner, Approximating point process likelihoods with GLIM, Journal of the Royal Statistical Society. Series C 41 (1992) 31–38.
[3] T. Chu, J. Zhu, H. Wang, Penalized maximum likelihood estimation and variable selection in geostatistics, Annals of Statistics 39 (2011) 2607–2625.
[4] R. Condit, Tropical Forest Census Plots, Springer-Verlag and R.G. Landes Company, Berlin, Germany, and Georgetown, Texas, 1998.
[5] N. Draper, H. Smith, Applied Regression Analysis, Wiley, New York, 1998.
[6] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Annals of Statistics 32 (2004) 407–451.
[7] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, Journal of the American Statistical Association 96 (2001) 1348–1360.
[8] C. Gaetan, X. Guyon, Spatial Statistics and Modeling, Springer, New York, 2010.
[9] T. Hesterberg, N.H. Choi, L. Meier, C. Fraley, Least angle and l_1 penalized regression: a review, Statistics Surveys 2 (2008) 61–93.
[10] H.-C. Huang, N.-J. Hsu, D. Theobald, F.J. Breidt, Spatial Lasso with applications to GIS model selection, Journal of Computational and Graphical Statistics 19 (2010) 963–983.

[11] S.P. Hubbell, R. Condit, R.B. Foster, 2005. Barro Colorado Forest Census Plot Data. URL http://ctfs.arnarb.harvard.edu/webatlas/datasets/bci.
[12] S.P. Hubbell, R.B. Foster, S.T. O'Brien, K.E. Harms, R. Condit, B. Wechsler, S.J. Wright, S. Loo de Lao, Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest, Science 283 (1999) 554–557.
[13] I.A. Ibragimov, R.Z. Has'minskii, Statistical Estimation: Asymptotic Theory, Springer, New York, 1981.
[14] J. Møller, R. Waagepetersen, Statistical Inference and Simulation for Spatial Point Processes, Chapman & Hall, Boca Raton, 2004.
[15] J. Møller, R. Waagepetersen, Modern statistics for spatial point processes, Scandinavian Journal of Statistics 34 (2007) 643–684.
[16] S. Rathbun, N. Cressie, Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson processes, Advances in Applied Probability 26 (1994) 122–154.
[17] R. Tibshirani, Regression shrinkage and selection via the Lasso, Journal of the Royal Statistical Society. Series B 58 (1996) 267–288.
[18] H. Wang, G. Li, C.-L. Tsai, Regression coefficients and autoregressive order shrinkage and selection via the Lasso, Journal of the Royal Statistical Society. Series B 69 (2007) 63–78.
[19] H. Wang, J. Zhu, Variable selection in spatial regression via penalized least squares, Canadian Journal of Statistics 37 (2009) 607–624.
[20] J. Zhu, H.-C. Huang, P.E. Reyes, On selection of spatial linear models for lattice data, Journal of the Royal Statistical Society. Series B 72 (2010) 389–402.
[21] Z. Zhu, Y. Liu, Estimating spatial covariance using penalized likelihood with weighted L_1 penalty, Journal of Nonparametric Statistics 21 (2009) 925–942.
[22] H. Zou, The adaptive Lasso and its oracle properties, Journal of the American Statistical Association 101 (2006) 1418–1429.
[23] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society. Series B 67 (2005) 301–320.
[24] H. Zou, R. Li, One-step sparse estimates in nonconcave penalized likelihood models, The Annals of Statistics 36 (2008) 1509–1533. With discussion.