

Selection of Spatial-Temporal Lattice Models: Assessing the Impact of Climate Conditions on a Mountain Pine Beetle Outbreak

Perla E. REYES, Jun ZHU, and Brian H. AUKEMA

Insects are among the most significant indicators of a changing climate. Here we evaluate the impact of temperature, precipitation, and elevation on the tree-killing ability of an eruptive species of bark beetle in pine forests of British Columbia, Canada. We consider a spatial-temporal linear regression model and in particular, a new statistical method that simultaneously performs model selection and parameter estimation. This approach is penalized maximum likelihood estimation under a spatial-temporal adaptive Lasso penalty, paired with a computationally efficient algorithm to obtain approximate penalized maximum likelihood estimates. A simulation study shows that finite-sample properties of these estimates are sound. In a case study, we apply this approach to identify the appropriate components of a general class of landscape models which features the factors that propagate an outbreak. We interpret the results from ecological perspectives and compare our method with alternative model selection procedures.

Key Words: Autoregressive models; Bark beetle; Lattice model; Model selection; Penalized maximum likelihood; Spatial-temporal process.

1. INTRODUCTION

Insects are among the most significant indicators of a changing climate, as temperature is inextricably linked to activities such as foraging and reproductive success in these cold-blooded organisms. Deviations from seasonal precipitation norms, such as summer drought, can stress host plants and facilitate access by insects to host pools that would otherwise be inaccessible (Raffa et al. 2008). In forest ecosystems, rapid increases in insect populations facilitated by suitable climatic conditions can result in dramatic range shifts of herbivorous insects, affect subsequent host utilization patterns, and alter the frequency and severity of natural disturbance regimes such as wildfire (Battistia et al. 2006; Jenkins et al. 2008; Bentz et al. 2010).

Perla E. Reyes (✉) is Assistant Professor, Department of Statistics, Kansas State University, Manhattan, KS, USA (E-mail: pereyesc@ksu.edu). Jun Zhu is Professor, Department of Statistics and Department of Entomology, University of Wisconsin, Madison, WI, USA. Brian H. Aukema is Assistant Professor, Department of Entomology, University of Minnesota, St. Paul, MN, USA.

© 2012 International Biometric Society
Journal of Agricultural, Biological, and Environmental Statistics, Volume 17, Number 3, Pages 508–525
DOI: [10.1007/s13253-012-0103-0](https://doi.org/10.1007/s13253-012-0103-0)

Bark beetles comprise a group of forest insects that reproduce within the stems of mature trees. While the majority of species colonize and cull weakened trees, some species can erupt into immense populations that can kill otherwise healthy hosts over large regions (Raffa et al. 2008). Currently, for example, there is an enormous outbreak of mountain pine beetle [*Dendroctonus ponderosae* Hopkins (Coleoptera: Scolytidae; alt.: Scolytinae: Curculionidae)] covering more than 17 million hectares of mature lodgepole pine forest in the provinces of British Columbia and Alberta, Canada. This outbreak has been exacerbated by two factors; first, an abundance of susceptible pine, caused in part by altered disturbance regimes such as fire suppression; and second, a changing climate. The outbreak, exerting carbon impacts at a scale in the order of mega tonnes (Kurz et al. 2008), has expanded with each annual generation of insects. In the past five years, the beetles have breached a historic geoclimatic barrier of the Rocky Mountains into northwestern Alberta and now threaten to expand their range across the boreal forest to the east coast of North America (Safranyik et al. 2010; De la Giroday et al. 2011).

The scientific questions of interest and importance such as the nature of factors that propagate outbreaks can be addressed via building a general class of landscape models and identifying the most appropriate components of the general model. These include the most important covariates and, with strong temporal and spatial dependence through ecological processes of reproduction and dispersal, suitable spatial, temporal, and/or spatial-temporal dependence patterns as well. To address these scientific questions, we consider a general modeling framework for spatial-temporal lattice data and a likelihood-based approach to statistical inference. In particular, we apply a new method for the simultaneous selection of covariates and spatial-temporal dependence structure.

Spatial lattice models are important tools for the analysis of spatial lattice data and have been applied to many disciplines (see, e.g., Schabenberger and Gotway 2005). However, development of spatial-temporal lattice models is not as advanced as spatial lattice models. Although it has been an area of active research, the modeling and inference framework is generally Bayesian hierarchical modeling which is flexible and powerful (see, e.g., Banerjee, Carlin, and Gelfand 2004). In our opinion, however, Bayesian inference tends to be computationally intensive (see, e.g., Zheng and Zhu 2008) and model selection is not always adequately addressed in spatial-temporal modeling, possibly because of the high computational cost. In this regard, we take an alternative, maximum likelihood-based approach for the selection of spatial-temporal lattice models.

Variable selection via penalized methods for standard linear regression has gained popularity in the last decade or so. Innovations include least absolute shrinkage and selection operator (Lasso) (Tibshirani 1996), adaptive Lasso (Zou 2006), and penalized least squares or maximum likelihood under nonconcave penalty (Fan and Li 2001). For efficient computation, Efron et al. (2004) proposed least angle regression (LARS) algorithms, whereas Zou and Li (2008) developed one-step sparse estimation for approximation of solutions. While most penalized methods assume independence, some are becoming available for dependent data such as time-series data (see, e.g., Wang, Li, and Tsai 2007a) and spatial lattice data (see, e.g., Huang et al. 2010; Zhu and Liu 2009; Zhu, Huang, and Reyes 2010).

For selection of spatial-temporal lattice models, however, the methodology available is limited. To the best of our knowledge, our work is the first to employ penalized meth-

ods to spatial-temporal lattice models for simultaneous selection of covariates and spatial-temporal dependence structures. Furthermore, we tailor our new model selection technique toward the analysis of MPB outbreak data and evaluate the impact of climatic variables on the outbreaks across space and over time. In particular, we utilize innovations such as LARS algorithms and one-step sparse estimation when devising our computational algorithms to ensure computational efficiency. In addition, we contrast the new method with alternative approaches taken by practitioners.

The remainder of the paper is organized as follows. In Section 2, we introduce a general class of spatial-temporal lattice models with a flexible parameterization of spatial-temporal neighborhood. In Section 3, we describe spatial-temporal Lasso for model selection and related computational issues. A simulation study is given in Section 4. A detailed description of the scientific background and data analysis using spatial-temporal Lasso is given in Section 5, as well as comparison against alternative approaches in practice. A summary is given in Section 6.

2. STATISTICAL MODEL

2.1. SPATIAL-TEMPORAL LATTICE MODEL

Let $D_I = \{s_1, \dots, s_I\} \subset \mathbb{R}^d$ denote a spatial grid consisting of I sites s_i , for $i = 1, \dots, I$, which are sometimes viewed as representatives of the grid cells that partition a spatial domain of interest. Let $y_{i,t} = y(s_i, t)$ denote the response variable at site $s_i \in D_I$ and time t , for $i = 1, \dots, I, t = 1, \dots, T$. Further, $\mathbf{x}_{i,t} = (x_{1,i,t}, \dots, x_{J,i,t})'$ is a J -dimensional vector of covariates at site s_i time t . Consider a linear regression model

$$y_{i,t} = \mathbf{x}'_{i,t} \boldsymbol{\beta} + \varepsilon_{i,t}, \quad (2.1)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)'$ is a J -dimensional vector of regression coefficients. We model the error term by a spatial-temporal autoregressive model (Section 6.8, Cressie 1993). In particular, let

$$\boldsymbol{\varepsilon}_t = \sum_{l=0}^L \mathbf{C}_l \boldsymbol{\varepsilon}_{t-l} + \mathbf{v}_t, \quad (2.2)$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{I,t})'$ denotes an I -dimensional vector of errors at time t for $t = 1, \dots, T$, $L \geq 0$ is a pre-specified maximum time lag, and \mathbf{C}_l for $l = 0, \dots, L$ are $I \times I$ matrices consisting of $c_{i,i'}^{(l)}$ with $i, i' = 1, \dots, I$. The matrices \mathbf{C}_0 and \mathbf{C}_l ($l > 0$) comprise autoregressive coefficients among spatial neighbors at the same time point and at two time points that are l time lags apart, respectively. Furthermore, $\mathbf{v}_t = (v_{1,t}, \dots, v_{I,t})' \sim \text{iid } N(\mathbf{0}, \sigma^2 \mathbf{I}_I)$ consists of iid noise with mean 0 and variance component σ^2 . Under this assumption, the error term in (2.1) follows a zero-mean Gaussian process with

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Gamma}), \quad (2.3)$$

where $\boldsymbol{\varepsilon} = (\varepsilon_{1,1}, \dots, \varepsilon_{I,T})'$ denotes an N -dimensional vector of errors, and $\boldsymbol{\Gamma}$ is an $N \times N$ covariance matrix consisting of $\text{cov}(\varepsilon_{i,t}, \varepsilon_{i',t'})$, for $i, i' = 1, \dots, I, t, t' = 1, \dots, T$, and $N = IT$.

There are different ways of setting the initial values of $\boldsymbol{\epsilon}_{1-l}$ for $l = 1, \dots, L$. One approach is to set them to zero $\boldsymbol{\epsilon}_{1-l} = \mathbf{0}$, but this would result in different equations (2.2) for the initial L time points and, thus, possible bias. Another approach that circumvents this boundary effect is to let $\boldsymbol{\epsilon}_{1-l} = \boldsymbol{\epsilon}_{T+1-l}$ for $l = 1, \dots, L$. It can be shown that, under the zero initial values for the error terms, the covariance matrix is

$$\boldsymbol{\Gamma} = \sigma^2(\mathbf{I}_N - \mathbf{C})^{-1}(\mathbf{I}_N - \mathbf{C}')^{-1}, \tag{2.4}$$

where \mathbf{C} is a block-lower-triangular matrix with \mathbf{C}_0 as the diagonal blocks and \mathbf{C}_l as the l th subdiagonal blocks for $l = 1, \dots, L$, whereas under the alternative initial values for the error terms, the covariance matrix is

$$\boldsymbol{\Gamma} = \sigma^2(\mathbf{I}_N - \mathbf{C} - \mathbf{G})^{-1}(\mathbf{I}_N - \mathbf{C}' - \mathbf{G}')^{-1}, \tag{2.5}$$

where $\mathbf{G} = \begin{pmatrix} \mathbf{G}_1 & \mathbf{G}_2 \\ \mathbf{G}_3 & \mathbf{G}_4 \end{pmatrix}$, \mathbf{G}_2 is an $IL \times IL$ upper-triangular matrix with \mathbf{C}_L as the diagonal blocks and \mathbf{C}_{L-l} as the l th subdiagonal blocks for $l = 1, \dots, L - 1$, and \mathbf{G}_1 , \mathbf{G}_3 and \mathbf{G}_4 are matrices of zero's with dimensions $IL \times I(T - L)$, $I(T - L) \times I(T - L)$, and $I(T - L) \times IL$, respectively.

2.2. SPATIAL-TEMPORAL NEIGHBORHOOD STRUCTURE

For a given site i , we let $\mathcal{N}(i)$ be its neighborhood and let $\mathcal{N}(i) = \bigcup_{k=1}^K \mathcal{N}_k(i)$, where $\{\mathcal{N}_k(i) : k = 1, \dots, K\}$ are neighborhoods that partition $\mathcal{N}(i)$, $i = 1, \dots, I$ (Zhu, Huang, and Reyes 2010). On a regular grid, the k th-order neighbors in $\mathcal{N}_k(i)$ of a given site i can be defined as the k th-nearest neighbors in terms of distance between two sites, for $k = 1, \dots, K$. For example, $\mathcal{N}_1(i)$ consists of the four nearest neighbors in the north, south, west, and east, $\mathcal{N}_2(i)$ consists of the four second-nearest neighbors in the northwest, north-east, southwest, and southeast, etc. The number of neighbors is not necessarily four at higher orders.

We consider the following parameterization for modeling spatial-temporal dependence:

$$\mathbf{C}_l = \sum_{k=0}^K \theta_{k,l} \mathbf{W}_{k,l}, \tag{2.6}$$

where $l = 0, \dots, L$, $\theta_{k,l}$ is an unknown spatial-temporal coefficient, and $\mathbf{W}_{k,l} = [w_{i,i'}^{k,l}]_{i,i'=1}^I$ is an $I \times I$ matrix consisting of pre-specified spatial-temporal weights for the k th-order neighborhood and l th-order time lag, where $k = 0, \dots, K$ and $l = 0, \dots, L$. We assume that the weights are symmetric in the sense that $w_{i,i'}^{k,l} = w_{i',i}^{k,l}$ for all $i' \neq i$; $k = 1, \dots, K$ and $l = 0, \dots, L$. We set $\theta_{0,0} \equiv 0$ and $\mathbf{W}_{0,l} \equiv \mathbf{I}_I$ for $l \geq 1$ in order that at time lag $l = 0$, $\mathbf{C}_0 = \sum_{k=1}^K \theta_{k,0} \mathbf{W}_{k,0}$ features spatial autocorrelation among neighbors via spatial-only coefficients $\theta_{k,0}$ for $k = 1, \dots, K$; and that at time lag $l \geq 1$, $\mathbf{C}_l = \theta_{0,l} \mathbf{I}_I + \sum_{k=1}^K \theta_{k,l} \mathbf{W}_{k,l}$ features spatial-temporal autocorrelation via temporal-only coefficients $\theta_{0,l}$ for $l = 1, \dots, L$ and spatial-temporal coefficients $\theta_{k,l}$ for $k \geq 1$ and $l \geq 1$.

The parameterization (2.6) is general and quite flexible. It features spatial-temporal interaction, as it allows autocorrelation between site i at time t and site i' at time t' , provided that sites i and i' are spatial neighbors and t and t' are within L time points apart. This

general form will be referred to as a space-time interaction model. Moreover, by constraining the individual spatial-temporal coefficients $\{\theta_{k,l}\}$, (2.6) gives rise to other simpler spatial-temporal models. For example, by letting the spatial-temporal coefficients $\theta_{k,l} = 0$ for $k \geq 1, l \geq 1$, the space-time interaction model in (2.6) is reduced to a space-time separable model, where temporal autocorrelation is allowed, but only for a same site. By setting $\theta_{k,l} = 0$ for $k \geq 1, l \geq 1$ and furthermore $\theta_{0,l} = 0$, spatial autocorrelation is allowed but there is no temporal autocorrelation.

3. STATISTICAL INFERENCE

3.1. SPATIAL-TEMPORAL LASSO

Let $\boldsymbol{\theta} = (\theta_{1,0}, \dots, \theta_{K,0}, \dots, \theta_{1,L}, \dots, \theta_{K,L}, \theta_{0,1}, \dots, \theta_{0,L})'$ denote an R -dimensional vector of spatial-temporal coefficients, where $R = (K + 1)(L + 1) - 1$. Henceforth, we replace the double index in $\theta_{k,l}$ with a single index θ_r , for $r = 1, \dots, R$, except where double indexing aids interpretation. Let $\boldsymbol{\gamma} = (\boldsymbol{\theta}', \sigma^2)'$, we sometimes use $\boldsymbol{\Gamma}_\boldsymbol{\gamma}$ to emphasize the parameterization of $\boldsymbol{\Gamma}$ by $\boldsymbol{\gamma}$. Let $\mathbf{y} = (y_{1,1}, \dots, y_{I,T})'$ denote an N -dimensional vector of response variables and let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_J]$ denote an $N \times J$ design matrix, where $\mathbf{x}_j = (x_{j,1,1}, \dots, x_{j,I,T})'$ denotes an N -dimensional vector of the j th covariate with $j = 1, \dots, J$. Thus, by (2.1) and (2.3),

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Gamma}_\boldsymbol{\gamma}). \quad (3.1)$$

We now consider selection of covariates and spatial-temporal dependence structure. For selection of covariates, our method will determine which regression coefficients are nonzero. For selection of a spatial-temporal dependence structure, we utilize the parameterization in (2.6) and determine which of the spatial-temporal coefficients are nonzero.

Let $\boldsymbol{\eta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ denote a $(J + R + 1)$ -dimensional vector of model parameters consisting of both regression coefficients and spatial-temporal coefficients. Under (3.1), the log-likelihood function is

$$\begin{aligned} \log L(\boldsymbol{\eta}; \mathbf{y}, \mathbf{X}) &= \text{const} - (1/2) \log |\boldsymbol{\Gamma}_\boldsymbol{\gamma}| - (1/2)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Gamma}_\boldsymbol{\gamma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &\equiv \text{const} + \ell(\boldsymbol{\eta}). \end{aligned}$$

We let $\hat{\boldsymbol{\eta}}_{\text{MLE}} = \arg \max_{\boldsymbol{\eta}} \ell(\boldsymbol{\eta})$ denote the maximum likelihood estimates (MLE) of $\boldsymbol{\eta}$.

We consider the following penalized log-likelihood function:

$$Q(\boldsymbol{\eta}) = \ell(\boldsymbol{\eta}) - N \sum_{j=1}^J \lambda_j |\beta_j| - N \sum_{r=1}^R \tau_r |\theta_r|, \quad (3.2)$$

where the last two terms are adaptive Lasso penalty on the coefficients, $\{\lambda_j\}_{j=1}^J$ are regularization parameters for the regression coefficients $\boldsymbol{\beta}$, and $\{\tau_r\}_{r=1}^R$ are regularization parameters for the spatial-temporal coefficients $\boldsymbol{\theta}$. We let $\hat{\boldsymbol{\eta}}_{\text{PMLE}} = \arg \max_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$ denote the penalized maximum likelihood estimates (PMLE) of $\boldsymbol{\eta}$.

Let $\hat{\boldsymbol{\eta}}^{(0)} = (\hat{\boldsymbol{\beta}}^{(0)'}, \hat{\boldsymbol{\gamma}}^{(0)'})'$ denote an initial value of $\boldsymbol{\eta}$, which is set to the MLE $\hat{\boldsymbol{\eta}}_{\text{MLE}}$. Given that $\boldsymbol{\eta} \approx \hat{\boldsymbol{\eta}}^{(0)}$, we approximate the penalized log-likelihood function (3.2) up to a constant by

$$Q^*(\boldsymbol{\eta}) = (\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}^{(0)})' \frac{\partial \ell(\hat{\boldsymbol{\eta}}^{(0)})}{\partial \boldsymbol{\eta}} - (1/2)(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}^{(0)})' \boldsymbol{I}(\hat{\boldsymbol{\eta}}^{(0)})(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}^{(0)}) - N \sum_{j=1}^J \lambda_j |\beta_j| - N \sum_{r=1}^R \tau_r |\theta_r|, \tag{3.3}$$

where $\boldsymbol{I}(\boldsymbol{\eta}) = E_{\boldsymbol{\eta}}\{-\frac{\partial^2 \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}\}$ is an expected information matrix (Zhu, Huang, and Reyes 2010). We propose to approximate $\hat{\boldsymbol{\eta}}_{\text{PMLE}}$ by

$$\hat{\boldsymbol{\eta}}^{(1)} = \arg \max_{\boldsymbol{\eta}} \{Q^*(\boldsymbol{\eta})\}. \tag{3.4}$$

Since the expected information matrix is block diagonal with $\boldsymbol{I}(\boldsymbol{\eta}) = \text{diag}\{\boldsymbol{I}(\boldsymbol{\beta}), \boldsymbol{I}(\boldsymbol{\gamma})\}$, we obtain $\hat{\boldsymbol{\beta}}^{(1)}$ and $\hat{\boldsymbol{\gamma}}^{(1)}$ separately. That is,

$$\hat{\boldsymbol{\beta}}^{(1)} = \arg \min_{\boldsymbol{\beta}} \left\{ -(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(0)})' \frac{\partial \ell(\hat{\boldsymbol{\eta}}^{(0)})}{\partial \boldsymbol{\beta}} + (1/2)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(0)})' \boldsymbol{I}(\hat{\boldsymbol{\beta}}^{(0)})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(0)}) + N \sum_{j=1}^J \lambda_j |\beta_j| \right\}. \tag{3.5}$$

It can be shown that the solution of (3.5) can be attained equivalently by

$$\hat{\boldsymbol{\beta}}^{*(1)} = \arg \min_{\boldsymbol{\beta}^*} \left\{ (1/2)(\mathbf{y}^* - \mathbf{X}^* \boldsymbol{\beta}^*)' (\mathbf{y}^* - \mathbf{X}^* \boldsymbol{\beta}^*) + N \sum_{j=1}^J |\beta_j^*| \right\}, \tag{3.6}$$

where $\mathbf{y}^* = (\mathbf{A}^{-1})' \{ \frac{\partial \ell(\hat{\boldsymbol{\eta}}^{(0)})}{\partial \boldsymbol{\beta}} + \boldsymbol{I}(\hat{\boldsymbol{\beta}}^{(0)})' \hat{\boldsymbol{\beta}}^{(0)} \}$, $\mathbf{X}^* = \mathbf{A} \text{diag}\{\lambda_j^{-1}\}_{j=1}^J$, $\boldsymbol{\beta}^* = \text{diag}\{\lambda_j\}_{j=1}^J \boldsymbol{\beta}$, and $\boldsymbol{I}(\hat{\boldsymbol{\beta}}^{(0)}) = \mathbf{A}' \mathbf{A}$. Hence, $\hat{\boldsymbol{\beta}}^{(1)} = \text{diag}\{\lambda_j^{-1}\}_{j=1}^J \hat{\boldsymbol{\beta}}^{*(1)}$.

Next,

$$\hat{\boldsymbol{\gamma}}^{(1)} = \arg \min_{\boldsymbol{\gamma}} \left\{ -(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}^{(0)})' \frac{\partial \ell(\hat{\boldsymbol{\eta}}^{(0)})}{\partial \boldsymbol{\gamma}} + (1/2)(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}^{(0)})' \boldsymbol{I}(\hat{\boldsymbol{\gamma}}^{(0)})(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}^{(0)}) + N \sum_{r=1}^R \tau_r |\theta_r| \right\}. \tag{3.7}$$

Given that σ^2 is not subject to any penalty, we let

$$\mathbf{X}_r^{**} = \tau_r^{-1} (\mathbf{B}_r - c_r \mathbf{B}_{R+1}), \quad r = 1, \dots, R, \quad \text{and} \quad \mathbf{X}_{R+1}^{**} = \mathbf{B}_{R+1},$$

where $c_r = \mathbf{B}'_{R+1} \mathbf{B}_r / \mathbf{B}'_{R+1} \mathbf{B}_{R+1}$, for $r = 1, \dots, R$, and $\boldsymbol{I}(\hat{\boldsymbol{\gamma}}^{(0)}) = \mathbf{B}' \mathbf{B}$. It follows that $\mathbf{X}_{R+1}^{**'} \mathbf{X}_r^{**} = \mathbf{0}$ for $r = 1, \dots, R$. Let $\mathbf{y}^{**} = (\mathbf{B}^{-1})' \{ \frac{\partial \ell(\hat{\boldsymbol{\eta}}^{(0)})}{\partial \boldsymbol{\gamma}} + \boldsymbol{I}(\hat{\boldsymbol{\gamma}}^{(0)})' \hat{\boldsymbol{\gamma}}^{(0)} \}$. It can be shown

that the solution of σ^2 in (3.7) has a closed form $(\hat{\sigma}^{*2})^{(1)} = \mathbf{X}_{R+1}^{**'} \mathbf{y}^{**} / \mathbf{X}_{R+1}^{**'} \mathbf{X}_{R+1}^{**}$, where $\sigma^{*2} = \sum_{r=1}^R c_r \theta_r + \sigma^2$. Furthermore,

$$\hat{\boldsymbol{\theta}}^{*(1)} = \arg \min_{\boldsymbol{\theta}^*} \left\{ (1/2) (\mathbf{y}^{**} - \mathbf{X}^{**} \boldsymbol{\theta}^*)' (\mathbf{y}^{**} - \mathbf{X}^{**} \boldsymbol{\theta}^*) + N \sum_{r=1}^R |\theta_r^*| \right\}, \quad (3.8)$$

where $\mathbf{X}^{**} = [\mathbf{X}_1^{**}, \dots, \mathbf{X}_R^{**}]$ and $\boldsymbol{\theta}^* = \text{diag}\{\tau_r\}_{r=1}^R \boldsymbol{\theta}$. Hence, $\hat{\boldsymbol{\theta}}^{(1)} = \text{diag}\{\tau_r^{-1}\}_{r=1}^R \hat{\boldsymbol{\theta}}^{*(1)}$ and $(\hat{\sigma}^2)^{(1)} = (\hat{\sigma}^{*2})^{(1)} - \sum_{r=1}^R c_r \hat{\theta}_r^{(1)}$.

Let $\hat{\boldsymbol{\eta}}_{\text{APMLE}} = \hat{\boldsymbol{\eta}}^{(1)}$ denote the approximate penalized maximum likelihood estimates (APMLE) of $\boldsymbol{\eta}$, where $\hat{\boldsymbol{\eta}}^{(1)} = (\hat{\boldsymbol{\beta}}^{(1)'}, \hat{\boldsymbol{\gamma}}^{(1)'})'$. Equations (3.6) and (3.8) can be solved by a LARS algorithm, and thus the computation is efficient. Although Equation (3.4) can be iterated until convergence, a one-step solution is preferred here because it is computationally efficient and the estimates still possess desirable asymptotic properties, as is explained in the next section.

3.2. STANDARD ERRORS AND OTHER COMPUTATIONAL ASPECTS

In Reyes (2010) Chapter 3, the existence, consistency, and sparsity of the PMLE were established. In addition, a central limit theorem for the PMLE of the nonzero-valued regression and spatial-temporal coefficients were given. Furthermore, the consistency of the APMLE and its asymptotic normality at the rate of $N^{1/2}$ were established, with the same limiting distribution as that of the PMLE. Thus, although $\hat{\boldsymbol{\eta}}_{\text{APMLE}}$ obtained from our computational algorithm is a local optimum and is not necessarily a global optimum, it has the desirable asymptotic properties. Interestingly, the asymptotic framework requires that either the number of time points of observation tend to infinity or the spatial lattice, but not necessarily both. Thus it includes the practical situation where the spatial lattice is fixed, but observations are made repeatedly over longer periods of time. Accordingly, we have

$$\text{var}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{\mathcal{I}}(\boldsymbol{\beta})^{-1}, \quad \text{var}(\hat{\boldsymbol{\gamma}}) \approx \boldsymbol{\mathcal{I}}(\boldsymbol{\gamma})^{-1},$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta})$ and $\boldsymbol{\mathcal{I}}(\boldsymbol{\gamma})$ are obtained from the expected information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\eta})$.

Since $\text{E}_{\boldsymbol{\eta}}\{-\partial^2 \ell(\boldsymbol{\eta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'\} = \mathbf{0}$, the expected information matrix becomes

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\eta}) = \text{diag}\{\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}), \boldsymbol{\mathcal{I}}(\boldsymbol{\gamma})\}, \quad (3.9)$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) = \text{E}_{\boldsymbol{\eta}}\{-\frac{\partial^2 \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\} = \mathbf{X}' \boldsymbol{\Gamma}^{-1} \mathbf{X}$ and the (r, r') th entry of $\boldsymbol{\mathcal{I}}(\boldsymbol{\gamma}) = \text{E}_{\boldsymbol{\eta}}\{-\frac{\partial^2 \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}\}$ is $(1/2) \text{tr}(\boldsymbol{\Gamma}^r \boldsymbol{\Gamma} \boldsymbol{\Gamma}^{r'} \boldsymbol{\Gamma})$, with $\boldsymbol{\Gamma}^r = \frac{\partial \boldsymbol{\Gamma}^{-1}}{\partial \gamma_r}$ for $r, r' = 1, \dots, R+1$ where, for ease of presentation, $\gamma_{R+1} = \sigma^2$. Evaluating (3.9) at the APMLE, we obtain estimates of the covariance matrix, which are used in Section 5.2 to compute the standard errors of AMPLE.

To estimate the regularization parameters, $\{\lambda_j\}_{j=1}^J$ and $\{\tau_r\}_{r=1}^R$, we let

$$\lambda_j = \lambda \log(N) (N |\hat{\beta}_j|)^{-1}, \quad \tau_r = \tau \log(N) (N |\hat{\theta}_r|)^{-1}, \quad (3.10)$$

where $\hat{\beta}_j$ and $\hat{\theta}_r$ are the entries of $\hat{\boldsymbol{\beta}}_{\text{MLE}}$ and $\hat{\boldsymbol{\theta}}_{\text{MLE}}$. The dimension reduction in (3.10) is useful, as now only two regularization parameters instead of $J + R$ need to be estimated

(Zhu, Huang, and Reyes 2010). To determine λ and τ , we compute Bayesian information criterion (BIC),

$$\text{BIC}(\lambda, \tau) = -2\ell(\hat{\eta}; \lambda, \tau) + e(\lambda, \tau) \log N, \tag{3.11}$$

where $e(\lambda, \tau) = \sum_{j=1}^J I\{\hat{\beta}_j \neq 0\} + \sum_{r=1}^R I\{\hat{\theta}_r \neq 0\}$, for all combinations of λ and τ (Wang, Li, and Tsai 2007b). We select the combination that has the smallest BIC value. Since we utilize a LARS-type algorithm, we may obtain parameter estimates at different λ and τ in one path.

4. SIMULATION STUDY

We conduct simulation to examine the finite-sample properties of APMLE. Consider two square lattices with sizes 5×5 and 10×10 . The corresponding lattices sizes are $I = 25$ and 100 . For each lattice size, we consider two total time points $T = 5$ and 10 . That is $(T = 5, I = 25)$, $(T = 10, I = 25)$, $(T = 5, I = 100)$, and $(T = 10, I = 100)$, corresponding to sample sizes $N = 125, 250, 500$, and $1,000$, respectively. For the sample size $N = 500$, we consider an additional case of $(T = 20, I = 25)$ to compare with the case of $(T = 5, I = 100)$.

For linear regression, we set seven covariates to follow a standard normal distribution. Covariates are constructed as follows to have spatial, temporal and cross-covariate correlation. First, generate the j th covariate $\mathbf{u}_j = (u_{j,1,1}, \dots, u_{j,I,T})'$, such that it has an exponential covariance function with no nugget and range parameter 1. That is, $\text{cov}(u_{j,i,t}, u_{j,i',t'}) = \exp\{-d(s_i, s_{i'})\}$, where $d(s_i, s_{i'})$ is the distance between sites i and i' , $i, i' = 1, \dots, I$. Next, we construct $\mathbf{z}_{i,t} = (z_{1,i,t}, \dots, z_{7,i,t})'$ by letting $\mathbf{z}_{i,t} = \mathbf{A}\mathbf{u}_{i,t}$, where $\mathbf{u}_{i,t} = (u_{1,i,t}, \dots, u_{7,i,t})'$, $\mathbf{A}\mathbf{A}' = [\rho_A^{|j-j'|}]_{j,j'=1}^7$, and $\rho_A = 0.5$. Finally, the temporal correlation is induced by letting $\mathbf{x}_{j,i} = \mathbf{B}\mathbf{z}_{j,i}$, where $\mathbf{x}_{j,i} = (x_{j,i,1}, \dots, x_{j,i,T})'$ for the j covariate at site i , $\mathbf{z}_{j,i} = (z_{j,i,1}, \dots, z_{j,i,T})'$, $\mathbf{B}\mathbf{B}' = [\rho_B^{|t-t'|}]_{t,t'=1}^T$, and $\rho_B = 0.5$. The regression coefficients are set to $\boldsymbol{\beta} = (4, 3, 2, 1, 0, 0, 0)'$. We standardize each covariate to have mean 0 and variance 1, as well as the response variable to have mean 0. Thus there will be no intercept term.

The error term is assumed to have mean 0 and follow the covariance function (2.4). The spatial-temporal coefficients are set to $\boldsymbol{\theta} = (0.2, 0, 0.05, 0, 0.1)'$. That is, for both time lags $l = 0, 1$, the true neighborhood structure is of the first order, and the true temporal structure is of the first order. In model fitting, however, we allow both first and second order for the spatial neighborhood structure.

For each combination of I and T , a total of 100 data sets are simulated. We compute APMLEs by the algorithm in Section 3. Table 1 summarizes the results. In particular, we compute an average number of correctly identified non-zero values coefficients, as well as zero-valued coefficients, for both regression and spatial-temporal coefficients. For a non-zero-valued coefficient, we compute an average of the APMLE and compare that with the true parameter value for evaluating accuracy of the APMLE. From the APMLEs, we also compute a standard deviation for assessing precision of the estimates. We present the estimate of the variance component even though it is not subject to penalty.

Table 1. Average number of correctly identified zero and nonzero regression coefficients $\{\beta_j\}$ and spatial-temporal coefficients $\{\theta_{k,l}\}$, mean approximate penalized maximum likelihood estimate (MAPMLE), and standard deviation (SD).

Grid size I	25	25	25	100	100	
Time points T	5	10	20	5	10	
Sample size N	125	250	500	500	1000	Truth
<i>Regression coefficients</i>						
Selection of covariates						
# nonzero β_j	4.00	4.00	4.00	4.00	4.00	4
# zero β_j	1.76	1.78	1.81	1.86	1.79	3
MAPMLE and SD						
β_1	3.99	3.99	4.00	4.00	4.00	4.00
SD	0.0106	0.0075	0.0043	0.0044	0.0021	
β_2	3.02	3.02	3.00	3.01	3.00	3.00
SD	0.0377	0.0170	0.0077	0.0051	0.0029	
β_3	1.98	2.01	2.01	2.00	2.00	2.00
SD	0.0389	0.0121	0.0052	0.0057	0.0031	
β_4	0.99	0.99	1.00	0.99	1.01	1.00
SD	0.0197	0.0071	0.0039	0.0064	0.0031	
<i>Spatial-temporal coefficients</i>						
Selection of spatial-temporal dependence						
# nonzero $\theta_{k,l}$	1.86	2.19	2.52	2.37	2.86	3
# zero $\theta_{k,l}$	1.45	1.48	1.50	1.66	1.23	2
MAPMLE and SD						
$\theta_{1,0}$	0.20	0.20	0.20	0.20	0.21	0.20
SD	0.0011	0.0004	0.0002	0.0002	0.0004	
$\theta_{1,1}$	0.02	0.04	0.04	0.03	0.05	0.05
SD	0.0024	0.0023	0.0014	0.0013	0.0006	
$\theta_{0,1}$	0.06	0.09	0.10	0.08	0.12	0.10
SD	0.0096	0.0081	0.0058	0.0054	0.0051	
<i>Variance component</i>						
MAPMLE and SD						
σ^2	0.92	0.98	0.99	0.98	1.02	1.00
SD	0.0226	0.0085	0.0040	0.0044	0.0046	

As the lattice size I increases or the number of time points T increases, variable selection improves in terms of identification of both zero-valued and non-zero-valued coefficients. Our method seems to identify the non-zero-valued coefficients with more accuracy than the zero-valued coefficients. There appears to be no major difference in the results between the combination of I and T for the sample size $N = 500$. For estimation of the non-zero-valued coefficients, both accuracy and precision improve as sample size increases when either the lattice size I increases or the number of time points T increases.

5. IMPACT OF CLIMATE ON A MOUNTAIN PINE BEETLE OUTBREAK

5.1. SCIENTIFIC BACKGROUND

The mountain pine beetle (MPB) is an insect native to western North America. Approximately the size of a grain of rice, female beetles tunnel into the bark of mature pines, emitting aggregation pheromones that attract mates. Eggs are laid in galleries underneath the bark, where the developing progeny feed within the water-conducting tissues. This feeding, in concert with fungi vectored by the adult beetles, can kill mature trees when beetles are at outbreak levels. Outbreaks typically occur when there is an abundance of suitable host trees and environmental conditions permit beetle populations to increase. Temperature is typically the most critical factor, serving as a cue for emergence of the annual flight of adults and governing overwintering success. The insects overwinter as larvae within the trees, where prolonged periods of temperatures under -40°C may exert lethal mortality to the population. Ongoing research is aimed at elucidating the nature of various factors in outbreaks of MPB, such as habitat heterogeneity, climate, reproduction, and dispersal. Identifying and understanding the key factors could result in reliable models that would greatly facilitate management and planning of pine forests.

Our case study concerns an outbreak of MPB in a study region of British Columbia, Canada, from 1977–1986, which spread over almost 800,000 ha of mature pine forest before collapsing due to abnormally cold autumnal weather (Stahl et al. 2006; Aukema et al. 2008). We present the MPB data as a case study for five reasons. First, it allows study of a recent outbreak from initiation to spread and propagation to collapse phase. Second, the geographic area is relatively uniform with mean elevation across the Plateau region of 1345 m (Aukema et al. 2008), which may foster uniform dispersal and, consequently, be captured by spatial dependence. Third, temporal dependence may be suitable for modeling the annual generation times exhibited by the majority of the population during this outbreak. Fourth, temperature variation is extreme, spanning more than 70°C , which allows good inference on environmental covariates. Finally, alternate analytical frameworks exist for this data set, which provide a valuable comparative basis (Aukema et al. 2008). Robust statistical tools that can provide inference and prediction of outbreak spread behavior are critical to providing management advice, especially given the scope of the current outbreak (Safranyik et al. 2010).

5.2. DATA DESCRIPTION

The data were collected as follows. Annual aerial surveys of killed trees, whose crowns fade to red within a year of being killed by mountain pine beetle, were overlaid on a spatial lattice of 469 cells at approximate 12 km resolution sensu Aukema et al. (2008). The covariates we consider in our model include the average elevation of the cell and several temperature terms. The temperature terms are defined from interpolations of observed climate station data according to Stahl, Moore, and McKendry (2006). Variables include the minimum, mean, and maximum temperatures for each cell over each calendar year.

As well, we include variables defined according to the ecology of MPB per Safranyik, Shrimpton, and Whitney (1975). We calculate the mean August temperature for each cell, as beetles typically emerge to seek new hosts during late summer. Two degree day terms, DD and DDEG, reflect accumulated heat days above 5.5 °C from August to the end of the growing season (e.g., 50 % egg hatch requires 306 °C degree days) and from previous August through current July (e.g., a single annual generation requires approximately 833 °C degree days). Finally, a precipitation term is included, although is of uncertain utility given the low reliability of interpolating precipitation from point weather stations across mountainous regions.

In previous work, the focus was on binary data that represent presence and absence of outbreak in any given cell as the response variable. While such a coarse response yielded useful results at a landscape scale, we seek to take advantage of finer details of the spatial pattern that may be indicative of the eruptive potential of the insect. In the previous approach using binary data aggregated to the cell level, a cell with several dozen small point infestations across a 10 km cell appears the same as a cell with only one patch of trees killed by this insect. Here, we count the number of patches of trees killed by MPB to obtain an infestation intensity per cell. For the binary data, spatial-temporal autologistic regression was applied to construct landscape models, where Monte Carlo maximum likelihood estimation was developed for parameter estimation (Aukema et al. 2008; Zhu et al. 2008), an improvement over the pseudolikelihood estimation (Zhu, Huang, and Wu 2005). For model selection, a two-step approach was taken. First, an information criterion was used to determine a suitable order for the spatial and temporal neighborhood, and then covariates were selected by backward elimination.

Here, in contrast, we expand these previous analysis and use intensity of infestation in each cell. This is a more ecologically realistic approach to studying outbreak dynamics of bark beetles, because outbreaks typically commence as a single mass-attacked tree, growing to patches of 10–20 trees in subsequent years, before coalescing into large gray areas where trees have lost their needles. Moreover, equipped with the new method, we are able to identify the more important covariates and appropriate spatial-temporal neighborhood structure by penalized maximum likelihood. There are at least two advantages of our new approach. First, our model selection procedure has desired asymptotic properties in terms of consistency, sparsity, and asymptotic normality; and second, the accompanying computational algorithm for our method is efficient and thus more feasible for practical use. Because the previous algorithm required Monte Carlo simulation, even though computation was faster than Bayesian inference, it was still time consuming and sometimes numerically unstable (Zheng and Zhu 2008).

Figure 1 shows the geographic location of the study region and the average digital elevation in the 12 km by 12 km cells of the grid in our study area. The MPB infestation intensity on the log scale is mapped for each of the 10 years from 1977 to 1986 in Figure 2. The infestation level intensified in the early 1980s and dropped sharply after 1985. For the analysis, infestation intensity on the log scale is used as response variable to compensate for the skewness and the possible non-normality of infestation intensity. Among the covariates, strong collinearity exists. For example, elevation tends to be negatively correlated

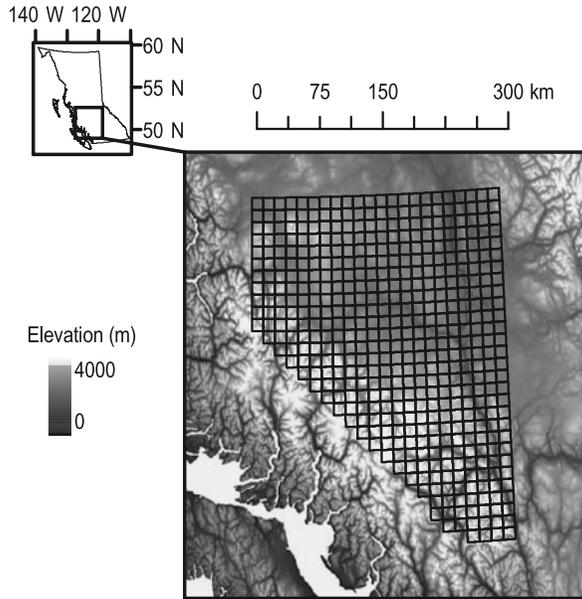


Figure 1. Top: Study region in British Columbia of Canada; bottom: map of elevation on a spatial lattice overlaying the study region.

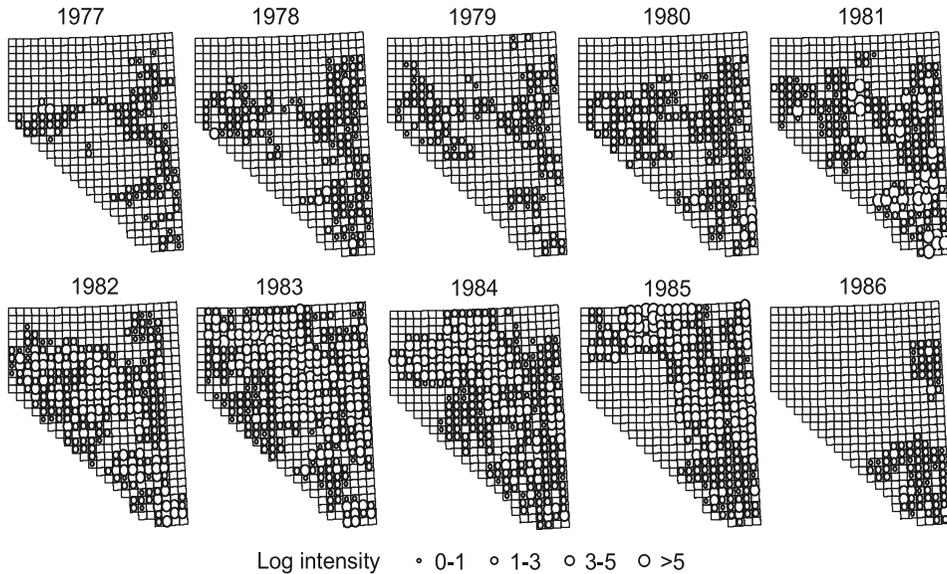


Figure 2. Map of log intensity of mountain pine beetle infestation from 1977 to 1986. The size of a circle reflects the magnitude of log intensity. Unmarked cells correspond to absence of infestation.

with temperature and different metrics of warm temperature are positively correlated with each other. As before, we standardize each covariate to have mean 0 and variance 1, as well as the response variable to have mean 0.

Table 2. Nonzero approximate penalized maximum likelihood estimate and standard error in parentheses of model parameters in the study of a mountain pine beetle outbreak with selection of both covariates and spatial-temporal dependence structure, covariates only, and spatial-temporal dependence structure only.

Model Selection		Covariates & Dependence	Covariates Only	Dependence Only
Covariates				
Elevation	β_1	-0.11 (0.07)	-0.29 (0.04)	-0.29 (0.07)
Temp min	β_2	-	0.01 (0.06)	0.01 (0.06)
Temp max	β_3	0.13 (0.06)	0.34 (0.06)	0.34 (0.06)
Temp mean	β_4	-	0.28 (0.11)	0.28 (0.10)
August temp mean	β_5	-	-0.23 (0.08)	-0.23 (0.07)
DD	β_6	-	-0.12 (0.08)	-0.12 (0.07)
DDEGG	β_7	-	-0.12 (0.07)	-0.12 (0.08)
Precip	β_8	-	0.06 (0.03)	0.06 (0.03)
Time Lag 0				
Spatial 1st	$\theta_{1,0}$	0.17 (0.005)	0.16 (0.005)	0.17 (0.005)
2nd	$\theta_{2,0}$	-	-0.05 0.004	-
Time Lag 1				
Time Lag 1	$\theta_{0,1}$	0.61 (0.02)	0.21 (0.02)	0.61 (0.02)
Spatial 1st	$\theta_{1,1}$	-	0.04 0.008	-
2nd	$\theta_{2,1}$	-0.10 (0.008)	-0.12 (0.007)	-0.10 (0.008)
Time Lag 2				
Time Lag 2	$\theta_{0,2}$	-	0.21 (0.02)	-
Spatial 1st	$\theta_{1,2}$	-	0.09 (0.009)	-
2nd	$\theta_{2,2}$	-	-0.14 (0.008)	-
Variance component				
	σ^2	0.75 (0.02)	1.27 (0.03)	0.75 (0.02)
BIC		3660	5366	3741

5.3. MODEL SELECTION VIA SPATIAL-TEMPORAL LASSO

Now, we apply the PMLE method described in Section 3 to perform model selection and parameter estimation. For spatial-temporal dependence, we consider up to two time lags and up to two orders of neighbors. Regarding the initial values of $\boldsymbol{\epsilon}_{1-l}$ for $l = 1, \dots, L$, the results under the zero initial values are shown in Table 2, as they outperformed the fit under the alternative, boundary-effect adjusted initial values, based on BIC values. Three practical situations are considered for model selection. In the first case, covariates and spatial-temporal dependence structures are selected simultaneously. In the second case, only covariates are selected, but the spatial-dependence structure is fixed at the highest order two for both time lags and neighborhoods. In the third case, only spatial-temporal

dependence structures are selected, but not the covariates. Among the three situations, the first one has the smallest BIC value. Regarding the normality assumption made here, a graphical inspection of the residuals did not produce evidence against it.

In the first case where both covariates and spatial-temporal dependence structures are selected, two covariates are selected: elevation and maximum temperature. The APMLE of the coefficient for elevation is -0.11 with an estimated standard error (SE) of 0.07 , whereas the APMLE of the coefficient for maximum temperature is 0.13 with an estimated SE of 0.06 . That is, there is a negative relationship between elevation and MPB infestation intensity, but the relationship is positive between maximum temperature and MPB infestation intensity. It suggests that higher elevations are associated with fewer patches of trees killed by MPB, possibly because regions at higher elevations tend to have a cooler climate and less pine. It also supports the hypothesis that warmer temperatures are associated with more MPB infestation, as flights of adult bark beetles in search of new trees are modulated by warmer temperature thresholds (Powell et al. 2000). In the second case where only the covariates are selected, all the covariates are kept and the parameters estimates are close to the third case where covariates are not selected. However, signs of some of the regression coefficient estimates are counter-intuitive. For example, the APMLE of the coefficient for August mean temperature is -0.23 with a SE of 0.08 . This may well be a consequence of collinearity among the covariates. Moreover, the BIC value for the second case is larger than the other two cases. Although the first and the third cases have similar BIC values, the first case has smaller BIC, and appears to be the better model.

For spatial and temporal dependence structure, the two cases involving spatial-temporal dependence selection choose the same neighborhood structure. Only the first-order spatial neighbors are selected for time lag 0 (i.e. within a same year), whereas only the same site and the second-order spatial neighbors are selected at time lag 1 (i.e. from a previous year). At time lag 2 (i.e. from the year before last), none of the coefficients are selected. Clearly, there is spatial-temporal interaction and a lack of space-temporal separability in the sense that not all $\theta_{k,l} = 0$ for $k \geq 1, l \geq 1$. Furthermore, the shift from the first-order to the second-order spatial dependence gives some evidence of gradual dispersion over time. The negative coefficient associated with second-order spatial neighbors may seem counter-intuitive for a building outbreak, but could reflect a tradeoff with the other larger positive coefficients of time lag 0 and 1 as discrete patches of dead trees begin to coalesce into larger areas.

5.4. COMPARISON AGAINST OTHER APPROACHES

In the presence of complex spatial-temporal dependence structure, there currently is no standard technique for model selection. For comparison, we consider a two-step approach. First, BIC is used to choose an appropriate spatial-temporal dependence structure from a list of candidates. Then, with the chosen spatial-temporal dependence structure in the error model, a backward elimination based again on BIC is applied to select covariates.

We consider this alternative technique for several reasons. First, it is common practice to combine linear regression with information criteria (Venables and Ripley 2002). Second, BIC is a widely used information criterion for model assessment and backward elimination

is commonly used for the selection of covariates. Third, without separating the selection of covariates and dependence structures, it is impractical to rely solely on BIC for model selection especially when the number of covariates is large. Finally, in previous work, a similar approach taken has produced reasonable results for binary responses (see, e.g., Aukema et al. 2008; Zhu et al. 2008).

For the error model in (2.2), there are many candidate spatial-temporal dependence structures that can be specified. However, for each of the possible dependence structures, all of the model parameters need to be estimated and the BIC values computed. Such a task can be not only computationally costly, but also numerically unstable, especially for complex models. Thus, for simplicity, we focus on only two types of dependence structure, namely, a spatial autoregressive (SAR) model and a temporal autoregressive model up to the L th order (AR(L)). The SAR model can be defined by setting $L = 0$ and $C_0 = \delta W$, where $W = [w_{i,i'}]_{i,i'=1}^I$ is an $I \times I$ matrix consisting of spatial weights according to a neighborhood structure. The AR(L) model can be specified by $C_0 = \mathbf{0}$ and $C_l = \rho_l I_l$ for $l = 1, \dots, L$.

Candidate models can be specified by changing the neighborhood structure for the SAR model and by proposing different values of L for AR(L). Here, for W on a regular grid, we consider a first-order neighborhood that comprises the four nearest neighbors (north, south, west, and east), or combine the first-order and the second-order neighborhoods that comprise the nearest and the second-nearest neighbors (northwest, northeast, southwest, and southeast). For the temporal case, either $L = 1$ or $L = 2$.

Maximum likelihood was used for estimation of the parameters of candidate models. The resulting MLEs are presented in Table 2. Based on BIC, models with only spatial dependence are better than models with only temporal dependence. In fact, the model with the smallest BIC is a SAR model that includes the first-order and second-order neighbors. Using this dependence structure, we then applied a backward elimination to select covariates which excluded elevation from the model. The estimates produced are presented in the last column of Table 3.

On the one hand, with positive estimated coefficients, higher mean temperature and August mean temperature appear to be associated with higher levels of MPB infestation. The estimated spatial coefficient shows a positive dependence between neighboring sites, which is supported by the pattern seen in Figure 2. In fact, this is not unexpected, as temperature works to synchronize populations with the same density-dependent structure across broad spatial areas (Aukema et al. 2006), and these insects can disperse several kilometers in search of new host trees when at outbreak levels (Robertson et al. 2009; De la Giroday et al. 2011). On the other hand, with negative estimated coefficients, higher values of minimum and maximum temperature, DD, DDEG, and precipitation are associated with lower levels of MPB infestation. Some of these results are counter-intuitive.

Although the BIC value of the model developed by this alternative approach is smaller than the BIC of the best model selected by spatial-temporal Lasso, the latter resulted in a simpler model. It is also ecologically tractable and does not suffer as much from counter-intuitive interpretation of the regression coefficients likely induced by collinearity among the covariates. Similar climatic signatures are apparent in analogous spatial-temporal lat-

Table 3. Maximum likelihood estimate and standard error in parentheses of model parameters in the study of a mountain pine beetle outbreak. The errors are assumed to independent, AR(2), or SAR(2). Also, a two-step model selection is applied such that the dependence structure is selected first and then the covariates via backward elimination.

Model		Independent Error	AR(2)	SAR 1st order	SAR 1st & 2nd order	SAR 1st & 2nd order Backward
Covariates						
Elevation	β_1	-0.25 (0.03)	-0.29 (0.05)	-0.27 (0.05)	0.04 (0.04)	-
Temp min	β_2	0.19 (0.04)	0.34 (0.04)	0.15 (0.08)	-0.40 (0.08)	-0.60 (0.07)
Temp max	β_3	0.48 (0.05)	0.49 (0.04)	0.47 (0.10)	-0.41 (0.10)	-0.15 (0.08)
Temp mean	β_4	0.28 (0.08)	0.08 (0.07)	0.28 (0.16)	1.22 (0.15)	1.31 (0.12)
August temp mean	β_5	-0.39 (0.06)	-0.43 (0.05)	-0.38 (0.12)	-0.11 (0.13)	0.17 (0.11)
DD	β_6	-0.15 (0.06)	-0.18 (0.05)	-0.15 (0.12)	0.23 (0.13)	-0.35 (0.11)
DDEG	β_7	-0.18 (0.05)	-0.04 (0.05)	-0.19 (0.12)	-1.02 (0.13)	-0.50 (0.11)
Precip	β_8	0.07 (0.02)	0.04 (0.02)	0.05 (0.05)	-0.61 (0.06)	-0.26 (0.05)
Dependence Structure						
Time Lag 1	ρ_1		0.30 (0.01)			
Time Lag 2	ρ_2		0.29 (0.02)			
Spatial	δ			0.19 (0.003)	0.11 (0.001)	0.11 (0.001)
Variance	σ^2	1.93 (0.04)	1.51 (0.03)	1.21 (0.03)	0.85 (0.02)	0.60 (0.01)
BIC		7842	6722	4232	3835	3287

tice models examining effects of climate change within the current, ongoing outbreak (Sambaraju et al. 2011).

6. CONCLUSIONS

We have considered a spatial-temporal linear regression model and in particular, a new statistical method that simultaneously performs model selection and parameter estimation via a spatial-temporal adaptive Lasso. In a case study, we have evaluated the impact of climate conditions on the tree-killing ability of an eruptive species of bark beetle in pine forests of British Columbia, Canada. In particular, we have applied this approach to identify the appropriate components of a general model that features the factors that propagate an outbreak of MPB and interpret the results from ecological perspectives. A comparison has been made against an alternative, two-step model selection procedure.

The new method requires specification of the model in its most general form. The technique produces estimates of a reduced model, where covariates and dependence structures can be selected simultaneously. The resulting estimates have good asymptotic and finite-sample properties that other ad-hoc procedures may not possess. In addition, alternative approaches like the one described in Section 5.4 involve a great deal of trial and error and the results still seem to be unstable. Depending on whether the covariates or the dependence structure is selected first in a two-step procedure, the final results may vary.

ACKNOWLEDGEMENTS

Funding has been provided for this research from a USDA Cooperative State Research, Education and Extension Service (CSREES) Hatch project and the TRIA project of Genome Centers of British Columbia, Alberta, and Canada.

[Published Online September 2012.]

REFERENCES

- Aukema, B. H., Carroll, A. L., Zhu, J., Raffa, K. F., Sickley, T. A., and Taylor, S. W. (2006), "Landscape Level Analysis of Mountain Pine Beetle in British Columbia, Canada: Spatiotemporal Development and Spatial Synchrony Within the Present Outbreak," *Ecography*, 29, 427–441.
- Aukema, B. H., Carroll, A. L., Zheng, Y., Zhu, J., Raffa, K. F., Moore, R. D., and Stahl, K. (2008), "Movement of Outbreak Populations of Mountain Pine Beetle: Influences of Spatiotemporal Patterns and Climate," *Ecography*, 31, 348–358.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman and Hall.
- Battistia, A., Stastny, M., Buffo, E., and Larsson, S. (2006), "A Rapid Altitudinal Range Expansion in the Pine Processionary Moth Produced by the 2003 Climatic Anomaly," *Global Change Biology*, 12, 662–671.
- Bentz, B. J., Regnier, J., Fettig, C. J., Hansen, E. M., Hayes, J. L., Hicke, J. A., Kelsey, R. G., Negron, J. F., and Seybold, S. J. (2010), "Climate Change and Bark Beetles of the Western United States and Canada: Direct and Indirect Effects," *Bioscience*, 60, 602–613.
- Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley, revised edition.
- De la Giroday, H.-M. C., Carroll, A. L., Lindgren, B. S., and Aukema, B. H. (2011), "Association of Landscape Features With Dispersing Mountain Pine Beetle Populations During a Range Expansion Event in Western Canada," *Landscape Ecology*, 26, 1097–1110.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), "Least Angle Regression" (with discussion), *Annals of Statistics*, 32, 407–499.
- Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360.
- Huang, H.-C., Hsu, N.-J., Theobald, D., and Breidt, F. J. (2010), "Spatial LASSO With Applications to GIS Model Selection," *Journal of Computational and Graphical Statistics*, 19, 963–983.
- Jenkins, M. J., Hebertson, E. G., Page, W., and Jorgensen, C. A. (2008), "Bark Beetles, Fuels, Fire and Implications for Forest Management in the Intermountain West," *Forest Ecology and Management*, 254, 16–34.
- Kurz, W. A., Dymond, C. C., Stinson, G., Rampley, G. J., Neilson, E. T., Carroll, A. L., Ebata, T., and Safranyik, L. (2008), "Mountain Pine Beetle and Forest Carbon Feedback to Climate Change," *Nature*, 452, 987–990.
- Powell, J. A., Jenkins, J. L., Logan, J. A., and Bentz, B. J. (2000), "Seasonal Temperature Alone Can Synchronize Life Cycles," *Bulletin of Mathematical Biology*, 62, 977–998.
- Raffa, K. F., Aukema, B. H., Bentz, B. J., Carroll, A. L., Hicke, J. A., Turner, M. G., and Romme, W. H. (2008), "Cross-Scale Drivers of Natural Disturbances Prone to Anthropogenic Amplification: The Dynamics of Bark Beetle Eruptions," *Bioscience*, 58, 501–518.
- Reyes, P. E. (2010), "Selection of Spatial and Spatial-Temporal Linear Models for Lattice Data". PhD thesis, University of Wisconsin-Madison.
- Robertson, C., Nelson, T. A., Jelinski, D. E., Wulder, M. A., and Boots, B. (2009), "Spatial-Temporal Analysis of Species Range Expansion: The Case of the Mountain Pine Beetle *Dendroctonus Ponderosae*," *Journal of Biogeography*, 36, 1446–1458.
- Safranyik, L., Shrimpton, D. M., and Whitney, H. S. (1975), "An Interpretation of the Interaction Between Lodgepole Pine, the Mountain Pine Beetle and Its Associated Blue Stain Fungi in Western Canada," in *The Biology*

- and Epidemiology of the Mountain Pine Beetle in Lodgepole Pine Forests*, ed. D. M. Baumgartner, pp. 406–428. Management of Lodgepole Pine Ecosystems Symposium Proceedings, Washington State University Cooperative Extension Service, Pullman, Washington
- Safranyik, L., Carroll, A. L., Regniere, J., Langor, D. W., Riel, W. G., Shore, T. L., Peter, B., Cooke, B. J., and Nealis, V. G. S. W. T. (2010), “Potential for Range Expansion of Mountain Pine Beetle Into The Boreal Forest of North America,” *The Canadian Entomologist*, 142, 415–442.
- Sambaraju, K. R., Carroll, A. L., Zhu, J., Stahl, K., Moore, R. D., and Aukema, B. H. (2011), “Climate Change Could Alter the Distribution of Mountain Pine Beetle Outbreaks in Western Canada,” *Ecography*, 35, 211–223.
- Schabenberger, O., and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Boca Raton: Chapman and Hall.
- Stahl, K., Moore, R. D., and McKendry, I. G. (2006), “Climatology of Winter Cold Spells in Relation to Mountain Pine Beetle Mortality in British Columbia, Canada,” *Climate Research*, 32, 13–23.
- Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G., and McKendry, I. G. (2006), “Comparison of Approaches for Spatial Interpolation of Daily Air Temperature in a Large Region With Complex Topography and Highly Variable Station Density,” *Agricultural and Forest Metereology*, 139, 224–236.
- Tibshirani, R. (1996), “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society, Series B*, 58, 267–288.
- Venables, W. N., and Ripley, B. D. (2002), *Modern Applied Statistics with S* (4th ed.), Berlin: Springer.
- Wang, H., Li, G., and Tsai, C.-L. (2007a), “Regression Coefficients and Autoregressive Order Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society, Series B*, 69, 63–78.
- Wang, H., Li, R., and Tsai, C.-L. (2007b), “Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method,” *Biometrika*, 94, 553–568.
- Zheng, Y., and Zhu, J. (2008), “Markov Chain Monte Carlo for Spatial-Temporal Autologistic Regression Model,” *Journal of Computational and Graphical Statistics*, 17, 123–127.
- Zhu, J., Huang, H.-C., and Reyes, P. (2010), “On Selection of Spatial Linear Models for Lattice Data,” *Journal of the Royal Statistical Society, Series B*, 72, 389–402.
- Zhu, J., Huang, H.-C., and Wu, J. (2005), “Modeling Spatial-Temporal Binary Data Using Markov Random Fields,” *Journal of Agricultural, Biological, and Environmental Statistics*, 10, 212–225.
- Zhu, Z., and Liu, Y. (2009), “Estimating Spatial Covariance Using Penalized Likelihood With Weighted L_1 Penalty,” *Journal of Nonparametric Statistics*, 21, 925–942.
- Zhu, J., Zheng, Y., Carroll, A. L., and Aukema, B. H. (2008), “Autologistic Regression Analysis of Spatial-Temporal Binary Data via Monte Carlo Maximum Likelihood,” *Journal of Agricultural, Biological, and Environmental Statistics*, 13, 84–98.
- Zou, H. (2006), “The Adaptive LASSO and Its Oracle Properties,” *Journal of the American Statistical Association*, 101, 1418–1429.
- Zou, H., and Li, R. (2008), “One-Step Sparse Estimates in Nonconcave Penalized Likelihood Models” (with discussion), *Annals of Statistics*, 36, 1509–1566.