Modeling Fire Frequency with Negative Binomial Spatial Regression Models

Yolanda Li* (Victoria), Steve Taylor (Pacific Forestry), and Farouk Nathoo* (Victoria)

* Department of Mathematics and Statistics, University of Victoria, Victoria, B.C.

University of Victoria

1. Introduction

Fire plays an important role in Canada's forested ecosystems. It helps to maintain forest health and diversity; however, it can have undesirable negative effects on public safety, health and property. Forest fire has numerous causes such as dry weather and human behavior. Moreover, large areas of dead forest due to mountain pine beetle outbreak in British Columbia may lead to more severe wildfires. Therefore, it is important to study the distribution of forest fire and its relationship to these factors.

2. Data Structure

Study Area: British Columbia (divided into $I = 1712$ homogeneous grid cells)

Response: N_i, total fire counts in each region over 44 years

Covariates (regional specific): area affected by mountain pine beetle (MPB) outbreak, area of forest covering, area of pine leading stands, number of roadways and drought climate

Spatial Information: an adjacency matrix W, coding adjacencies of partitioning grid cells

3. Exploratory Analysis

Figure 1: Total Fire Counts

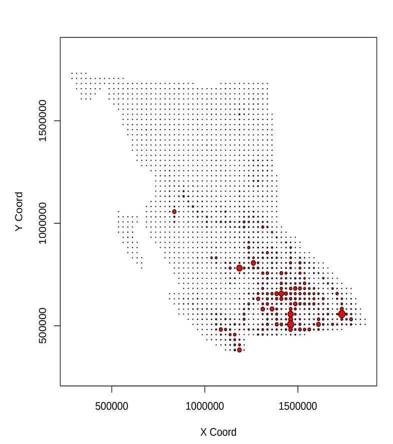


Figure 2: MPB Outbreak

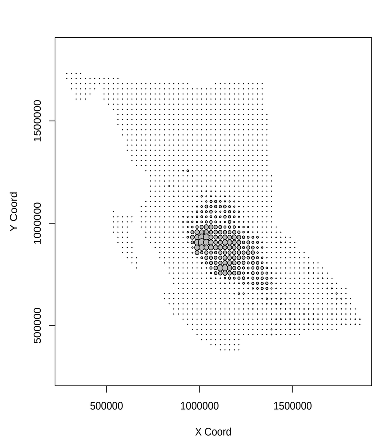


Figure 3: Covariates

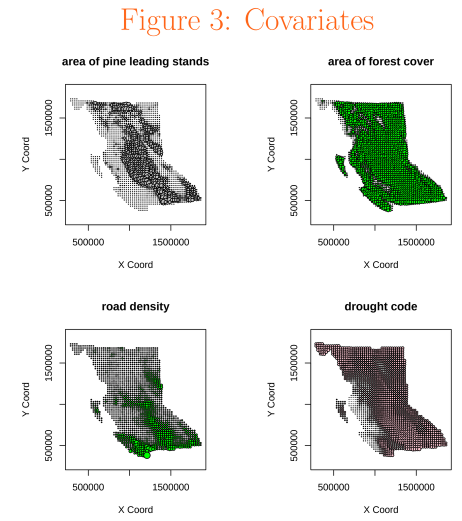
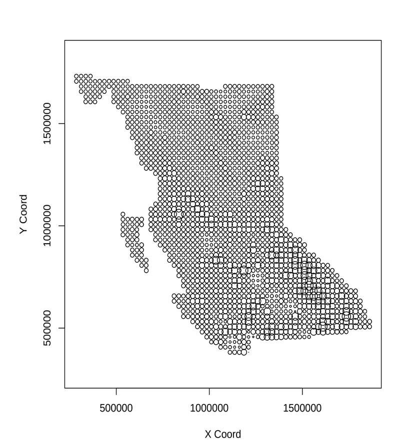


Figure 4: Deviance of Residuals



- Figure 1-3: spatial distribution of response and covariates (larger circle indicates a higher value)
- Figure 4: deviance of residuals under the standard log-linear Poisson regression model → not randomly distributed → data are spatially correlated
- Overdispersed count data (Alexander *et al.*, 2000)

4. Overdispersed Spatial Count Model

$$N_i \mid \lambda_i, a \sim \text{Negbin}(\lambda_i, a) \qquad i = 1, ..., I$$

where N_i is the total fire count in region i, λ_i describes the mean and $a \geq 0$ is the dispersion parameter ($a \to 0$ yields the Poisson(λ_i)) (Lawless, 1987).

Under a Bayesian hierarchical framework, we can use an equivalent Poisson-Gamma mixture representation:

$$N_i \mid \nu_i, \lambda_i \sim \text{Poisson}(\nu_i \lambda_i)$$

$$\log(\lambda_i) = \boldsymbol{\beta}^T \mathbf{X}_i + b_i$$

$$\mathbf{b} = (b_1, ..., b_n) \text{ are spatial random effects, } \mathbf{b} \mid \sigma^2_{\mathbf{b}} \sim \text{CAR}(\sigma^2_{\mathbf{b}})$$

$$\nu_i \mid a \overset{i.i.d}{\sim} \text{Gamma}(\tfrac{1}{a}, a) \text{ accommodate extra Poisson variation}$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients and $\mathbf{X}_i$ is a vector of covariates for region i. The conditional autoregressive model (CAR) (Besag, 1974) employed for $\mathbf{b}$ accounts the spatial effect of region i conditionally on its neighboring regions based on the adjacency matrix W.

Identifiability: Note the random effects b_i and ν_i are not uniquely identified; however the sum $\alpha_i = b_i + \log(\nu_i)$ is identified. In order to deal with the weak identifiability issue among ν_i and b_i, we monitor α_i instead.

Model estimation of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, a, \sigma^2_b)$ given the data is carried out using a Markov Chain Monte Carlo Algorithms programed in Matlab.

Posterior Summaries (NegBin with CAR)

	Mean	Std. Deviation	2.50%	97.5%
β_0 (Intercept)	2.6879	0.0145	2.6587	2.7157
β_1 (MPB)	−0.0335	0.0553	−0.1446	0.0746
β_2 (DCmean)	0.8457	0.0811	0.6941	1.0119
β_3 (forest)	0.8316	0.0377	0.7577	0.9058
β_4 (pine)	−0.3491	0.0418	−0.4313	−0.2655
β_5 (road)	0.4220	0.0425	0.3376	0.5058
$\tau = 1/\sigma^2_{\mathbf{b}}$ (CAR precision)	0.3221	0.0143	0.2949	0.3508
a (overdispersion)	0.0002	0.0010	0.0000	0.0020

p_D and DIC

model	p_D	DIC
Poisson(λ_i), $log(\lambda_i) = \boldsymbol{\beta}^T X_i$	6	95292
Poisson(λ_i), $log(\lambda_i) = \boldsymbol{\beta}^T X_i + b_i$, $b_i \overset{i.i.d}{\sim} \text{N}(0, 1/\tau)$	1476	10849
Poisson(λ_i), $log(\lambda_i) = \boldsymbol{\beta}^T X_i + b_i$, $\mathbf{b} \sim \text{CAR}(1/\tau)$	1315	10697
NegBin(λ_i, a), $log(\lambda_i) = \boldsymbol{\beta}^T X_i$	1403	10710
NegBin(λ_i, a), $log(\lambda_i) = \boldsymbol{\beta}^T X_i + b_i$, $\mathbf{b} \sim \text{CAR}(1/\tau)$	1295	10678

Comment: The Negative Binomial model with spatial random effects is preferred since this model has the smallest DIC.

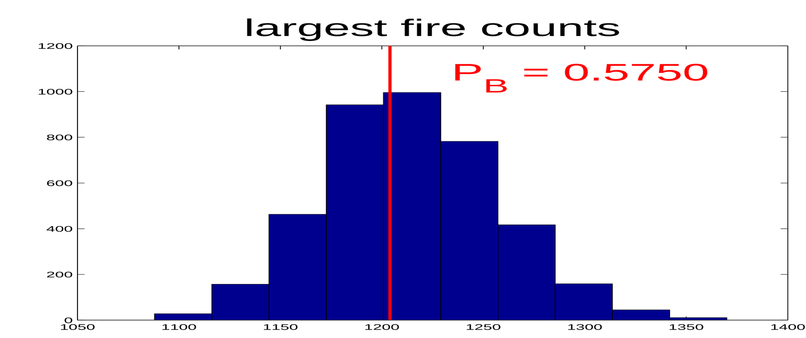
5. Goodness of Fit

Posterior Predictive Checking: (Gelman *et al.*, 2004)

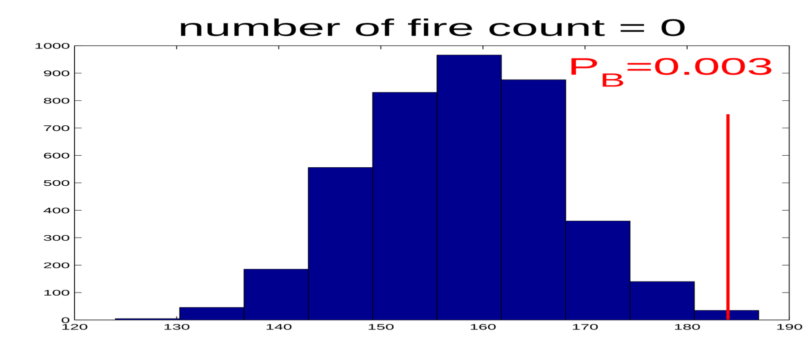
- J=4000 replicated data sets, N^{rep}, generated from posterior predictive distribution
- Posterior Predictive P-value is defined as

$$P_B = \Pr(\text{T}(\mathbf{N}^{rep}, \boldsymbol{\theta}) \geq \text{T}(\mathbf{N}, \boldsymbol{\theta}) \mid \mathbf{N})$$

where T(.) is a test quantity that is a scalar summary of parameters and data

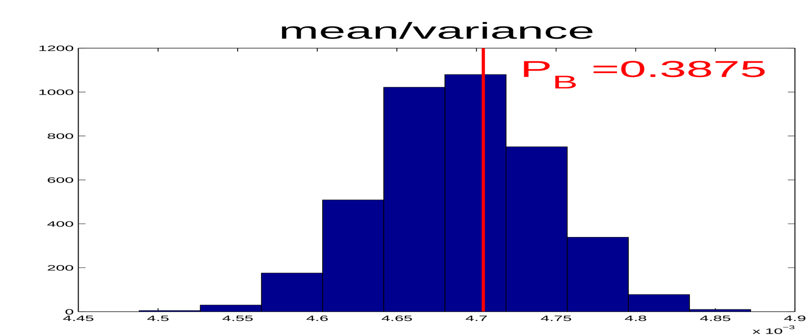


Extremes: T = the largest value of fire counts in each data set



Heavy Zeros: $T = \sum_{i=1}^{I} I(N^{rep}_{ij} = 0)$

I(.) is the indication function



Overdispersion: $T = \frac{\text{mean}(\mathbf{N}^{rep}_j)}{\text{variance}(\mathbf{N}^{rep}_j)}$

Comment: No evidence of lack of fit for extremes and overdispersion aspects; however, not adequately detail with zeros.

6. Future Work

- Use zero heavy Poisson Mixture (Hougaard *et al.*, 1997) to deal with the heavy tail property of the data.

  $f^*(y \mid \lambda, \rho) = \rho f(y \mid \lambda) + (1 - \rho) I\{y = 0\}$ where $\rho \in [0, 1]$, $f(y \mid \lambda)$ is the Poisson pmf

- Move towards both spatial and temporal modeling.

References:

- Alexander, N., Moyeed, R. and Sander, J. (2000). Spatial modeling of individual-level parasite counts using the negative binomial distribution. *Biostatistics* **1**, 453-463.
- Besag, J. (1974). Spatial interaction and the statistical analysis of life systems (with discussion). *Journal of the Royal Statistical Society, Series B* **48**, 192-236.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd edn. New York: CRC.
- Hougaard, P., Lee, T. M. and Whitmore, A. G. (1997). Analysis of overdispersed count data by mixtures of poisson variables and poisson processes. *Biometrics* **53**, 1225-1238.
- Lawless, F. J. (1987). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics* **15**, 209-225.

- Research supported by grants from NSERC and GEOIDE.