Spatial-Temporal Generalized Linear Models with Bark Beetle and Other Damage Causing Agents in the Rocky Mountains Example

Kimberly A. Kaufeld*

Abstract
Zhu et al (2008) developed a binary spatial-temporal autologistic regression model which accounts for spatial and temporal dependence at discrete time points. It uses logistic regression to model a response variable on explanatory variables and autoregression on responses from spatial neighborhoods. This research extends the work of Zhu et al's autologistic binary model to generalized linear models such as nominal multinomial models and models for ordinal response data where the spatial grid changes at each time point. The data are measured repeatedly based on spatial distances over discrete time points. A spatial neighborhood structure is constructed and ordered with respect to the adjacency of the initial site. A spatial-temporal autologistic regression model draws samples using Monte Carlo estimation using a Gibbs Sampler to obtain estimates of the model parameters. A dataset of bark beetle and multiple damage causing agents in the Rocky Mountain Forest Region from 2005-2009 is used to demonstrate the methodology.

Key Words: Spatial-temporal, Multinomial, Generalized Linear Models, Bark Beetles

1. Introduction

Types of data that provide when and where data were collected are called spatial-temporal data. The spatial component is the location or where the data were recorded and the temporal component is when the data were recorded, at a discrete time. Spatial-temporal data have an important statistical characteristic where the observations that are closer in space and time are more similar than those that are further apart (Cressie, 2011). This implies that the data are not statistically independent as there is dependence in both time and space, temporal and spatial dependence. Spatial-temporal generalized linear models, models that have binary, multinomial, poisson, or normal response variables, account for both spatial and temporal dependence in the data to infer cause and effect relationships.

In 1972, Besag developed a set of spatial models called the auto-models. The models are based upon a conditional distribution in which the current site, s_i, depends upon all the other sites, s_{-i} on a lattice. The conditional probability distribution for each site belongs to an exponential family, i.e Poisson, Binomial, Normal, which model a response variable on a set of explanatory variables while accounting for spatial correlation. However, the temporal component for these models are stationary in that they only account for a single year.

Zhu et al. (2005) developed a spatial-temporal autologistic regression model (STARM) to account for both spatial and temporal dependence on discrete time intervals. The model is an extension of the atemporal version of the autologistic model developed by Besag (1972, 1974) and later used by Gumpertz et al. (1997) and Huffer and Wu (1998) using more efficient computation methods. The model captures the both the spatial and the temporal dependence simultaneously in order to capture the correlation over space and time. It uses logistic regression to model three things, (1) the response variable on explanatory variables (2) the autoregression on responses from spatial neighborhoods or locations due to the spatial dependence among sites and (3) the autoregression of the temporal term due

*University of Northern Colorado, 518 McKee Hall, Greeley, CO 80634

to different discrete times. It incorporates spatial correlation while modeling the relationship between the spatial binary response and the explanatory variables (Cressie, 1993, Zhu et al, 2005, Zheng and Zhu, 2008). The autologistic model by Zhu et al (2005) uses binary responses over space in time, however, oftentimes there are multiple categories that need to be modeled spatially and temporally.

This paper extends the spatial-temporal binary autologistic model developed by Zhu et al (2005), Zheng and Zhu (2008), and Zhu et al (2008), to multinomial responses as well as spatial-temporal models where the spatial grid changes with time. This paper is presented as follows. Section 2 presents the spatial-temporal autologistic model proposed by Zhu et al (2005) as well as a modified spatial-temporal autologistic model that has a grid that changes with time. Section 3 introduces the nominal multinomial and ordinal multinomial spatial-temporal models on a moving grid. In section 4, statistical inference using the Monte Carlo Maximum Likelihood method is presented. A real data example is provided in section 5. Discussion and conclusions are given in section 6.

2. Spatial-Temporal Autologistic Regression Model

2.1 Notation

The binary spatial-temporal autologistic model is used when there are dichotomous responses. It models the binary data on a spatial lattice repeatedly over time while accounting for the spatial and temporal dependence simultaneously (Zhu et al, 2005). Let the response variable, Y_{it}, denote a binary response at time t and site i. Let $i = 1, \cdots, I$ represent the sites on a spatial lattice collected over $t = 1, \cdots, T$ discrete times. The response variable at the t^{th} time of the i^{th} site is represented as $Y_{it} = 0$ or 1, where $\boldsymbol{Y}_t = (Y_{1t}, \cdots, Y_{It})'$ represent the binary responses for a given time. The k explanatory variables, denoted X_{itk}, are represented at the i^{th} site and at time t where $k = 1, \cdots, K$. The neighbors of order ℓ adjacent to the i^{th} site are represented by $\mathcal{N}_i^{(l)}$.

2.2 Spatial-Temporal Autologistic Regression Model (STARM)

The spatial-temporal autologistic model developed by Zhu et al (2005) has a spatial frame that stays the same for each time point. It models the response variable, Y_{it} at times $1, \cdots, t$, a conditional distribution of $\boldsymbol{Y_t}$ that depends on the most recent, S times, $t - 1, \cdots, t - S$. For each time, t, it is assumed that the response variable follows a Markov random field under a specified spatial neighborhood. A logistic regression model is used, where Y_{it} is Bernoulli due to the binary response. The conditional probability is specified as

$$p_{it} = P(Y_{it} = 1 | Y_{jt} : J \in \mathcal{N}_i, \boldsymbol{Y}_{t'} : t' = t-1, \cdots, t-S)$$

for $t = S+1, \cdots, T$ and the probability of success is defined as

$$p_{it} = \frac{\exp\{\eta_{it}\}}{1 + \exp\{\eta_{it}\}}.$$

The probabilities, p_{it}, are modeled using a logit link and the following systematic component

$$\eta_{it} = \sum_{k=0}^{K} \theta_k X_{itk} + \frac{1}{2}\left[\sum_{l=1}^{L} \phi_l \sum_{\mathcal{N}_i} Y_{jt}\right] + \sum_{s=1}^{S} \gamma_s Y_{i,t-s}$$

The distribution of the model at the i^{th} space is written as $f_{it}(Y_{it} = 1 | Y_{jt} : J \in \mathcal{N}_i, \boldsymbol{Y}_{t'} : t' = t-1, \cdots, t-S)$, which uses a logit transformation to model the probabilities defined as

$$\text{logit}(p_{it}) = \sum_{k=0}^{K} \theta_k X_{itk} + \frac{1}{2}\left[\sum_{l=1}^{L} \phi_l \sum_{\mathcal{N}_i} Y_{jt}\right] + \sum_{s=1}^{S} \gamma_s Y_{i,t-s} \quad (1)$$

where $\theta_0, \cdots, \theta_K$ are the regression coefficients, $\psi_1, \cdots, \phi_L$ are the spatial autoregressive coefficients, and $\gamma_1, \cdots, \gamma_S$ are the temporal autoregressive coefficients.

2.3 Modified Spatial-Temporal Autologistic Model

The modified spatial-temporal autologistic model has a spatial grid that changes for each time point, moving grid. It has the same binary responses and conditional probability as the STARM model but has a modified systematic component defined as

$$\eta_{it} = \sum_{k=0}^{K} \theta_k X_{itk} + \frac{1}{2}\left[\sum_{l=1}^{L} \phi_l \sum_{\mathcal{N}_i} \frac{1}{d_j} Y_{jt}\right] + \sum_{s=1}^{S} \frac{1}{d_j} \gamma_s Y_{i,t-s}. \quad (2)$$

The moving grid adds an additional component, d_j, which represents the euclidean distance for each site with respect to the magnitude of the distance between the current i^{th} site and any other site j. The component is added to both the spatial and the temporal term. The weight for the spatial component is added due to the irregularity of the spatial frame. The temporal term is assigned a weight due to the varying spatial frame each year. Since the spatial frame moves for each time, the previous time responses, $Y_{i,t-s}$ are found based upon the minimum euclidean distance to the current years response. Previous years observations that are closest to the current time point are assigned higher weights and those further away are assigned lower weights.

The joint distribution over $i = 1, \cdots I$ sites for a given time point, t can be defined based upon the Hammersley-Clifford theorem (Cressie, 1993) as

$$\begin{aligned} p(\boldsymbol{Y}_t : \boldsymbol{Y}'_t) &= \frac{\exp(Q(\boldsymbol{Y}_t))}{\sum_{Z \in \Omega} \exp(Q(\boldsymbol{Z}_t; \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma}))} \\ &= c(\boldsymbol{Y}_{t-1}, \cdots, \boldsymbol{Y}_{t-S} : \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma})^{-1} \times \exp \sum_{i=1}^{I} Y_{it}\eta_{it} \end{aligned}$$

where $c(\boldsymbol{Y}_{t-1}, \cdots, \boldsymbol{Y}_{t-S} : \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma})^{-1}$ is an unknown normalizing constant.

3. Multinomial Spatial-Temporal Autologistic Model

3.1 Notation

The multinomial spatial-temporal autologistic model is used when the data are categorical responses. We would like to model data from R categories on a spatial lattice, measured repeatedly over time, while accounting for the spatial and temporal dependence simultaneously. Let $i = 1, \cdots, I$ represent the sites on a spatial lattice collected over $t = 1, \cdots, T$ discrete times. The response variable, Y_{itr}, measured at the i^{th} site at time t for the r^{th} category, is represented similarly to a Bernoulli trial. Let $Y_{itr} = 1$ if the response is in the r^{th} category and $Y_{i,t,r} = 0$ if the response is not in category r. The categorical responses for a given time are represented as $Y_{itr} = (Y_{i1r}, \cdots Y_{ItR})$. The k explanatory variables, denoted X_{itrk}, are represented at the i^{th} site, time t and category r where $k = 1, \cdots, K$. The neighbors of order ℓ adjacent to the i^{th} site are represented by $\mathcal{N}_i^{(l)}$.

3.2 Nominal Multinomial Model Specification

The spatial-temporal multinomial model uses a multinomial logistic regression where the conditional distribution of Y_{itr} is Multinomial. The conditional probability is

$$p_{itr} = P(Y_{its} = r | Y_{jtr} : J \in \mathcal{N}_i, \boldsymbol{Y}_{t'r} : t' = t-1, \cdots, t-S),$$

where S is the most recent time point. For each time, t, it is assumed that the response variable follows a Markov random field under a specified spatial neighborhood. The euclidean distance for each site, denoted d_j, represents the magnitude of the distance between the current i^{th} site and any other site j. The weight for the spatial component is added due to the irregularity of the spatial frame. The temporal term is assigned a weight due to the varying spatial frame each year. Since the spatial frame moves for each time, the previous time responses, $Y_{i,t-s}$ are found based upon the minimum euclidean distance to the current years response. Previous years observations that are closest to the current time point are assigned higher weights and those further away are assigned lower weights. The counts at the R categories of Y are multinomial with probabilities $p_{it1}, \cdots p_{itR}$ and with the systematic component, η_{itr},

$$\eta_{itr} = \sum_{r=1}^{R-1} \sum_{k=0}^{K} \theta_{kr} X_{itrk} + \sum_{r=1}^{R-1} \frac{1}{2} \left[\sum_{l=1}^{L} \phi_{lr} \sum_{j \in \mathcal{N}_i^{(l)}} \frac{1}{d_j} Y_{jtr} \right] + \sum_{r=1}^{R-1} \sum_{s=1}^{S} \frac{1}{d_j} \gamma_{sr} Y_{i,t-s,r}$$

and logit link,

$$\text{logit}(p_{itr}) = \eta_{itr}$$

where $\theta_{0r}, \cdots, \theta_{Kr}$ denote the regression coefficients, $\phi_{r1}, \cdots, \phi_{Lr}$ the spatial autoregressive coefficients, and $\gamma_{1r}, \cdots, \gamma_{Sr}$ the temporal autoregressive coefficients.

In the nominal multinomial model, the probability of selecting the r^{th} category is defined as

$$p_{itr} = \frac{\exp\{\eta_{itr}\}}{1 + \sum_{r=1}^{R-1} \exp\{\eta_{itr}\}}.$$

3.2.1 Joint Probability Distribution

The conditional distribution with respect to the most recent S time for a multinomial response at the i^{th} space and r^{th} category is

$$f_{itr}(Y_{itr} = r | Y_{jtr} : J \in \mathcal{N}_i, \boldsymbol{Y}_{t'r} : t' = t-1, \cdots, t-S) = p_{itr}^{\mu}(1 - p_{itr})^{(1-\mu)},$$

where $\mu = \sum_{r=1}^{R-1} Y_{itr}$.

Using a log link, the conditional distribution can be defined as

$$\begin{aligned} \ln[f_{itr}(Y_{itr} = r | Y_{r,j,t})] &= \sum_{r=1}^{R-1} Y_{itr} \ln p_{itr} + \left(1 - \sum_{r=1}^{R-1} Y_{itr}\right) \ln(1 - p_{itr}) \\ &= \sum_{r=1}^{R-1} Y_{itr} \ln\left(\frac{p_{itr}}{1 - p_{itr}}\right) + \ln(1 - p_{itr}), \end{aligned} \quad (3)$$

where $1 - p_{itr} = p_{itR}$ and $\ln\left(\frac{p_{itr}}{1-p_{itr}}\right) = \eta_{rit}$. Substituting into equation (3),

$$\ln[f_{itr}(Y_{itr} = r|Y_{jtr})] = \sum_{r=1}^{R-1} Y_{itr}\eta_{rit} - \ln(1 + \sum_{r=1}^{R-1} \exp\{\eta_{itr}\}). \tag{4}$$

The joint probability is specified by the Hammersley-Clifford theorem, which states that the joint probability of the spatial-temporal process will be well defined. The following derivation with respect to the joint distribution of the nominal multinomial is provided.

Suppose that the probability structure is dependent only upon contributions from cliques containing no more than two sites. The negpotential function (Besag, 1972) with respect to the sites is

$$Q(Y) = \sum G(Y_{itr}) + \sum G_{itjr}(Y_{itr}, Y_{jtr}), \tag{5}$$

where $G(Y_{itr}) \equiv 0$ unless the sites i and j are neighbors due to the pairwise-only dependencies between sites. In this case, the G functions are defined as

$$G_{itr}(Y_{itr}) = \ln\left(\frac{f_{itr}(Y_{itr}|\boldsymbol{Y}^*_{jtr})}{f_{itr}(Y^*_{itr}|\boldsymbol{Y}_{jtr}*)}\right), \tag{6}$$

and

$$G_{ijtr}(Y_{itr}, Y_{jtr}) = \ln\left(\frac{f_{itr}(Y_{itr}|Y_{jtr}, \boldsymbol{Y}^*_{itr,jtr}) f_{itr}(Y^*_{itr}|\boldsymbol{Y}^*_{itr})}{f_{itr}(Y^*_{itr}|Y_{jtr}, \boldsymbol{Y}^*_{itr,jtr}) f_{itr}(Y_{itr}|\boldsymbol{Y}^*_{itr})}\right). \tag{7}$$

In the multinomial model, when one response is selected all others are 0, which implies that $\boldsymbol{Y}^* = \boldsymbol{0}$. Substituting the conditional distribution, equation (4), into equations (6) and (7), the G functions for the nominal multinomial model are

$$G(Y_{itr}) = \sum_{r=1}^{R-1} Y_{itr}\left(\sum_{k=0}^{K} \theta_{kr}X_{itrk} + \sum_{s=1}^{S} \frac{1}{d_i}\gamma_{sr}Y_{i,t-s,r}\right) \tag{8}$$

and

$$G_{ijr}(Y_{itr}, Y_{jtr}) = \frac{1}{2}\left[\sum_{r=1}^{R-1} Y_{itr}\left(\sum_{l=1}^{L} \frac{1}{d_i}\phi_{lr} \sum_{j\in\mathcal{N}_i^{(l)}} Y_{jtr}\right)\right] \tag{9}$$

Hence, using equations (8) and (9), the negpotential function, eqn (5), becomes

$$Q(\boldsymbol{Y}) = \sum_{i=1}^{I}\sum_{r=1}^{R-1} Y_{itr}\left(\sum_{k=0}^{K} \theta_{kr}X_{itrk} + \sum_{s=1}^{S} \frac{1}{d_j}\gamma_{sr}Y_{i,t-s,r} + \frac{1}{2}\left[\sum_{l=1}^{L} \frac{1}{d_j}\phi_{lr} \sum_{j\in\mathcal{N}_i^{(l)}} Y_{jtr}\right]\right). \tag{10}$$

This implies that the joint probability distribution is

$$\begin{aligned} p(\boldsymbol{Y}_t : \boldsymbol{Y}'_t) &= \frac{\exp(Q(\boldsymbol{Y}_t))}{\sum_{Z\in\Omega}\exp(Q(\boldsymbol{Z}_t; \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma}))} \\ &= c(\boldsymbol{Y}_{t-1}, \cdots, \boldsymbol{Y}_{t-S} : \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma})^{-1} \times \sum_{i=1}^{I} Y_{itr}\eta_{itr} \end{aligned}$$

3.3 Ordinal Multinomial Spatial-Temporal Autologistic Model Specification

The ordinal multinomial spatial-temporal autologistic model is used when the data are ordered categorical responses. The data from R different categories a spatial lattice are measured repeatedly over time, where the R^{th} category is greater than all the other categories. The spatial-temporal ordinal multinomial response has a conditional distribution from an ordinal multinomial distribution. In this model, the probabilities are ordered in increasing probability where $P(Y_{it} \leq R)$ has the following two properties; $P_{\leq R}$ =1 and $p_{\leq 1} \leq p_{\leq 2} \leq \cdots \leq p_{\leq R}$. Assume the conditional probability

$$p_{rit} = P(Y_{itr} = r | Y_{r,j,t} : J \in \mathcal{N}_i, \boldsymbol{Y}_{r,t'} : t' = t-1, \cdots, t-S),$$

has a cumulative logit link where S is the most recent time. For each time, t, it is assumed that the response variable follows a Markov random field under a specified spatial neighborhood. The euclidean distance for each site and category, denoted d_j, represent the magnitude of the distance between each i^{th} site any other site j.The weight for the spatial component is added due to the irregularity of the spatial frame. The temporal term is assigned a weight due to the varying spatial frame each year. Since the spatial frame moves for each time, the previous time responses, $Y_{i,t-s}$ are found based upon the minimum euclidean distance to the current years response. Previous years observations that are closest to the current time point are assigned higher weights and those further away are assigned lower weights. Let η_{rit} denote the systematic component,

$$\eta_{itr} = \sum_{r=1}^{R-1} \sum_{k=0}^{K} \theta_{kr} X_{itrk} + \frac{1}{2} \left[\sum_{l=1}^{L} \phi_{lr} \sum_{j \in \mathcal{N}_i^{(l)}} \frac{1}{d_j} Y_{jtr} \right] + \sum_{s=1}^{S} \frac{1}{d_j} \gamma_{sr} Y_{i,t-s,r},$$

where $\theta_{0r}, \cdots, \theta_{Kr}$ are the regression coefficients, $\phi_{1r}, \cdots, \phi_{Lr}$ are the spatial autoregressive coefficients, and $\gamma_{1r}, \cdots, \gamma_{Sr}$ are the temporal autoregressive coefficients. The probability of success for the r^{th} category is defined as

$$p_{itr} = \frac{\exp\{\eta_{itr}\}}{1 + \exp\{\eta_{itr}\}} - \sum_{k=1+r}^{R-1} p_k.$$

The cumulative logit can be applied to the conditional probabilities,

$$\begin{aligned} \text{logit}(p_{\leq rit}) &= \ln\left(\frac{p_{\leq rit}}{1 - p_{\leq rit}}\right) \\ &= \sum_{r=1}^{R-1} \sum_{k=0}^{K} \theta_{kr} X_{ritk} + \frac{1}{2} \left[\sum_{l=1}^{L} \phi_{lr} \sum_{j \in \mathcal{N}_i^{(l)}} \frac{1}{d_j} Y_{jtr} \right] + \sum_{s=1}^{S} \frac{1}{d_j} \gamma_{sr} Y_{i,t-s,r}, \end{aligned}$$

for $r = 1, \cdots, R$.

3.3.1 Joint Probability Distribution

The model is similar to the nominal multinomial model, however, the slopes do not depend on the r categories. A general assumption of the ordinal multinomial model is that each model for the r categories has equal slopes otherwise termed the proportional odds model. Suppose that the probability structure is dependent only upon contributions from cliques containing no more than two sites then the negpotential function (Besag 1972) is defined the same way as in the nominal multinomial models Eqn (8), (9) and (10).

The negpotential function, eqn (5) is defined using equations (8) and (9) as

$$Q(\boldsymbol{Y}) = \sum_{i=1}^{I}\sum_{r=1}^{R-1} Y_{itr}\left(\sum_{k=0}^{K}\theta_{kr}X_{itrk} + \sum_{s=1}^{S}\frac{1}{d_j}\gamma_{sr}Y_{i,t-s,r} + \frac{1}{2}\left[\sum_{l=1}^{L}\frac{1}{d_j}\phi_{lr}\sum_{j\in\mathcal{N}_i^{(l)}} Y_{jtr}\right]\right). \quad (11)$$

This implies that the joint probability distribution is

$$\begin{aligned} p(\boldsymbol{Y}_t : \boldsymbol{Y}'_t) &= \frac{\exp(Q(\boldsymbol{Y}_t))}{\sum_{Z\in\Omega}\exp(Q(\boldsymbol{Z}_t;\boldsymbol{\theta}))} \\ &= c(\boldsymbol{Y}_{t-1},\cdots,\boldsymbol{Y}_{t-S}:\boldsymbol{\theta})^{-1}\times\sum_{i=1}^{I}Y_{itr}\eta_{itr} \end{aligned} \quad (12)$$

4. Statistical Inference

Statistical inference is based upon the likelihood function of the conditional distribution. The conditional probability has been specified in section 2 as

$$p(\mathbf{Y}_t|\mathbf{Y}'_t : t-1,\cdots,t-S) = c(\mathbf{Y}_{t-1},\cdots,\mathbf{Y}_{\mathbf{t-S}};\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\gamma})^{-1}\times\exp\{\eta_{itr}\}$$

The conditional distribution depends on either the Bernoulli, for the binary case where $R = 2$, or the Multinomial distribution, where the response vector $[\mathbf{Y}_{S+1},\cdots\mathbf{Y}_T]^T$ is conditioned on the response at the first S time points, $[\mathbf{Y}_1,\cdots\mathbf{Y}_S]^T$. Since each time point has to be taken into account, each time point is considered independent as stated in the Hammersley-Clifford Theorem (Hammersley and Clifford, 1971).

The likelihood function of the binary, nominal multinomial and ordinal multinomial responses for $\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\gamma}$ is found as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta},\boldsymbol{\phi},\gamma) =& \mathcal{L}(\boldsymbol{\theta},\boldsymbol{\phi},\gamma,\mathbf{Y}_S+1,\cdots\mathbf{Y}_T|\mathbf{Y}_1,\cdots\mathbf{Y}_S) = \\ &= \prod_{i=S+1}^{T}[c(\mathbf{Y}_{t-1},\cdots,\mathbf{Y}_{t-S},\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\gamma})]^{-1}\times\prod_{i=S+1}^{T}\exp\left\{\sum_{r=1}^{R}Y_{it}\eta_{itr}\right\} \\ &= \left[\prod_{i=S+1}^{T}c(\mathbf{Y}_{t-1},\cdots,\mathbf{Y}_{t-S},\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\gamma})\right]^{-1}\times\exp\sum_{i=S+1}^{T}\sum_{r=1}^{R}Y_{it}\eta_{itr}. \end{aligned}$$

The log likelihood for $\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\gamma}$ is

$$\begin{aligned} \ell(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\gamma}) &= \log\left[\prod_{i=S+1}^{T}c(\mathbf{Y}_{t-1},\cdots,\mathbf{Y}_{t-S},\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\gamma})\right]^{-1}\times\exp\left\{\sum_{i=S+1}^{T}\sum_{r=1}^{R}Y_{it}\eta_{itr}\right\} \\ &= -\log\left[\sum_{i=S+1}^{T}c(\mathbf{Y}_{t-1},\cdots,\mathbf{Y}_{t-S},\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\gamma})\right] + \sum_{i=S+1}^{T}\sum_{r=1}^{R}Y_{it}\eta_{itr}, \end{aligned} \quad (13)$$

where $\boldsymbol{Z_t}$ will be denoted as

$$\boldsymbol{Z_t} = \sum_{r=1}^{R}\sum_{i=1}^{I}\left(\sum_{k=0}^{K}X_{itk}Y_{it} + \frac{1}{2}\left[\sum_{l=0}^{L}\sum_{j\in\mathcal{N}_i^{(l)}}Y_{it}Y_{jt}\right] + \sum_{s=1}^{S}Y_{it}Y_{i,t-s}\right),$$

part of the systematic component. Let $\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma} = (\theta_{0r}, \cdots, \theta_{Kr}, \phi_{1r}, \cdots, \phi_{Lr}, \gamma_{1r}, \cdots \gamma_{Sr})^T$, where K is the number of explanatory variables, L the lth order neighbor and S the latest time. The log likelihood, equation (12), becomes

$$\ell(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma}) = -\log\left[\sum_{i=S+1}^{T} c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-S}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma})\right] + \sum_{i=S+1}^{T} \boldsymbol{Z_t}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma}). \quad (14)$$

The log likelihood is usually easier to estimate, however, due to the normalizing constant the log likelihood becomes difficult to approximate.The normalizing constant, $c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-S}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma})$, does not have a closed form so direct maximization of the likelihood, Eqn (13), requires an approximation of $c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-S}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma})$. The Monte Carlo Maximum Likelihood Method (MCML) will be used to estimate the parameters. The MCML method approximates an expectation from the sample mean of a function of simulated random variables. The simulated random variables will be the normalizing constants with respect to the temporal terms. The MCML process uses an iterative method that randomly samples from a specific probability distribution, the Bernoulli, or Multinomial likelihood function, through simulation to estimate the value of the parameters. The parameter values are then found by averaging the simulated estimates. By the law of large numbers, the estimated parameters from MCML will be close to the true parameter values.

The Monte Carlo Maximum Likelihood method uses a Markov chain, which takes randomly sampled values of a distribution. As the sample increases the actual value is identified from the mean of all the values. The samples are neither independent nor identically distributed; however the probability converges in distribution to the actual distribution as if the samples are iid (Geyer, 1992). Monte Carlo simulation it is a way of making sure that certain values will have more impact on the parameter being estimated, by the “important” values being sampled more frequently. In this way the variance of the parameter values, $\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma}$ can be reduced.

Importance sampling is a common method used in the case where the estimation is difficult. It is based upon the property that the likelihood of two ratios can be written as an expectation of a density. This changes the probability such that estimation is easier to calculate. A reference parameter can be used to reduce the variation in the model and thus provide more accurate measurements of the model parameters. It is essentially adding a value such that the expected value of ψ and P_t is 1. Let $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\gamma})$, the log likelihood function, $\boldsymbol{\Theta}$ is updated by using a importance sampling shift as follows

$$\ell(\boldsymbol{\Theta}) - \ell(\boldsymbol{\psi}) = \sum_{t=S+1}^{T} (\boldsymbol{\Theta} - \boldsymbol{\psi})' \boldsymbol{Z}_t - \sum_{t=S+1}^{T} \log \frac{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\Theta})}{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\psi})}, \quad (15)$$

where $\boldsymbol{\psi} = (\psi_0, \cdots, \psi_{K+L+S})'$ represents the regressive, spatial and temporal reference parameters.

Now, assuming that the difference in the log likelihood functions is approximately zero equation (15) can be rewritten as

$$0 = \sum_{t=S+1}^{T} (\boldsymbol{\Theta} - \boldsymbol{\psi})' \boldsymbol{Z}_t - \sum_{t=S+1}^{T} \log \frac{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\Theta})}{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\psi})}$$

$$\Rightarrow \sum_{t=S+1}^{T} \frac{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\Theta})}{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\psi})} = \sum_{t=S+1}^{T} \frac{\exp\{\boldsymbol{\Theta' Z_t}\}}{\exp\{\boldsymbol{\psi' Z_t}\}}$$

The likelihood of two ratios is an expectation of a density by importance sample so the ratio of the normalizing constants is

$$\frac{c(\mathbf{Y}_{t-1},\cdots,\mathbf{Y}_{\mathbf{t-S}};\mathbf{\Theta})}{c(\mathbf{Y}_{t-1},\cdots,\mathbf{Y}_{\mathbf{t-S}};\boldsymbol{\psi})} = E_{\psi}\left[\frac{\exp\{\boldsymbol{\Theta'Z_t}\}}{\exp\{\boldsymbol{\psi'Z_t}\}}\right] \tag{16}$$

for a time point t and the expectation is with respect to the conditional distribution evaluated at the reference parameter.

Monte Carlo estimation techniques will be used on the ratio of the normalizing constants, equation (16), as an expectation can be found and it is known that the normalizing constants do not have a closed form and have to be approximated as

$$\frac{1}{M}\sum_{m=1}^{M}\frac{\exp\{\boldsymbol{\Theta'Z_t^{(m)}}\}}{\exp\{\boldsymbol{\psi'Z_t^{(m)}}\}},$$

where $m = 1,\cdots M$, the number of iterations using a Monte Carlo Gibbs sampler, and $\boldsymbol{Z}_t^{(m)}$ is the m^{th} set of Monte Carlo samples of the response vector, $\boldsymbol{Y}_t$. Since the normalizing constant $\boldsymbol{\psi}$ is unknown a Gibbs sampler is used to generate Monte Carlo samples according to the full conditional distribution, as specified by equation (12).

Using the log likelihood ratio via importance sampling and Monte Carlo estimation, an approximation of the log likelihood ratio, equation (15) can be found as

$$\begin{aligned}\ell(\mathbf{\Theta}) - \ell(\boldsymbol{\psi}) &= \sum_{t=S+1}^{T}(\mathbf{\Theta}-\boldsymbol{\psi})'\boldsymbol{Z}_t - \sum_{t=S+1}^{T}\log\frac{c(\mathbf{Y}_{t-1},\cdots,\mathbf{Y}_{\mathbf{t-S}};\mathbf{\Theta})}{c(\mathbf{Y}_{t-1},\cdots,\mathbf{Y}_{\mathbf{t-S}};\boldsymbol{\psi})}\\ &\approx \sum_{t=S+1}^{T}(\mathbf{\Theta}-\boldsymbol{\psi})'\boldsymbol{Z}_t - \sum_{t=S+1}^{T}\log\left[\frac{1}{M}\sum_{m=1}^{M}\frac{\exp\{\boldsymbol{\Theta'Z_t^{(m)}}\}}{\exp\{\boldsymbol{\psi'Z_t^{(m)}}\}}\right]\\ &\equiv \ell(\mathbf{\Theta};\boldsymbol{\psi}). \end{aligned} \tag{17}$$

The log likelihood, $\ell(\boldsymbol{\psi})$ is free of $\mathbf{\Theta}$ so it is possible to maximize equation (17) with respect to $\mathbf{\Theta}$ and obtain the maximum likelihood estimators $\hat{\mathbf{\Theta}}$ of $\mathbf{\Theta}$.

In order to estimate the reference parameter $\boldsymbol{\psi}$, Maximum Pseudo-Likelihood (MPL) techniques can be used (Zhu, 2005). In the case that the MPL is not close to the MLE, the MLE may become difficult to obtain. An alternative method is to use a stochastic approximation algorithm as specified by Gu and Zhu (2001).

5. Example

In this section, the modified spatial-temporal autologistic model and nominal multinomial models from section 2-4 are applied to the Rocky Mountain Forest Service Data collected by the United States Forest Service. The data are composed of polygons variable in size that record the presence of various types of damage causing agents in North Central Rocky Mountain Colorado area from 2005-2009. The number of polygons, represented as the i sites in the dataset change each year. That is, for a response Y_{itr}, $i = 918, 1254, 1314, 713, 877$ for $t = 2005, 2006, 2007, 2008, 2009$, respectively. The binary response represented as $Y_{it} = 1$ for the presence of bark beetle and $Y_{it} = 0$ for the presence of multiple types of damage (Five Needle Pine Decline and Subalpine Fir Mortality) outbreak at the i^{th} polygon and the t^{th} year. The nominal multinomial response, Y_{itr}, is represented by three categories, $R = 3$, multiple types of damage, mountain pine beetle, and spruce beetle presence.

Sites were considered neighbors if the corresponding sites are within 5 km of each other, based upon the average maximum distance that a bark beetle can move from location to location. For model parameter estimation a Gibbs sample with a total of 5,000 Monte Carlo samples with a burn in period of 100 are used. The trace plots of the Monte Carlo samples of the response, Y_{itr}, indicated that an adequate amount of samples were used.

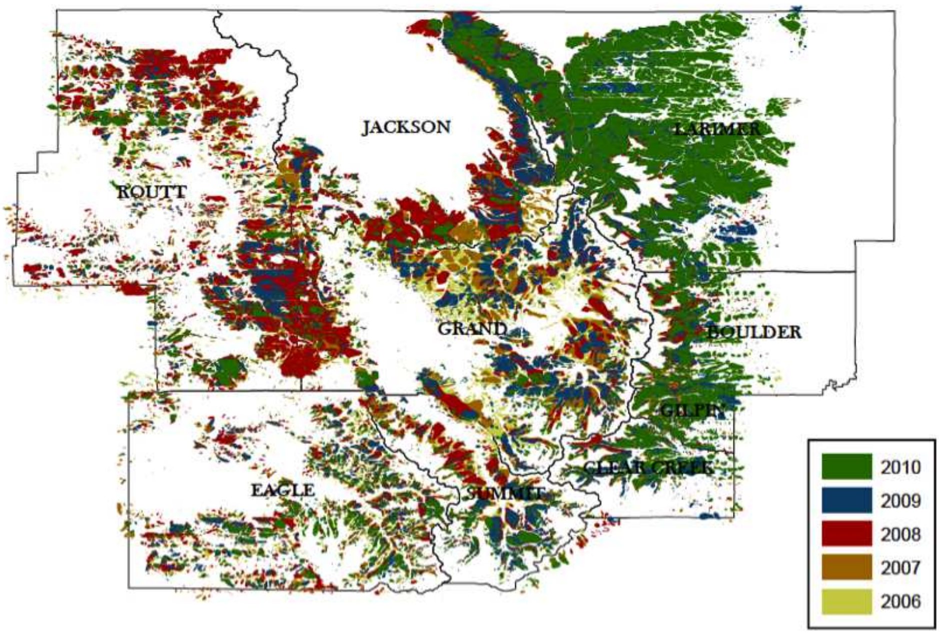

Figure 1: Mountain Pine Beetle and Other Damage

Figure 1 displays the presence of bark beetles in North Central Colorado. Figure 2 is a time-series plot of the number of sites that experienced either a bark beetle or multi-damage outbreak in a year, for each of the years 2005 to 2009.

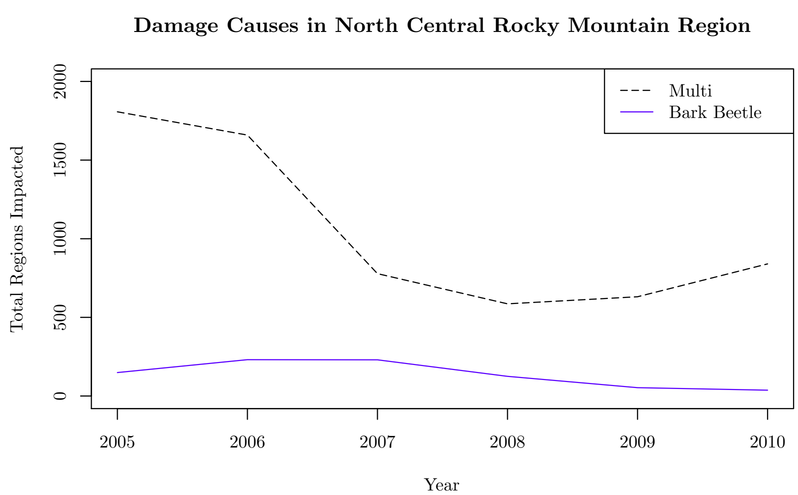

Figure 2: Mountain Pine Beetle and Other Damage

In Figure 3 a time-series plot of the number of sites that experienced an outbreak of either Mountain Pine Beetles (MPB), Spruce beetles, or Multiple types of damages (Multi) within Larimer County, a subset of the North Central Colorado dataset. The time-series displays that when there is a peak in multiple types of damage there are fewer bark beetle outbreaks whereas when there are fewer outbreaks of multiple types of damage then the bark beetle outbreaks tend to be high. This might indicate that the trees are more susceptible to outbreaks after a disease or other type of infestation.

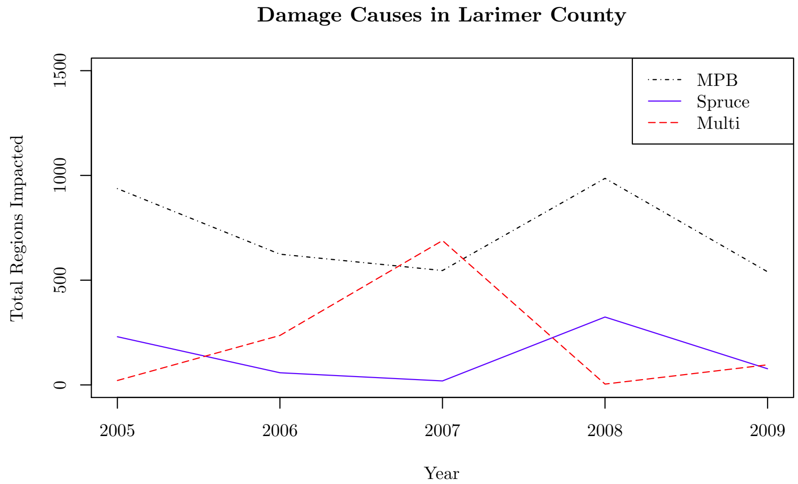

Figure 3: Mountain Pine Beetle and Other Damage

The explanatory variable that was used for the binary model is dead trees per acre (TPA), a continuous variable, values ranging from 0.1 to 166.67 dead trees per acre. The binary model uses the modified spatial-temporal autologistic model specified in section 2.3 and described in section 3, the MLE of the parameters of the model and the standard errors were computed (Table 1). For the binary model, there is a strong negative relation between the Multiple Types of Damage and Trees per Acre. This indicates that the higher the dead trees per acres (TPA) the lower the probability the outbreak is from Multiple Damage Causing Agents verses Bark Beetle outbreaks. There is also strong evidence of positive spatial dependence at the first-order neighborhood as well as positive temporal dependence.

Table 1: Multiple Damage Causing Agents (Binary) Estimates

	Parameter	Estimate	SE
Intercept	θ_0	-81.14	8.76
X_1 TPA	θ_1	-2067.06	24.77
Spatial	ϕ	696.15	17.52
Temporal	γ	704.74	17.52

The nominal multinomial model used dead trees per acre and tree type (Pine, Fir), an indicator variable as explanatory variables. The nominal multinomial spatial-temporal model specified in section 3.2, was used to estimate the MLE of the parameters of the model and the standard errors (Table 2). The results are comparable to the binary estimates as there is still strong evidence of positive spatial dependence at the first-order neighborhood as well as positive temporal dependence for each of the categories compared to the baseline Multi-Damage outbreaks. However, it is important to note that the magnitude of the spatial and temporal dependence is different for the Spruce Beetle and the Mountain Pine Beetle (MPB) outbreaks. There is a higher degree of temporal and spatial dependence for the MPB outbreaks verses the Spruce Beetle. The model also indicates the higher the dead trees per acres (TPA) the lower the probability of outbreak of Multi-Damage outbreaks, consistent with the binary model. The unusually low standard errors for the temporal dependence of the Multi-Damage verses the MPB indicates a lack of variability of outbreaks across time. The low standard error for the spatial dependence of Multi-Damage verses the Spruce beetle indicates that there is little spatial variability in the spatial frame which requires further investigation.

Table 2: Nominal Multinomial Model Estimates (Baseline: Multi-Damage)

	Parameter	Estimate	SE
Intercept (Spruce)	θ_{01}	-203.74	26.76
Intercept (MPB)	θ_{02}	-221.0	46.34
X_{11} Pine (Spruce)	θ_{11}	697.17	1.28
X_{12} Pine (MPB)	θ_{12}	781.03	46.34
X_{21} TPA (Spruce)	θ_{21}	-3087.24	35.4
X_{22} TPA (MPB)	θ_{22}	-3266.29	139
Spatial (Spruce)	ϕ_{11}	701.87	< 0.0001
Spatial (MPB)	ϕ_{12}	802.02	113.5
Temporal (Spruce)	γ_{11}	401.34	26.76
Temporal (MPB)	γ_{12}	581.61	< 0.0001

6. Discussion and Conclusions

The spatial-temporal autologistic regression model proposed by Zhu et al. (2005) has been extended to to multinomial responses as well as models that exist on a spatial grid that change at each time point. The modified spatial-temporal and multinomial autologistic models as specified in sections 2.3, 3.2 and 3.3 deal with the difficulty of matching up previous years observations to current observations by selecting the previous years observation that is closest to the current years observation. The models also assign weights for the spatial and temporal components based upon the distance so the observations that are closest have the most weight and those further away have the least weight due to the irregularity of the spatial grid. A real example has been used to demonstrate the methodology of the models. In both the binary and multinomial models there existed positive spatial and temporal dependence. The multinomial model, however, displayed that groups or categories can have different spatial dependence structures and should be accounted for in the model. In this type of situation it would be less appropriate to use a single spatial component as it would ignore the individual categories spatial dependence.

The estimates for both the binary and multinomial models were found to be sensitive to initial values in the MCML procedure. In the binary case, there was an additional sensitivity to indicator variables as it caused instability in the model. Further investigation into the initial value estimates and indicator sensitivity for the models will be conducted. For future work, different models and estimation techniques should be compared.

REFERENCES

Besag, J. (1972), "Nearest-Neighbour Systems and the Auto-logistic Model for Binary Data," *Journal of the Royal Statistical Society,* Series B, 34, 75-83.

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society,* Series B, 36, 192-236.

Besag, J. (1975), "Statistical Analysis of Non-lattice Data," *The Statistician*, 24, 179-195.

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

Cressie, N., and Wikle, C. (2011), *Statistics for Spatio-Temporal Data*, Hoboken, NJ: Wiley.

Geyer, C.J., and Thompson, E.A. (1992), "Constrained Monte Carlo Maximum Likelihood for Dependent Data," (with discussion), *Journal of the Royal Statistical Society*, Series B, 54, 657-699.

Gu, M.G., and Zhu, H.T. (2001), "Maximum Likelihood Estimation for Spatial Models by Markov Chain Monte Carlo Stochastic Approximation," *Journal of the Royal Statistical Society*, Series B, 63, 339-355.

Gumpertz, M. L, Graham, J.M and Ristaino, J.B. (1997), "Autologistic Model of Spatial Pattern of Phytophthora Epidemic in Bell Pepper: Effects of Soil Variables on Disease Presence," *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 131-156.

Hammersley, J. M. and Clifford, P. (1971), "Markov Fields on Finite Graphs and Lattices," (Unpublished).

Huffer, F.W., Wu, H. (1998), "Markov Chain Monte Carlo for Autologistic Regression Models with Application to the Distribution of Plant Species," *Biometrics*, 54, 509-524.

Zheng, Y., and Zhu, J. (2008), "Markov Chain Monte Carlo for a Spatial-Temporal Autologistic Regression Model," *Journal of Computational and Graphical Statistics*, 17, 123-137.

Zhu, J., Huang, H., and Wu, J. (2005), "Modeling Spatial-Temporal Binary Data Using Markov Random Fields,"*Journal of Agricultural, Biological, and Environmental Statistics*, 10, 212-225.

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.