



7º CONGRESO FORESTAL ESPAÑOL

**Gestión del monte: servicios
ambientales y bioeconomía**

26 - 30 junio 2017 | Plasencia
Cáceres, Extremadura

7CFE01-406

Edita: Sociedad Española de Ciencias Forestales
Plasencia. Cáceres, Extremadura. 26-30 junio 2017
ISBN 978-84-941695-2-6

© Sociedad Española de Ciencias Forestales

Estudio de la ocurrencia y gravedad de los incendios forestales mediante modelos de regresión distribucional aditivos estructurados

RIOS -PENA, L.¹, CADARSO - SUAREZ, C.², KNEIV, T.³ Y MAREY-PÉREZ, M.F.⁴

¹ Departamento de Biología de la Conservación. Estación Biológica de Doñana-CSIC, C/Américo Vespucio 26, 41092 Sevilla

² Unidad de Bioestadística. Departamento de Estadística e Investigación Operativa. Facultad de Medicina, Universidad de Santiago de Compostela, C/San Francisco s/n, 15782 Santiago de Compostela

³ Chair of Statistics, Georg-August-University Göttingen, Humboldtalle 3, 37073 Göttingen, Germany.

⁴ GI-1716 PROEPLA, Proyectos y Planificación. Universidad de Santiago de Compostela. Escuela Politécnica Superior, campus universitario s/n 27002, Lugo.

Resumen

Cuando se estudia el fenómeno empírico de los incendios forestales, se puede distinguir entre dos aspectos importantes: la ocurrencia de incendios forestales en una ubicación específica en un punto de tiempo dado y la gravedad de los incendios forestales que se mide, por ejemplo, por el porcentaje de superficie quemada en ese lugar. Este estudio propone el uso de modelos de regresión aditivos estructurados basados en la distribución beta inflada cero-uno para estudiar simultáneamente la ocurrencia de incendios forestales y el área quemada. La distribución beta ofrece una manera conveniente de estudiar el porcentaje de superficie quemada cuando los porcentajes están delimitados entre cero y uno y además la inflación con ceros y unos nos permite tratar observaciones sin zona quemada o con el 100% área quemada como casos especiales. La metodología propuesta se aplicó a una gran base de datos de incendios forestales en Galicia. Los resultados mostraron escenarios de riesgo de incendio para las variables meteorológicas estudiadas, es decir, las condiciones climáticas espaciales y temporales influyen en la cantidad y características de los incendios forestales. Estos resultados son información esencial a la hora de elaborar planes de prevención y lucha contra incendios forestales.

Palabras clave

Incendios forestales, regresión aditiva estructurada, efectos espaciales y temporales, simulación de Monte Carlo, Galicia.

1. Introducción

Cada vez existe un mayor interés en desarrollar modelos de predicción de incendios que permitan determinar con más precisión dónde y cómo van a ser esos incendios. La mejora en la toma de datos y especialmente disponer de modelos matemáticos más flexibles y adaptados a la realidad son las claves para los futuros desarrollos de modelos de predicción (RÍOS-PENA et al. 2015; 2017).

Conocer la probabilidad de ocurrencia de incendios forestales es muy interesante desde la perspectiva de la gestión de riesgos, también es importante cuantificar el área quemada de cada incendio en cada lugar. Como representar la importancia de los incendios, se puede llevar a cabo mediante el porcentaje de área quemada (LOEPFE et al., 2014, HERNÁNDEZ et al., 2015, RUFFAULT et al., 2016). En nuestro trabajo proponemos estudiar simultáneamente la ocurrencia de incendios y el área afectada por los incendios forestales de forma conjunta.

2. Objetivos

El objetivo principal del presente trabajo es analizar y validar un modelo de regresión aditivo estructurado como predictor simultáneo de la ocurrencia y gravedad de la superficie afectada por los incendios forestales.

3. Metodología

3.1. Material

Se creó una base de datos propia a partir de los incendios forestales del Ministerio Agricultura y Pesca, Alimentación y Medio Ambiente (MAPAMA). Nuestra base de datos contiene información sobre el lugar (coordenadas geográficas) fecha y situación meteorológica (extraída de la estaciones climatológicas más próximas) para el momento del incendio. El período de estudio se centró en la primera quincena de agosto de 2006 en la que se produjeron un total de 2060 incendios forestales. La mayoría de los incendios se centran en las regiones costeras y el sureste de Galicia.

Se elaboró una base de datos espacial formada por las rejillas de 1×1 km para toda la superficie de Galicia y calculando la superficie quemada en cada cuadrícula. Alrededor del 99% de las celdas de la rejilla están completamente libres de incendios en el período de observación.

Se asociaron las siguientes variables a cada una de las 30.158 celdas de 1×1 km que se muestran en la tabla 1.

Tabla 1. Variables seleccionadas a partir de la base de datos para el análisis

| Variable | Descripción |
|---------------------|---|
| Número de incendios | Coordenadas geográficas del punto de ignición |
| burnt | Porcentaje de la celda quemada (%) |
| temp | Media diaria de la temperatura en el punto de ignición (°C) |
| hr | Humedad relativa media diaria en el punto de ignición (%) |
| dia | Día del mes |
| cdconc | Código municipal de la cuadrícula |

3.2. Metodología estadística

La distribución beta inflada cero-uno se define como la mezcla de un componente continuo después de una distribución beta y dos masas puntuales en cero y uno tal que finalmente se obtiene la densidad continua discreta mixta, de acuerdo a la siguiente ecuación:

$$p(y) = \left\{ \begin{array}{ll} \pi_0 & y = 0 \\ (1 - \pi_0 - \pi_1) \left(\frac{1}{B(a, b)} \right) y^{a-1} (1 - y)^{b-1} & 0 < y < 1 \\ \pi_1 & y = 1 \end{array} \right\}$$

Donde $B(a, b)$ indican la función beta.

$$B(a, b) = \int_0^1 y^{a-1} (1 - y)^{b-1} dy$$

Con parámetros $a > 0$ y $b > 0$ donde π_0 y $\pi_1 \in (0, 1)$ y donde las probabilidades de que centro de masas esté en cero y en uno, respectivamente, responde a la ecuación $\pi_0 + \pi_1 < 1$. OSPINA & FERRARI (2010) propusieron una parametrización alternativa de la distribución beta inflada cero-uno que permite una interpretación más fácil en el contexto de la regresión. En primer lugar, la parte continua correspondiente a una distribución beta está parametrizada en base a la expectativa.

$$\mu = \frac{a}{a+b} \in (0,1)$$

Y a la escala del parámetro.

$$\mu = \frac{a}{a+b} \in (0,1)$$

En una segunda etapa, las probabilidades de las masas puntuales en cero y uno son reemplazadas por $\nu > 0$ y $\tau > 0$ determinadas a partir de las ecuaciones.

$$\pi_0 = \frac{\nu}{1+\nu+\tau}, \pi_1 = \frac{\tau}{1+\nu+\tau}$$

En este caso ν y τ son proporcionales al porcentaje de superficie quemada en cada celda, respectivamente. Parametrizando de esta forma se evita la restricción ocasionada por $\pi_0 + \pi_1 < 1$.

A partir de la regresión distributiva (KLEIN et al., 2015) y de los modelos aditivos generalizados de localización, escala y forma (RIGBY & STASINOPOULOS, 2005), se pueden relacionar cada uno de los cuatro parámetros de la distribución beta inflada cero-uno a predictores de regresión, de la siguiente forma.

$$\mu_i = \frac{\exp(\eta_i^\mu)}{1 + \exp(\eta_i^\mu)}, \sigma_i = \frac{\exp(\eta_i^\sigma)}{1 + \exp(\eta_i^\sigma)}, \nu_i = \exp(\eta_i^\nu), \tau_i = \exp(\eta_i^\tau)$$

Una de las principales ventajas de la regresión aditiva estructurada (FAHRMEIR et al., 2004) consiste en la obtención de modelos generales lineales y aditivos generalizados con una estructura más flexible para los predictores de regresión, como se muestra en la siguiente ecuación.

$$\eta = \beta_0 + f_1(x) + \dots + f_j(x) \quad (1)$$

β_0 es la intercepción global, y las funciones f_1, \dots, f_j representan efectos de regresión de diferentes tipos de la covariable genérica vector x . Cada una de las funciones del predictor aditivo se aproxima ahora mediante una combinación lineal de funciones $B_1(x), \dots, B_K(x)$ tal que,

$$f(x) = \sum_{k=1}^K \beta_k B_k(x)$$

Donde β_k denota los correspondientes coeficientes de base. En la notación matricial, un predictor aditivo estructurado puede ser expresado como.

$$\eta = \beta_0 \mathbf{1} + \mathbf{Z}_1 \beta_1 + \dots + \mathbf{Z}_j \beta_j$$

Donde η se obtiene uniendo todas las observaciones individuales, $\mathbf{1}$ se refiere a un vector de unos, \mathbf{Z}_j surge de la evaluación de las funciones de base en las covariables observadas y β_j es el vector de coeficientes de base correspondientes.

Para asegurar las propiedades favorables de las estimaciones de función como la suavidad en caso de efectos no lineales y espaciales o contracción en caso de efectos aleatorios, nos basamos en

una formulación bayesiana donde las distribuciones anteriores se asignan a los vectores de coeficientes de regresión. Generalmente, éstas son dadas por distribuciones normales multivariadas.

$$p(\beta_j | \tau_j^2) \propto \exp\left(\frac{-1}{2\tau_j^2} \beta_j' K_j \beta_j\right)$$

Donde τ_j^2 es la varianza previa que determina el impacto de la distribución previa sobre las estimaciones de función mientras que K_j es la matriz de precisión previa que implementa las suposiciones de suavidad/dispersión. Desde una perspectiva frecuencial, las distribuciones anteriores corresponden a penalizaciones de la forma $\lambda_j \beta_j' K_j \beta_j$ donde $\lambda_j \geq 0$ es el parámetro de suavizado que determina el impacto de la penalización.

Para determinar los componentes del modelo se utilizan tres casos específicos de efectos de la clase de predictores aditivos estructurados que nos permiten tener en cuenta los efectos no lineales, espaciales y aleatorios. Para los efectos no lineales, nos basamos en splines penalizados (EILERS & MARX, 1996, BREZGER & LANG, 2006), donde se utilizan funciones de base B-spline para aproximar el efecto de una covariable continua. Como valor predeterminado, se consideraron B-splines cúbicos basados en 20 nudos equidistantes. Como distribución previa, se utilizó una distribución aleatoria de segundo orden que, desde una perspectiva frecuencial, se corresponde con penalizar las diferencias de segundo orden de los parámetros para las funciones de base adyacentes.

Para los efectos espaciales se utilizó una distribución aleatoria de Markov, asignando un coeficiente de regresión separado a cada región de un conjunto discreto de ubicaciones espaciales. Esto implica que las funciones de base en este caso corresponden a simples funciones de indicador para las diferentes regiones. Para reforzar la suavidad espacial, asumimos un campo aleatorio de Markov gaussiano anterior para los coeficientes de base. Para los efectos aleatorios, asumimos que los coeficientes de regresión son normales, independientes e idénticamente distribuidos. Para el desarrollo de la metodología estadística se utilizó el programa BayesX 3.0.2.

4. Resultados

Se asume la distribución beta inflada cero-uno con 4 parámetros de distribución $E(y_i | y_i > 0) = \mu_i \in (0, 1), \sigma_i^2 \in (0, 1)$ proporcional a $Var(y_i | y_i > 0), \nu_i$ y τ_i donde $P(y_i = 0) = \frac{\tau_i}{1 + \nu_i + \tau_i}$ y $P(y_i = 1) = \frac{\nu_i}{1 + \nu_i + \tau_i} > 0$.

Se estimó el siguiente modelo,

$$\eta_i = \beta_0 + f_1(dia_i) + temp_i \beta_2 + f_3(hr_i) + f_{spat}(cdconc_i)$$

Donde f_1 ajusta la relación entre el día en que ocurre un incendio forestal y el período de tiempo estudiado, β_2 ajusta la temperatura diaria media, f_3 ajusta la humedad relativa diaria, y f_{spat} captura la heterogeneidad espacial. β_0 representa el nivel global del predictor, f_1 y f_3 son funciones no lineales y se modelan como splines cúbicos penalizados con 20 nudos internos y orden aleatorio de segundo orden, el efecto espacial f_{spat} se basa en 315 municipios gallegos y se descompone en una parte lisa modelada por un campo aleatorio de Markov antes de capturar efectos que varían suavemente así como un efecto aleatorio espacial no estructurado que representa un patrón espacial localizado. A este último se le asigna un i.i.d. Gaussiano anterior al igual que los efectos individuales aleatorios específicos.

Para todos los parámetros de los cuatro candidatos de distribución, hemos comparado diferentes predictores basadas en la DIC. Encontramos que varios efectos tienen una influencia lineal sobre σ^2 ν y τ establecemos las siguientes estructuras predictoras para los cuatro parámetros de distribución:

$$\eta_i^\mu = \beta_0^\mu + f_1^\mu(hr_i) + temp_i\beta_2^\mu + f_3^\mu(dia_i) + f_{spat}^\mu(cdconc_i)$$

$$\eta_i^{\sigma^2} = \beta_0^{\sigma^2} + f_1^{\sigma^2}(hr_i) + temp_i\beta_2^{\sigma^2} + f_3^{\sigma^2}(dia_i)$$

$$\eta_i^\nu = \beta_0^\nu + f_1^\nu(hr_i) + f_2^\nu(temp_i) + f_3^\nu(dia_i) + f_{spat}^\nu(cdconc_i)$$

$$\eta_i^\tau = \beta_0^\tau + f_1^\tau(hr_i) + temp_i\beta_2^\tau + f_{spat}^\tau(cdconc_i)$$

La figura 1 muestra las estimaciones medias posteriores de los efectos no lineales sobre μ , σ^2 , ν y τ junto con intervalos de credibilidad al 80% y 95% centrados en torno a cero.

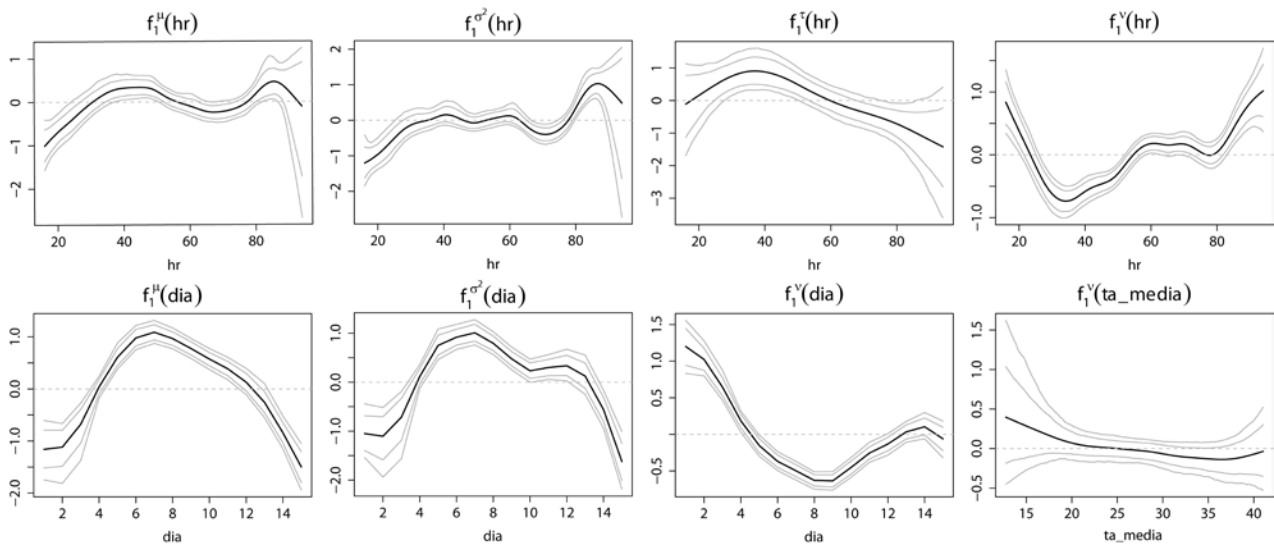


Figura1. Efectos de las covariables no lineales con los parámetros de la distribución.

Las tablas 2, 3 y 4 muestran los resultados obtenidos para los efectos lineales sobre μ (Tabla 1), sobre σ^2 (Tabla 2) y sobre ν (Tabla 3).

Table 2. Resumen de la distribución a posteriori de los efectos lineales sobre la μ en el modelo beta inflado cero-uno.

| Varname | pmean | pstd | Pqu2p5 | Pqu10 | pmed | Pqu90 | Pqu97p5 | Pcat95 | Pcat80 |
|---------|-------|------|--------|-------|-------|-------|---------|--------|--------|
| const | -3.17 | 0.48 | -4.13 | -3.78 | -3.18 | -2.53 | -2.24 | -1 | -1 |
| pp | 0.65 | 0.14 | 0.37 | 0.47 | 0.65 | 0.82 | 0.91 | 1 | 1 |
| temp | 0.01 | 0.02 | -0.02 | -0.01 | 0.01 | 0.04 | 0.05 | 0 | 0 |

Table 3. Resumen de la distribución a posteriori de los efectos lineales sobre la σ^2 en el modelo beta inflado cero-uno.

| Varname | pmean | pstd | Pqu2p5 | Pqu10 | pmed | Pqu90 | Pqu97p5 | Pcat95 | Pcat80 |
|---------|-------|------|--------|-------|-------|-------|---------|--------|--------|
| const | -2.83 | 0.49 | -3.80 | -3.46 | -2.83 | -2.18 | -1.87 | -1 | -1 |
| pp | 1.00 | 0.15 | 0.69 | 0.80 | 1.00 | 1.19 | 1.30 | 1 | 1 |
| temp | 0.04 | 0.02 | 0.00 | 0.01 | 0.04 | 0.06 | 0.07 | 1 | 1 |

Table 4. Resumen de la distribución a posteriori de los efectos lineales sobre la \mathcal{U} en el modelo beta inflado cero-uno.

| Vaname | pmean | pstd | Pqu2p5 | Pqu10 | pmed | Pqu90 | Pqu97p5 | Pcat95 | Pcat80 |
|--------|-------|------|--------|-------|------|-------|---------|--------|--------|
| const | 5.86 | 0.10 | 5.68 | 5.74 | 5.85 | 5.98 | 6.05 | 1 | 1 |
| pp | 0.25 | 0.07 | 0.12 | 0.16 | 0.25 | 0.35 | 0.40 | 1 | 1 |

La Figura 2 muestra los efectos espaciales medios posteriores del análisis sobre la aparición de los incendios forestales para cada uno de los parámetros de distribución μ , ν y τ

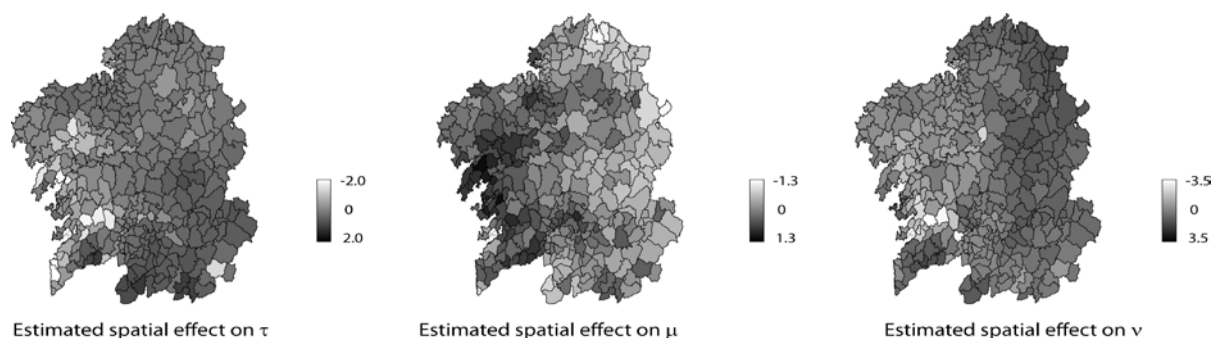


Figura2. Distribución de los efectos espaciales en los municipios de Galicia.

Los efectos espaciales estructurados muestran una tendencia al aumento de la incidencia de incendios en el Oeste y Sur de Galicia contra la presencia de incendios en la parte inferior de la superficie, en el Norte y Este de Galicia. A intervalos de confianza del 95% muestran municipios con alta probabilidad de ocurrencia de incendios (municipios en negro) y baja o nula probabilidad de incendios (municipios en blanco). Por lo tanto, el mapa sugiere que hay municipios que tienen incidencia significativamente menor de incendios y aparecen principalmente en el Norte y Noreste de Galicia antes de una gran ocurrencia de incendios en casi toda la costa oeste de Galicia. Todo ello de acuerdo con los datos del período analizado.

5. Conclusiones

Este trabajo constituye un nuevo enfoque para estudiar simultáneamente la ocurrencia y el área quemada de incendios forestales en el marco de la regresión distributiva aditiva estructurada validada con los datos de Galicia de agosto de 2006. El modelo propuesto considera la información espacial en forma de variable de localización que pueden acelerar el desarrollo de modelos de comportamiento contra incendios y resultar útiles para elaborar planes de prevención de incendios y extinción de incendios.

6. Bibliografía

BREZGER, A.; LANG, S.; 2006. Generalized structured additive regression based on Bayesian P-splines. *Comput. Stat. Data Anal.* 50, 967–991.

EILERS, P.H.C.; MARX, B. D.; 1996. Flexible smoothing with B-splines and penalties. *Stat. Sci.* 11, 89–121.

FAHRMEIR, L.; KNEIB, T.; LANG, S.; 2004. Penalized structured additive regression for space-time data: A Bayesian perspective. *Stat. Sin.* 14, 731–761.

HERNANDEZ, CH.; DROBINSKI, P.; TURQUETY, S.; DUPUY, J.L.; 2015. Size of wildfires in the Euro-Mediterranean region: observations and theoretical analysis. *Nat. Hazard Earth Sys.* 15, 1331-1341.

KLEIN, N.; KNEIB, T.; LANG, S.; SOHN, A.; 2015. Bayesian Structured Additive Distributional Regression with an Application to Regional Income Inequality in Germany. *Ann. Appl. Stat.* 9, 1024-1052.

LOEPFE, L.; RODRIGO, A.; LLORET, F.; 2014. Two thresholds determine climatic control of forest fire size in Europe and northern Africa. *Reg. Environ. Chang.* 14, 1395–1404.

OSPINA, R.; FERRARI, S.L.P.; 2010. A general class of zero-or-one inflated beta regression models. *Comput. Stat. Data Anal.* 56, 1609–1623.

RIGBY, R.A.; STASINOPOULOS, D.M.; 2005. Generalized additive models for location, scale and shape. *J. Appl. Stat.* 54, 507–554.

RÍOS-PENA, L.; CADARSO-SUÁREZ, C.; KNEIB, T.; MAREY-PÉREZ, M.F.; 2015. Applying Binary Structured Additive Regression (STAR) for predicting wildfire in Galicia, Spain. *Procedia Environ. Sci.* 27, 123 – 126.

RIOS-PENA, L.; KNEIB, T.; CADARSO-SUÁREZ, C.; MAREY-PÉREZ, M.F.; 2017. Predicting the occurrence of wildfires with binary structured additive regression models. *J. Environ. Manage.* 187, 154-165.

RUFFAULT, J.; MORON, V.; TRIGO, R.M.; CURT, T.; 2016. Daily synoptic conditions associated with large fire occurrence in Mediterranean France: evidence for a wind-driven fire regime. *Int. J. Climatol.* doi:10.1002/joc.4680.