Bayesian Hierarchical Modelling of Wildfires

Antónia Amaral Turkman (antonia.turkman@fc.ul.pt), K.F. Turkman (kfturkman@fc.ul.pt)

Lisbon University, Faculty of Sciences and CEAUL, Portugal

projects FCT/POCTI/2010 and PTDC/MAT/64353/2006 - FCT/PORTUGAL

Introduction

- Policy responses for local and global fire management as well as international green-gas inventories depend heavily on the proper understanding of the annual fire extend as well as its spatial variation across any given study area.
- We propose Bayesian methods to model jointly the probability of ignition and fire sizes.
- The data set on which we base our models and results consists of annual observations of several meteorological and topological explanatory variables, together with the percentage of land burned over a grid with resolution of 1 degree across Austalia and New Zealand.
- Our models and conclusions bring improvements on the results reported by Russell-Smith *et al.* (2007) based on a similar data set.

Data

The following covariates and response variable were observed in each pixel:

- X_1 :A line identifier (1-750 for 1998 ; 1-6750 for 1999-2006, from the repeated 750 grid cells for each of the 9 years)
- X_2 latitude of the middle of each 1^o grid cell
- X_3 longitude of the middle of each 1^o grid cell
- X_4: Dry Season Severity from TRMM satellite data (native resolution 0.25^o, aggregated to 1^o)
- X_5: Precipitation over the wet season (or growing season) from TRMM satellite data
- X_6 Human footprint (static)
- X_8 Tree landcover (from 0 to 1, 1 being 100%)(static)
- X_9: Grass and Shrubs landcover (static)
- X_{10} Agricultural Landcover (static)
- X_{11} Maximum number of consecutive dry pentads (pentad=5 days) over the corresponding year.
- Y: Estimated Burned Fraction over the year (dependent variable)

There are $n = 750$ pixels, each having $N = 9$ repetitions from 1998 to 2006.

Transformations

The data set immediately reveals some characteristics which need to be taken into account while modeling.

1. The data on the dependent variable Y, the percentage of land burned, contain unusually high number of 0's, that is, the data is zero-inflated.
2. Non-zero observations are highly skewed to the left with a fairly heavy right tail.
3. Spatial and temporal dependencies are too strong to assume uncorrelated error structure.

Several transformations for Y^+, the positive percentage of burned area, were considered.
The best fit was found for a normal approximation for $\log(\frac{Y^+}{1-Y^+})$.

Notation

Let

- $$Z(i,t) = \begin{cases} \log(\frac{Y(i,t)}{1-Y(i,t)}), & \text{if } Y(i,t) \neq 0, \\ 0, & \text{if } Y(i,t) = 0. \end{cases}$$
- $$R(i,t) = \begin{cases} 1, & \text{if there is a fire in pixel} i \text{ at instant } t; \\ 0, & \text{otherwise.} \end{cases}$$
- $\mathbf{X}(i,t) = (X_1(i,t), ..., X_k(i,t))$ represent the covariates.
- $p_1(i,t) = P[R(i,t) = 1]$ represent the probability that a fire is ignited at the pixel i, during year t.
- $\mathbf{V}(i,t) = (V_0(i,t), V_1(i,t))$ be a latent bivariate spatio-temporal process. Assume that V_0 is the latent process representing the unobserved explanatory variables having influence on the ignition process, whereas V_1 is the latent process representing the unobserved explanatory variables having influence on the fire sizes.
- $\mathbf{Z} = (Z(i,t), i = 1,2,...,750, t = 1,2,..,9)$, $\mathbf{X} = (\mathbf{X}(i,t), i = 1,2,...,750, t = 1,2,..,9)$, $\mathbf{R} = (R(i,t), i = 1,2,...,750, t = 1,2,..,9)$, $\mathbf{V} = (\mathbf{V}(i,t), i = 1,2,...,750, t = 1,2,..,9)$, $\boldsymbol{\Theta}$ be all the (random) model parameters to be defined later.

Likelihood

First level; Likelihood

We observe $\mathbf{Z}$ and $\mathbf{R}$ and we have the following model for the conditional likelihood

$$f(\mathbf{Z}, \mathbf{R}|\mathbf{X}, \mathbf{V}, \boldsymbol{\Theta}) = f(\mathbf{Z}|\mathbf{R}, \mathbf{X}, \mathbf{V}, \boldsymbol{\Theta}) f(\mathbf{R}|\mathbf{X}, \mathbf{V}, \boldsymbol{\Theta})$$

$$f(\mathbf{Z}, \mathbf{R}|\mathbf{V}, \mathbf{X}, \boldsymbol{\Theta}) = \prod_{t=1}^{9}\prod_{i=1}^{750} f\left(z(i,t), \mathbf{R}(i,t)|\mathbf{V}(i,t), \boldsymbol{\Theta}\right)$$

Assuming, as stated in the previous section, that for $0 < Y(i,t) < 1$, $Z(i,t)$ is normally distributed,we have

$$f(z(i,t)|\mathbf{R}, \mathbf{V}, \mathbf{X}, \boldsymbol{\Theta}) = \left[N(\mu(i,t), \sigma^2)\right]^{R(i,t)}.$$

where the conditional distribution of $R(i,t)$ is Bernoulli with $P[R(i,t) = 1] = p_1(i,t)$.

Link functions and latent processes

Link functions

$$\mu(i,t) = \beta_0 + \mathbf{X}(i,t)^T\beta + V_1(i) + \delta_1(t)$$

$$\log\frac{p_1(i,t)}{1-p_1(i,t)} = \eta_0 + \mathbf{X}(i,t)^T\boldsymbol{\eta} + V_2(i) + \delta_2(t),$$

where, **Latent processes**

$V_1(i) = \nu_1 W_0(i) + \nu_2 W_1(i)$, $\quad V_2(i) = W_1(i)$, with $W_0(i)$ and $W_1(i)$ independent CAR (conditionally autoregressive)models (Gelfand *et al.*, 2004).

$\delta_1(t)$ and $\delta_2(t)$, are temporal processes independent of each other, each one having the structure of a random walk process of order 1.

Prior specifications

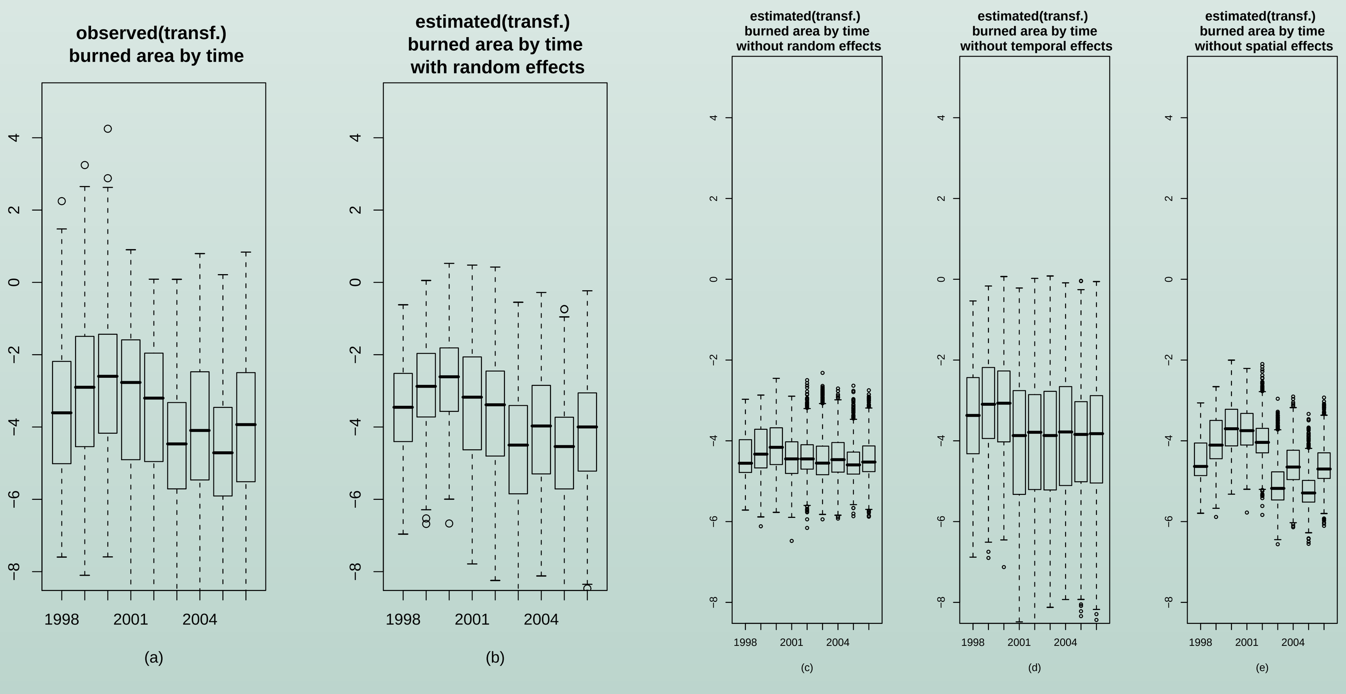
The model parameters and hyperparameters are $\tau = \frac{1}{\sigma^2}, \beta = (\beta_0, \beta_1, ..., \beta_k), \boldsymbol{\eta} = (\eta_0, \eta_1, ...\eta_k), \nu_1, \nu_2$, as well as the precisions of the latent processes involved. We assume diffuse independent priors for them, namely

1. β independent $N_0(0, 0.001)$, (note that the value 0.001 corresponds to the precision parameter of the normal distribution) $\boldsymbol{\eta}$ independent $N_0(0, 0.001)$.
2. $\nu_1 \sim N_0(1, 0.1)$, $\nu_2 \sim N_0(1, 0.1)$
3. $\sim Ga(0.01, 0.01)$ for precison parameters

Results and Conclusions

The model revealed that

- Climatological variables such as, dry season severity, precipitation over the wet season and the maximum number of consecutive dry pentads over the year, are factors that favor the increase in the fire size. Also higher percentage of tree land cover, or grass and shrubs land cover, increases risk of larger fires.
- Interactions between dry season severity with percentage of tree, grass and agriculture cover, as well as the amount of precipitation during the wet season with grass cover are important factors in explaining the size of the fire.
- Precipitation over the wet season, the maximum number of consecutive dry pentads over the year, percentage of tree, agricultural and grass cover, are very important factors in explaining ignition of fire, as well as the interaction between precipitation over the wet season and the percentage of grass cover. Dry season severity is an important factor for fire ignition when associated with the existence of trees.
- Model underestimates large fires. This suggests that large fires need to be modeled separately using the extreme value theory.
- Box plots below show that an improvement is obtained with the introduction of both the spatial and temporal random effects through the processes V_1 and δ_1. The temporal pattern, both in explaining the mean behavior of the percentage of burned area and the probability of ignition is well displayed



- The area under the ROC curve, which results from the logistic model for the probability of ignition, is 0.96. Predicting a fire when the probability of ignition is above 0.5, gives as result a sensitivity of 0.94, a specificity of 0.81, a positive predictive value of 0.92 and a negative predictive value of 0.85, revealing a very good performance.

References

1. Gelfand,A:, Schmit,A., Banerjee,S. and C. Firmans(2004) Nonstationary multivariate process modeling through spatially varying coregionalization. TEST VOL. 13, PP 263-312.
2. Russell-Smith,J. Yates,P., Whitehead,P.J, Smith,R.,Craig,R. Allan,G.E., Thackway,R., Frakes,I.,Cridland,S., Meyer,C.P. and A.M. Malcolm (2007) Bushfires 'down under': patterns and implications of contempory Australian landscape burning. International Journal of Wildland Fires, 16,361-377.