

Aplicación de un Modelo de Regresión Aditiva Estructurada (STAR) con respuesta binaria para la predicción de incendios forestales en Galicia. Agosto 2006.

RÍOS-PENA LAURA¹, CADARSO-SUÁREZ CARMEN MARÍA^{1,2}, KNEIB THOMAS³ y
MAREY-PÉREZ MANUEL FCO⁴.

¹Biostatech, Advice, Training & Innovation in Biostatistics, S.L. E-Mail:laura.rios.pena@gmail.com

²Unidad de Bioestadística, departamento de Estadística e IO. Facultad de Medicina. Universidade de Santiago de Compostela C/San Francisco s/n 15782-Santiago de Compostela. E-Mail:carmen.cadarso@usc.es

³Department of Statistics, Institute for Statistics and Econometrics, Georg-August-University Göttingen. Platz der Göttinger Sieben 5, 37073 Göttingen. E-Mail:thomas.kneib@wiwi.uni-goettingen.de

⁴Grupo de investigación PROEPLA, departamento de Ingeniería Agroforestal, Escola Politécnica Superior de Lugo. Universidade de Santiago de Compostela. C/ Benigno Ledo, Campus Universitario 27002 Lugo. E-Mail:manuel.marey@usc.es

RESUMEN

Los incendios forestales suponen uno de los principales problemas medioambientales a los que se enfrenta la sociedad actual. En particular, en Galicia el fuego es la principal causa de destrucción de los bosques. En este trabajo se ha utilizado la metodología estadística de Regresión Aditiva Estructurada (STAR) para la modelización de los incendios forestales. Esta metodología constituye una nueva aportación sobre los clásicos modelos de regresión logística y los más actuales GLM y GAM. Su ventaja se encuentra en la flexibilidad del modelado a la hora de incorporar variables espaciales y temporales directamente, mostrando un agregado en forma de mapa donde se explican las dependencias tanto para el conjunto de unidades (efectos estructurados) como para cada unidad (efectos no estructurados). La posibilidad de trabajar a escalas espaciales con nivel de resolución de vóxeles de $1Km \times 1Km$ por día unido a la posibilidad de mapear las predicciones en un rango de color, hace útil el modelo a efectos de planificación y gestión de incendios. Todo ello facilitará el desarrollo de modelos de comportamiento del fuego, que pueden ser de gran valor en la elaboración de planes de prevención y lucha contra el fuego.

Palabras y frases llave: Incendios forestales, regresión aditiva estructurada, modelo STAR, vóxel.

1. INTRODUCCIÓN

Un incendio forestal es cualquier fuego incontrolado en la vegetación combustible localizado en el campo o en una zona forestal. A nivel de paisaje, la ignición y propagación de los incendios forestales son el resultado de una compleja interacción entre el clima, la topografía y la cubierta vegetal (Mermoz et al, 2005.; Moreira et al., 2011). Las cuatro principales causas naturales de los incendios forestales son los relámpagos, las erupciones volcánicas, las chispas de la caída de rocas y la combustión espontánea (National Interagency Fire Centre, 2011). Sin embargo, muchos incendios se atribuyen a fuentes humanas (Pyne et al., 1996).

En el sur de Europa (Francia, Grecia, Italia, Portugal y España), la lucha contra el fuego es uno de los problemas más graves a los que se enfrentan los gestores ambientales (Badia-Perpiñán y Pallares-Barbera, 2006). En los últimos treinta años, un promedio de 49.838 incendios forestales han quemado aproximadamente 0,5 millones de hectáreas de bosques y otras tierras rurales cada año (San Miguel-Ayanz et al., 2012).

Últimamente, ha habido un aumento en el número de incendios producidos de forma intencionada (Catry et al., 2009; Chuvieco et al., 2010; González-Olabarria et al., 2011; Juan et al., 2012;

Romero-Calcerrada et al., 2008). Diferentes causas de incendio se han considerado, como el aumento de la superficie forestal (Badia et al., 2003; Catry et al., 2009); nuevos patrones del paisaje agrario (Amatulli et al., 2007; Catry et al., 2009; Corcoran et al., 2007; Fernandes, 2009; Kalabokidis et al., 2007; Martínez et al., 2009; Moreira et al., 2009 y 2011; Verde y Zêzere, 2010); los patrones del paisaje agrario tradicional (Vázquez y Moreno, 1998 y 2001); el abandono agrícola (Acosta et al., 2005; Aranzabal et al., 2008; Falcucci et al., 2007); la expansión de superficie de matorral (Bajocco y Ricotta, 2008; Díaz-Delgado et al., 2004, González et al., 2006; González y Pukkala, 2007; Koutsias et al., 2009; Moreira et al., 2009; Mouillot et al. 2005; Nunes et al., 2005, Sebastian-López et al., 2008); cambios socioeconómicos (Arianoutsou, 2001; Badia-Perpiñán y Pallares-Barbera, 2006; Cardille et al., 2001; Curt y Delcros, 2010; Loboda y Czizar 2007; Martínez et al., 2009; Romero-Calcerrada et al., 2010; Vasconcelos et al., 2001; Vilar et al., 2010; Wittenberg y Malkinson, 2009; Zea-Bermúdez et al., 2009) y los cambios en la distribución de la renta junto con un mayor nivel de desempleo (González-Olabarria et al., 2011).

Diferentes enfoques han sido considerados para resolver las dificultades del modelado de ignición de origen humano en Europa con un enfoque espacial explícito. Algunos estudios han intentado predecir la probabilidad de ocurrencia de incendios mediante modelos logísticos (Cardille et al., 2001; Catry et al., 2009; Chuvieco et al., 2010; Martínez et al., 2008; Martínez et al., 2009; Vasconcelos et al., 2001; Wittenberg y Malkinson., 2009), las redes neuronales artificiales (Vasconcelos et al., 2001; Vilar del Hoyo et al., 2007); pesos de evidencia (Amatulli et al., 2007; Romero-Calcerrada et al., 2008 y 2010); estadística multivariante (Kalabokidis et al., 2007); regresión lineal generalizada (Wood y Augustin, 2002); la distribución generalizada de Pareto (Zea-Bermúdez et al., 2009); la lógica difusa (Loboda y Czizar, 2007); modelos bayesianos (Verde y Zêzere, 2010); los modelos de Poisson (Benavet-Corai et al., 2007; Mandallaz y Ye, 1997; Wotton et al., 2003); los modelos aditivos generalizados (Vilar et al., 2010); densidad Kernel (Amatulli et al., 2007; de la Riva et al., 2004) y simulaciones de Monte Carlo (Conedera et al., 2011).

El presente trabajo persigue la obtención de un modelo predictivo de ocurrencia de incendios forestales a partir de variables climatológicas y de tipo de combustible. Se propone el empleo de técnicas de Regresión Aditivas Estructuradas (STAR) (Fahrmeir, L., Kneib, T., 2011) que constituyen una aportación metodológica sobre los clásicos modelos de regresión logística y los más actuales GLM y GAM en la flexibilidad del modelado a la hora de incorporar variables espaciales y temporales directamente, mostrando un agregado en forma de mapa donde se explican las dependencias tanto para el conjunto de unidades (efectos estructurados) como para cada unidad (efectos no estructurados).

Para generar este modelo predictivo a una resolución de $1\text{ Km} \times 1\text{ Km}$ asociando las cuadrículas de la rejilla del área de estudio al concello al que pertenecen, se toma como variable respuesta información para los incendios forestales. Para alcanzar este objetivo general se abordarán los siguientes objetivos específicos:

- Identificar aquellas variables independientes que representen los diversos factores de riesgo de incendio forestal (en este estudio se han considerado las variables climatológicas y de tipo de combustible existente en el suelo).
- Generar la variable dependiente en el modelo: En este trabajo se generan un tipo de variable respuesta para un mismo modelo de variables explicativas; la ocurrencia de incendios forestales por concello.
- Generar un modelo de regresión aditiva estructurada (STAR) de respuesta binaria para la variable respuesta.
- Validar los resultados obtenidos en el modelo de regresión aditiva estructurada (STAR) con respuesta binaria.

El propósito final de la línea de trabajo es desarrollar un modelo consistente espacio-temporal que puedan integrarse con facilidad en un sistema de predicción de riesgo más complejo que incluya otros factores (sociales, económicos, etc.) relacionados con la ocurrencia de incendios forestales.

2. ÁREA DE ESTUDIO Y BASE DE DATOS

Área de Estudio. La modelización de la ocurrencia de incendios debido a condiciones climáticas y de tipo de combustible para los incendios forestales ha sido elaborado para la Comunidad Autónoma de Galicia (ver Figura 1). El período de estudio comprende la primera quincena del mes de Agosto de 2006. Se ha elegido este período para asegurar la consistencia de los datos de incendios utilizados y para garantizar la robustez de los análisis estadísticos efectuados.

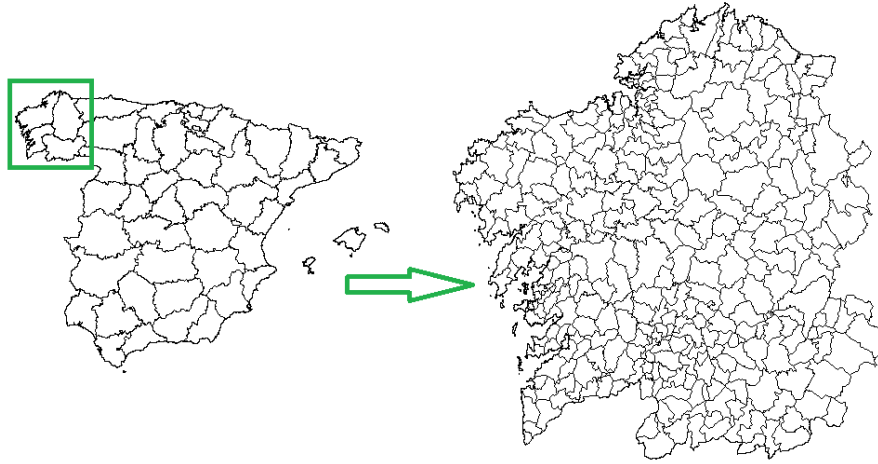


Figura 1: Mapa de España - Área de estudio por concellos

Base de datos. Los datos de incendios utilizados son una selección de la base de datos de incendios de la Consellería de Medio Rural y Mar de la Xunta de Galicia, correspondientes a la primera quincena de agosto de 2006, en la que se dieron unas circunstancias de gran número de incendios simultáneos.

Esta nueva base de datos está formada en un total de 362625 registros. La componente espacial, está formada por 30685 celdas que proceden de dividir el territorio gallego con una rejilla de $1Km \times 1Km$ asociando a cada una de las cuadrículas un total de 9 variables. Posteriormente, para los análisis y la presentación de los resultados, se ha utilizado un mapa “boundary” de los concellos de Galicia, la selección de esta escala de trabajo permite facilitar la interpretación de los resultados y un menor coste computacional. La componente temporal se corresponde a los 15 primeros días del mes de agosto de 2006 en los que se produjeron un total de 2060 incendios.

La variable respuesta es binaria: (1) ocurrencia de incendio forestal (0) no ocurrencia de incendio forestal. Las variables de estudio asociadas a cada cuadrícula son:

- **Variables meteorológicas:** temperatura media, humedad relativa y precipitaciones diarias, de las que se ha registrado dato en la estación meteorológica más cercana al incendio. Se ha establecido una restricción de distancia máxima de 30 Km.
- **Variables de tipos de modelos de combustible (Rothermel (1983)),** que se corresponden con la superficie en m^2 de cada uno de los 5 grupos de modelos (sin asignar modelo, pastos, matorral, hojarasca bajo arbolado y restos de cortas/operaciones selvícolas).

El cuadro 1 aporta una descripción de la variable respuesta (la primera), explicativas y de agregación espacial (última) que usaremos en nuestro modelo.

Variables	Descripción de las variables de estudio
numero_incendios	Ocurrencia de incendio forestal con categorías “no ocurrencia”(=0) y “ocurrencia” (=1)
ta_media	Temperatura media diaria registrada para cada cuadrícula en °C
hr	% de la humedad relativa media diaria para cada cuadrícula
pp	Precipitación diaria registrada en cada cuadrícula en l/m ²
mc_Sin_asignar_modelo	Formación heterogénea, no se asigna ningún modelo
mc_pastos	Formaciones de pastizales por cuadrícula en m ²
mc_matorral	Formaciones de matorral por cuadrícula en m ²
mc_hojarasca	Formaciones de hojarasca bajo bosque por cuadrícula en m ²
mc_restos	Formaciones de bosque con gran cantidad de restos leñosos muertos por cuadrícula en m ²
cdconc	Concellos de Galicia

Tabla 1: Variables generadas a partir del material de partida.

3. METODOLOGÍA ESTADÍSTICA

3.1 Modelo de Regresión Aditiva Estructurada (STAR)

El objetivo del análisis de regresión es medir la influencia de algunas variables x_j , $j = 1, \dots, q$, llamadas covariables, en una variable y llamada respuesta o variable independiente. El modelo más ampliamente usado es el modelo lineal clásico. Este modelo requiere una distribución gaussiana en la variable respuesta (o bajo supuestos menos estrictos que una continua). La relación entre la media condicional de la respuesta y las covariables es que asume:

$$E(y|x_1, \dots, x_q) = \gamma_0 + \gamma_1 x_1 + \dots + \gamma_q x_q = \gamma' x =: \eta \quad (1)$$

A través de su valor y signo algebraico, los coeficientes de regresión $\gamma_1, \dots, \gamma_q$ determinan la dirección y la fuerza de la influencia de sus covariables. El parámetro γ_0 es el término constante o intercepto. El parámetro η es referido como predictor lineal porque en la fórmula (1) es lineal en los coeficientes de regresión y la relación entre covariables y los valores esperados también son lineales.

Si la variable respuesta no es una distribución gaussiana pero pertenece a una familia exponencial univariante, el modelo lineal generalizado puede ser usado. Aquí, asumimos que el predictor lineal η y el valor esperado condicional están vinculados a través de una función de respuesta h , es decir:

$$E(y|x_1, \dots, x_q) = h(\gamma_0 + \gamma_1 x_1 + \dots + \gamma_q x_q) = h(\gamma' x) = h(\eta) \quad (2)$$

Generalmente, la función h se elige de tal manera que los valores de η son transformados al dominio del valor esperado. Para las respuestas de distribuciones gaussianas, los valores esperados puede adoptar todos los valores reales. Por lo tanto, una transformación no es necesaria y la función de identidad puede ser elegido por h , por ejemplo $h = \text{id}$. En este trabajo se reemplaza el predictor lineal

$$\eta := \gamma_0 + \gamma_1 x_1 + \dots + \gamma_q x_q = \gamma' x \quad (3)$$

por un predictor aditivo estructurado semiparamétrico (comparar con Fahrmeir, Kneib & Lang (2004)) de la forma

$$\eta := \gamma_0 + f_1(x_1) + \dots + f_q(x_q) + \gamma_1 u_1 + \dots + \gamma_p u_p = f_1(x_1) + \dots + f_q(x_q) + \gamma' u \quad (4)$$

La razón de usar un predictor semiparamétrico reside en los fuertes supuestos realizados por el predictor lineal. El predictor lineal supone:

1. Una influencia lineal de las covariables en el predictor o incluso en la respuesta en el caso gaussiano.

2. La independencia de las observaciones.

En muchas situaciones, sin embargo, las hipótesis no son adecuadas y son confrontadas con uno o más de los siguientes problemas:

- El efecto de algunas covariables continuas puede ser de una forma no lineal (desconocida).
- Las observaciones pueden ser espacialmente correlacionadas.
- Las observaciones pueden ser temporalmente correlacionadas.
- No puede haber heterogeneidad no observada entre individuos o unidades que no se explica por las covariables de las que se dispone.
- Puede haber interacción compleja entre dos variables continuas enditemize

El predictor aditivo estructurado (4) permite sustituir los efectos lineales $\gamma_j x_j$ por funciones $f_j(x_j)$. Las funciones f_j pueden ser de diferente tipo según los diferentes tipos de covariables x_j . Por ejemplo, el factor de predicción es capaz de modelizar los efectos no lineales de variables continuas o escalas de tiempo y puede manejar la información espacial o unidad específica. El predictor puede ser semiparamétrico, es decir, incluir una parte paramétrica $\gamma' u$ como en la ecuación ??, por lo que algunas covariables, en especial las variables categóricas, todavía puede ser modeladas por efectos lineales. Hay que tener en cuenta que las covariables se modelan linealmente por u_j con el fin de distinguirlas de otras covariables. La parte paramétrica $\gamma' u$ contiene también el término intercepto γ_0 .

Los modelos de regresión aditiva estructurada cubren una amplia gama de modelos diferentes. Algunos casos especiales que son muy conocidos en la literatura son los aditivos y modelos aditivos generalizados (Hastie y Tibshirani (1990), Rigby y Stasinopoulos (2005) o Wood (2006a)), modelos mixtos aditivos generalizados (Ruppert, Wand y Carroll (2003)), modelos geoaditivos (Fahrmeir y Lang (2001a) o Kammann y Wand (2003)), time-varying coefficient models (Hastie y Tibshirani (1993)), regresión geográficamente ponderada (Fotheringham, Brunson y Charlton (2002)) y modelos de iteración de tipo ANOVA (Chen (1993)).

3.2 Las componentes del modelo

Para cada tipo de covariable, existen una o más posibilidades de construir una función que represente adecuadamente la disposición de la información. Estas posibilidades con sus características específicas se definen en esta sección. Todas las funciones no lineales descritas en esta sección pueden ser escritas de forma general. Esto permite una igualdad de trato de todas las funciones no lineales para estimar los coeficientes de regresión y seleccionar las covariables relevantes. Esto significa que, para la inferencia y algoritmos de selección solo es necesario distinguir entre dos casos: efectos lineales y funciones no lineales.

Las características comunes de todas las funciones no lineales $f(x)$ son:

- El vector de evaluaciones de la función $\mathbf{f} = (f_1, \dots, f_n)'$ para n observaciones puede ser escrito como una combinación lineal de una $n \times p$ matriz de diseño $\mathbf{X}$ y un vector de coeficientes de regresión $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, es decir,

$$\mathbf{f} = \mathbf{X}\boldsymbol{\beta} \quad (5)$$

Esto significa que todas las funciones f son lineales en sus coeficientes de regresión.

- En un marco bayesiano, cada función f está provista de una distribución a priori. La distribución previa depende del tipo de covariable x y en suposiciones acerca de la suavidad de la función f . Esto conduce a diferentes tipos de funciones que se describirán posteriormente en detalle. Generalmente, los supuestos anteriores a cerca de f se pueden expresar mediante la aplicación de una distribución a priori de los coeficientes de regresión. La distribución es una distribución gaussiana adecuada o inadecuada de la forma

$$p(\boldsymbol{\beta}) \propto \exp\left(\frac{-1}{2\tau^2} \boldsymbol{\beta}' \mathbf{P} \boldsymbol{\beta}\right) \quad (6)$$

con un parámetro de varianza τ^2 y una matriz de penalización $\mathbf{P}$. Las distribuciones a priori de diferentes tipos de funciones se caracterizan por su matriz de penalización individual que contiene información sobre el tipo de función y suposiciones a cerca de la suavidad de la función. Si la matriz $\mathbf{P}$ es de rango deficiente, la distribución a priori es inadecuada, de lo contrario es apropiada. Existe una estrecha relación entre lo Bayesiano y el enfoque de verosimilitud penalizada. Suponemos que el predictor solo contiene la función f , es decir, $\eta = \mathbf{f} = \mathbf{X}\boldsymbol{\beta}$. En este caso la función de verosimilitud $L(y|\boldsymbol{\beta})$ y la función de verosimilitud logarítmica $l(y|\boldsymbol{\beta})$ solo contienen el parámetro vector y ningún otro parámetro. Entonces, la distribución a posteriori $p(\boldsymbol{\beta}|y)$ con el vector respuesta $y = (y_1, \dots, y_n)$ viene dado por

$$p(\boldsymbol{\beta}|y) \propto L(y|\boldsymbol{\beta}) \cdot p(\boldsymbol{\beta}) \quad (7)$$

El modo de esta distribución puede ser calculada a partir de la distribución logarítmica a posteriori

$$\log(p(\boldsymbol{\beta}|y)) \propto l(y|\boldsymbol{\beta}) + \log(p(\boldsymbol{\beta})) \propto l(y|\boldsymbol{\beta}) - \frac{1}{2\tau^2} \boldsymbol{\beta}' \mathbf{P} \boldsymbol{\beta} \quad (8)$$

La fórmula (8) es equivalente al logaritmo de verosimilitud penalizada donde la matriz de penalización $\mathbf{P}$ es la matriz utilizada, por lo tanto, la estimación de máxima verosimilitud y la forma de la distribución a posteriori son idénticas.

En el contexto de verosimilitud penalizada, en lugar del parámetro de varianza τ^2 generalmente se utiliza un parámetro de suavizado para controlar la suavidad de la función. Este parámetro de suavizado es definido como $\lambda := \phi/\tau^2$ (ver Green y Silverman (1994)), donde ϕ es el parámetro de escala de la distribución de la variable respuesta, es decir, $\phi = \tau^2$ y $\lambda := \frac{\sigma^2}{\tau^2}$ para el caso especial de respuesta con distribución gaussiana. La fórmula del logaritmo de verosimilitud penalizado, la cuál ha de ser maximizada para el cálculo de las estimaciones de $\boldsymbol{\beta}$, se define como

$$l_{pen}(y|\boldsymbol{\beta}) = \phi \cdot l(y|\boldsymbol{\beta}) - \frac{1}{2} \lambda \boldsymbol{\beta}' \mathbf{P} \boldsymbol{\beta} \quad (9)$$

- Todas las distribuciones a priori (??) incluyen un parámetro de varianza τ^2 que influye en la forma del efecto estimado. En este apartado consideramos el parámetro τ^2 o equivalente al parámetro de suavización λ para cada función, tanto no lineal como fija.

Efectos Lineales

Un predictor aditivo estructurado contiene a menudo una parte paramétrica que incluye variables $u_j, j = 1, \dots, q$ que han de ser modeladas linealmente. Además, el predictor siempre contiene un término intercepto γ_0 . Para el vector de parámetros de regresión $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_q)'$ para todos los efectos lineales incluidos en el término intercepto γ_0 , no utilizamos ninguna penalización. En este caso, obtenemos las estimaciones de máxima verosimilitud de los coeficientes.

Variables Categóricas

Con el fin de estimar el efecto de una variable categórica u con $k \geq 2$ categorías, la variable está representada por $k - 1$ variables dummy. Una de las categorías tiene que ser especificada como categoría de referencia. Sin restricción, el número de categorías $1, \dots, k$ utiliza la última categoría k como referencia.

- Codificación dummy. Las variables dummy $u_j, j = 1, \dots, k - 1$ se definen como

$$u_j = \begin{cases} 1, & \text{si } u = j \\ 0, & \text{en otro caso} \end{cases} \quad (10)$$

La categoría de referencia se indica mediante entradas de cero en todas las variables dummy. El efecto de la variable categórica es una combinación lineal de todas las variables dummy, es decir

$$\gamma_1 \cdot u_1 + \dots + \gamma_{k-1} \cdot u_{k-1}$$

y se añade a la parte paramétrica del predictor.

- Codificación efecto, que trabaja de forma similar a la anterior pero las variables u_j son ahora definidas por

$$u_j = \begin{cases} 1, & \text{si } u = j \\ -1, & \text{si } u = k \\ 0, & \text{en otro caso} \end{cases} \quad (11)$$

para $j = 1, \dots, k-1$. Esto conduce a una interpretación diferente de los coeficientes de regresión. Un parámetro para la categoría de referencia puede calcularse por

$$\gamma_k = - \sum_{j=1}^{k-1} \gamma_j$$

El intercepto representa el promedio de todas las categorías y parámetros γ_j y la diferencia entre esta media y la categoría j .

Covariables Continuas

En esta sección, consideramos que el modelo simple $\eta_i = f(x_i)$, $i = 1, \dots, n$, donde la función f se supone que es una función suave de una variable continua o escala de tiempo x . Para aproximar estas funciones no lineales, hay diferentes enfoques en la literatura, ya sea en función de los enfoques de probabilidad local (véase, por ejemplo Fan y Gijbels (1984) y Loader (1999)) o en una expansión de funciones de la base. En este trabajo se considerará el último caso.

1. B-Splines

Las funciones de base que se usan son funciones spline polinomiales que se definen a trozos sobre un conjunto de nodos. Los nodos se separan del rango de la variable x como

$$x_{min} = k_0 < \dots < k_r = x_{max}$$

Cada función de base, respectivamente, cada spline, es

- Un polinomio de grado l en el intervalo $[k_i, k_{i+1}]$, $i = 0, \dots, r-1$
- $l-1$ veces continuamente diferenciables en los nodos k_i (l veces en todos los otros puntos además de los nodos).

La función f puede ser escrita como una combinación lineal de las funciones de la base B_j , es decir

$$f(x_i) = \beta_1 \cdot B_1(x_i) + \dots + \beta_p \cdot B_p(x_i)$$

donde $p = l + r$ (ver De Boor (1978) o Dierckx (1995)). Los términos $B_j(x_i)$ denotan el valor de la j -th función de base evaluada en el punto de observación x_i y sirven como nuevas covariables. La propia función f también puede ser llamada spline, ya que tiene las mismas propiedades que se ha descrito anteriormente. En la notación matricial, cada fila i de la matriz de diseño $\mathbf{X} = (B_j(x_i))$ contiene las evaluaciones de las funciones de todas las funciones de base para el respectivo punto de observación x_i . El vector de la función de evaluación $\mathbf{f}$ está dado por $\mathbf{f} = \mathbf{X}\boldsymbol{\beta}$.

2. P-Splines

Una cuestión crucial con B-splines es la elección de los nodos, afectando tanto en el número de nodos como su ubicación. La pregunta es cuántos nodos deberían ser elegidos de modo que la función resultante no sea demasiado duro ni para suavizar. Este problema a menudo se llama equilibrio del sesgo y la varianza (ver Hastie y Tibshirani (1990)): muchos nodos resultado de una función en bruto que está cerca de los datos y por lo tanto tiene un pequeño sesgo. Sin embargo, la varianza de esta función es grande. Pocos nodos resultan en una función suave que tiene sólo una pequeña varianza pero un alto sesgo en su lugar. Un problema adicional cuando se eligen sólo unos pocos nodos es dónde colocar los nudos.

Para superar estos problemas, hay dos enfoques diferentes en la literatura: el primero se basa en la selección adaptativa del nodo donde los nodos son elegidos cuidadosamente, pero en las posiciones que dan lugar a una función suficientemente flexible. Enfoques bayesianos para la selección adaptativa del nodo se describen en Biller (2000). El segundo método utiliza una sanción dura. La idea es utilizar un número relativamente grande de funciones de la base para obtener la suficiente flexibilidad. La suavidad se consigue por un término de penalización que impone restricciones sobre el vector de parámetros β , como por ejemplo la reducción de los parámetros a cero o penalizar saltos abruptos entre los parámetros adyacentes. Para ese propósito, el logaritmo de verosimilitud se sustituye por un logaritmo de verosimilitud penalizado definido por:

$$l_{pen}(y|\beta_1, \dots, \beta_p) = \phi \cdot l(y|\beta_1, \dots, \beta_p) - \frac{1}{2} \cdot \text{penalty}(\lambda), \quad (12)$$

donde el equilibrio entre el sesgo y la varianza, es decir, entre la flexibilidad y la suavidad, es controlado por el parámetro de suavización λ .

Utilizamos la llamada P(enalised)-splines la cual fue introducida por Eilers y Marx (1996) y Marx y Eilers (1998) y que se basan en la base B-splines. Aquí 20-40 nodos son elegidos, por lo general equidistantes en el rango de x . Con el fin de asegurar la suavidad se utiliza un término de penalización diferente que consta de las diferencias cuadráticas de coeficientes adyacentes, es decir,

$$\text{penalty}(\lambda) = \lambda \cdot \sum_{j=k+1}^p (\Delta^k \beta_j)^2 = \lambda \cdot \beta' \mathbf{P} \beta, \quad (13)$$

donde Δ^k denota diferencias de orden k . Por lo general, las diferencias de orden $k = 1$ o $k = 2$ son usadas. Para nodos equidistantes, estos toman la forma:

$$\Delta^1 \beta_j = \beta_j - \beta_{j-1} \quad y \quad \Delta^2 \beta_j = \beta_j - 2\beta_{j-1} + \beta_{j-2} \quad (14)$$

Generalmente, las diferencias de orden k se pueden definir de forma recursiva como $\Delta^k \beta_j = \Delta^1(\Delta^{k-1} \beta_j)$ con $\Delta^0 \beta_j = \beta_j$. Por lo tanto, las diferencias de segundo orden se pueden calcular como

$$\Delta^2 \beta_j = \Delta^1 \beta_j + \Delta^1 \beta_{j-1} = \beta_j - \beta_{j-1} - (\beta_{j-1} - \beta_{j-2})$$

Mediante la definición $(p-k) \times p$ de matrices de diferencia $\mathbf{D}_k$, es posible escribir las diferencias para todos los parámetros en notación matricial usando el producto $\mathbf{D}_k \beta$. Para $k = 1$ y $k = 2$ las matrices $\mathbf{D}_k$ toma la forma

$$\mathbf{D}_1 = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix} \quad y \quad \mathbf{D}_2 = \begin{pmatrix} -1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & \ddots & \ddots & \ddots \\ & & & 1 & -2 & 1 \end{pmatrix}$$

Para la matriz de penalización $\mathbf{P}_k = \mathbf{D}_k' \mathbf{D}_k$ obtenemos

$$\mathbf{P}_1 = \begin{pmatrix} -1 & 1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix} \quad y \quad \mathbf{P}_2 = \begin{pmatrix} 1 & -2 & 1 & & & \\ -2 & 5 & -4 & 1 & & \\ 1 & -4 & 6 & -4 & 1 & \\ & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{pmatrix}$$

Un parámetro de suavización pequeño λ da lugar a una función con sesgo pequeño pero varianza grande. En el límite $\lambda \rightarrow 0$ no se produce penalización en absoluto. Por el contrario, un parámetro de suavizado grande se traduce en una función con alto sesgo y baja varianza. El límite $\lambda \rightarrow \infty$ conduce a resultados diferentes en función del orden k elegido: la penalización de las diferencias de primer orden conduce a una función constante y penalización de diferencias de segundo orden a un ajuste lineal. Una versión bayesiana de P-splines ha sido introducida por Lang y Brezger (2004) y Brezger y Lang (2006).

Escala de Tiempo

El efecto del tiempo a menudo se puede dividir en una tendencia suave y una componente estacional, es decir

$$f_{tiempo}(t) = f_{tend. suave}(t) + f_{comp. estacional}(t) \quad (15)$$

Con el fin de aproximar la función de tendencia suave podemos usar el mismo tipo de funciones no lineales como las covariables continuas, es decir, P-splines o paseos aleatorios (ver Fahrmeir y Lang (2001a)).

Covariables Espaciales

En esta sección se aborda el modelado de correlación espacial cuando se observan los puntos de datos en diferentes localizaciones. A menudo esta correlación espacial se puede atribuir a diferentes covariables no observables. La construcción de una función espacial donde las funciones de evaluación se correlacionan en todos los lugares es el objetivo en este apartado. A veces hay factores adicionales no observables, cuyo efecto es independiente el uno del otro en diferentes lugares.

$$f_{espacial}(s) = f_{estructurado}(s) + f_{no estructurado}(s) \quad (16)$$

Se considera el caso de que la covariable s representa una ubicación en regiones geográficas conectadas. En este caso, una función espacial suavizada puede ser modelizada por el campo aleatorio de Markov (MRF). Una parte importante en la construcción de los MRF es el conjunto de vecinos que deben de ser definidos para cada región s . Generalmente, los vecinos de un área s consiste en todas las regiones que comparten una frontera común con s . Para definiciones más complejas sobre vecindad, se puede consultar en Besag, York y Mollie (1991). La idea es que las regiones adyacentes son más similares que las ubicaciones arbitrarias.

La función de evaluación anterior $f_{espacial}(s) = \beta_s$ toma la forma

$$\beta_s | \beta_{s'}, s' \neq s \sim N \left(\frac{1}{N_s} \sum_{s' \in \delta_s} \beta_{s'}, \frac{\tau^2}{N_s} \right) \quad (17)$$

donde δ_s denota el conjunto de vecinos de la región s y $N_s = |\delta_s|$ el número de estos vecinos. Una contabilidad antes de la mejora de las irregularidades en el mapa se puede lograr por definir una versión ponderada similar a la de paseos aleatorios de una dimensión, es decir,

$$\beta_s | \beta_{s'}, s' \neq s \sim N \left(\sum_{s' \in \delta_s} \frac{w'_{ss}}{w_{s+}} \beta_{s'}, \frac{\tau^2}{w_{s+}} \right) \quad (18)$$

donde $w_{s+} = \sum_{s' \in \delta_s} w'_{ss}$ y los pesos w'_{ss} dependen de una medida de la distancia entre las regiones s y s' . Una medida de la distancia se puede especificar de acuerdo con uno de los siguientes ejemplos:

- Si uno supone siempre la misma distancia entre las regiones adyacentes, los pesos se convierten en $w'_{ss} = 1$ y la reduce a la Fórmula ?? anterior.
- Los pesos pueden ser elegidos proporcional a la longitud de la frontera común.
- Los pesos se pueden elegir inversamente proporcionales a la distancia euclídea $d(s, s')$ entre los centroides de dos regiones, es decir, $w_{ss'} \propto \exp(-d(s, s'))$.

Para las p regiones, la matriz de diseño X es una matriz de orden $n \times p$ indicando si la observación i pertenece a la región s ($X_{is} = 1$) o no ($X_{is} = 0$).

Hay que tener en cuenta que, la evaluación de la función de una región puede estimarse incluso si no hay observaciones disponibles para esta región. Esto es debido a los supuestos de suavidad incluidos en la distribución a priori. La MRF puede aplicarse también cuando un número relativamente pequeño de ubicaciones exactas $s = (s_x, s_y)$ están disponibles mediante la definición de una estructura de entorno simétrico. Para un gran conjunto de diferentes ubicaciones o si se requiere una estimación de la superficie, existen otros enfoques, más preferibles. Una posibilidad, no implementada para nuestros algoritmos de selección, son GRF (Gaussian Random Field) que asumen una función de correlación bidimensional al modelo de correlación espacial (ver Kammann y Wand (2003) o Kneib (2006)). Otra posibilidad de basar la estimación en splines tensor de productos penalizados bidimensionales. La desventaja de este enfoque (en contraste con un GRF) son las funciones de base anisotrópicas (ver Kneib (2006)). En este caso, las líneas de las funciones de base de los gráficos de contorno no forman círculos, sobre todo para un pequeño grado l . Esto implica, que las diferentes direcciones se tratan desigualmente.

Heterogeneidad no observada

En este apartado se tratan datos que consisten en observaciones repetidas de individuos o dentro de grupos, como grupos o regiones. No puede haber diferencias entre las unidades individuales o grupos que se deben a factores no observados. Para superar este problema, es posible estimar un efecto aleatorio que modela las diferencias entre cada unidad y la media global. Para ello usamos efectos aleatorios gaussianos i.i.d asumiendo los parámetros β_i , $i = 1, \dots, p$ para los individuos p que se distribuyen normalmente de forma independiente con un parámetro de varianza común, es decir,

$$\beta_i \sim N(0, \tau^2) \quad (19)$$

Aquí la distribución conjunta para β es una distribución normal adecuada. Sin embargo, se puede escribir en la misma forma general ?? como todas las otras distribuciones a priori mediante el uso de la matriz de identidad, es decir, $\mathbf{P}_{rand} = \mathbf{I}$. Esta matriz es de rango completo, de modo que el espacio nulo es de dimensión cero y sólo contiene el vector nulo $\mathbf{0} = (0, \dots, 0)'$.

La matriz de diseño $\mathbf{X}$ es de nuevo una matriz 0/1 de incidencia de orden $n \times p$. Si el efecto aleatorio se utiliza para estimar un efecto espacial no estructurado, la matriz de diseño del efecto aleatorio es exactamente idéntica a la que pertenece al efecto espacial estructurado. Un efecto aleatorio incluye un término constante al igual que todas las demás funciones univariantes que se describen en este capítulo. Pero a diferencia de otras funciones univariantes, los efectos aleatorios penalizan el término constante. Esto se puede ver a partir de la matriz de penalización cuyo espacio nulo contiene simplemente el vector nulo.

4. MODELIZACIÓN DE LOS INCENDIOS FORESTALES

En este apartado se explora la evolución de los incendios forestales de acuerdo a las variables meteorológicas y de tipo de combustible junto con el posible efecto espacial del concello.

El modelo se representa por:

$$\begin{aligned}
 \text{logit}(P_k) &= \\
 &= \log \left(\frac{P_k}{1 - P_k} \right) = \\
 &= \beta_0 + f_1(\text{altitud}) + f_2(\text{ta_media}) + f_3(\text{hr}) + f_4(\text{pp}) + \\
 &+ f_5(\text{Sin_Asignar_Modelo}) + f_6(\text{Pastos}) + f_7(\text{Matorral}) + \\
 &+ f_8(\text{Hojarasca}) + f_9(\text{Restos}) + \\
 &+ f_{\text{estr}}(\text{cdconc}) + f_{\text{no estr}}(\text{cdconc})
 \end{aligned} \tag{20}$$

Para simplificar la notación consideramos k como el índice del conjunto de voxels sobre el área de estudio y periodo, por lo tanto P_k denotan la probabilidad de ignición en el voxel k teniendo en cuenta el conjunto de covariables observadas correspondientes a cada voxel. Nuestros modelos seleccionados contiene un intercepto β_0 . Los términos f_i son funciones suavizadas que describen la relación no lineal entre las variables explicativas y el logit de probabilidad de la variable respuesta, estas funciones han sido estimadas usando P-splines cúbicos bayesianos de segundo orden con 20 nudos interiores. La f_{estr} es estimada usando campos aleatorios de Markov y $f_{\text{no estr}}$ usando i.i.d. efectos aleatorios Gaussianos.

Las Figuras 2, 3, 4 y 5 ilustran los efectos de las covariables para la ocurrencia de incendios forestal del modelo STAR de respuesta binomial. Además, la estimación de los grados de libertad del intercepto y las covariables junto con los intervalos de credibilidad al 80 y 95 % se muestran en la Tabla 2

Variables	grados libertad	IC al 95 %
cte	1	(-9.3666, -7.5362)
altitud	1.089	Ver Figura 2
ta_media	4.075	Ver Figura 2
hr	5.183	Ver Figura 3
pp	1.071	Ver Figura 3
Sin Asignar Modelo	3.643	Ver Figura 4
Pasto	1.009	Ver Figura 4
Matorral	1.014	Ver Figura 5
Hojarasca	2.149	Ver Figura 5
Restos	1.001	Ver Figura 5

Tabla 2: Estimación de los grados de libertad e intervalos de credibilidad al 80 y 95 %

Los efectos de las covariables altitud y ta_media se representan en la Figura 2. La covariable altitud presenta (ver panel izquierdo) un descenso acusado de la ocurrencia de incendios forestales a medida que aumenta su valor, esto es debido, a que en el periodo de estudio considerado en este trabajo, la mayoría de los incendios ocurrieron en zonas costeras en donde las altitudes tendieron a ser bajas. El efecto de la covariable ta_media (ver panel derecho) parece tener un efecto creciente a lo largo de toda la función, más acusado a temperaturas bajas, inferiores a 20 °C. A partir de esta temperatura, la ocurrencia de incendios para temperaturas comprendidas entre 20 y 28 °C se estabiliza, mostrando a partir de los 28 °C aproximadamente, un efecto creciente no tan drástico como en temperaturas bajas (< 20°C).

El efecto de la covariable hr (ver Figura 3, panel izquierdo) es creciente para valores bajos hasta un 30 % aproximadamente, donde este efecto pasa a ser decreciente y más acusado para valores altos de humedad relativa (entorno al 75 %). Incrementos de humedad relativa están asociados a un descenso de ocurrencia de incendios. El efecto de la covariable pp (ver Figura 3, panel derecho)

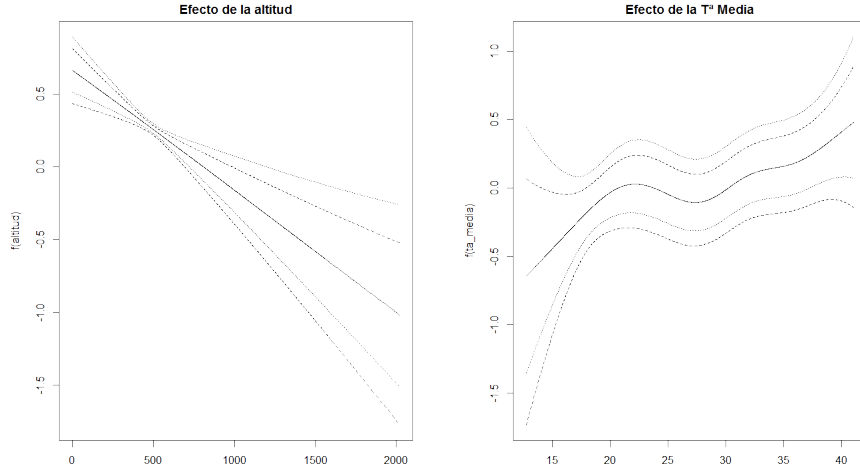


Figura 2: Estimación de los efectos suavizados de la altitud (figura izquierda) y ta_media (figura derecha) con sus correspondientes bandas de credibilidad punteadas al 80 y 95 % - Ocurrencia de incendios forestales totales

tiene un efecto prácticamente lineal y decreciente. La inexistencia de precipitaciones es condición necesaria pero no suficiente para la ocurrencia de incendios. Hay que señalar que las bandas de credibilidad para valores superiores a 5 l/m_2 se hacen más amplias, por lo que de nuevo, no se debe sobreinterpretar los efectos debido a la incertidumbre de estos resultados. La estimación de los grados de libertad para ambas covariables se recogen en la Tabla 2.

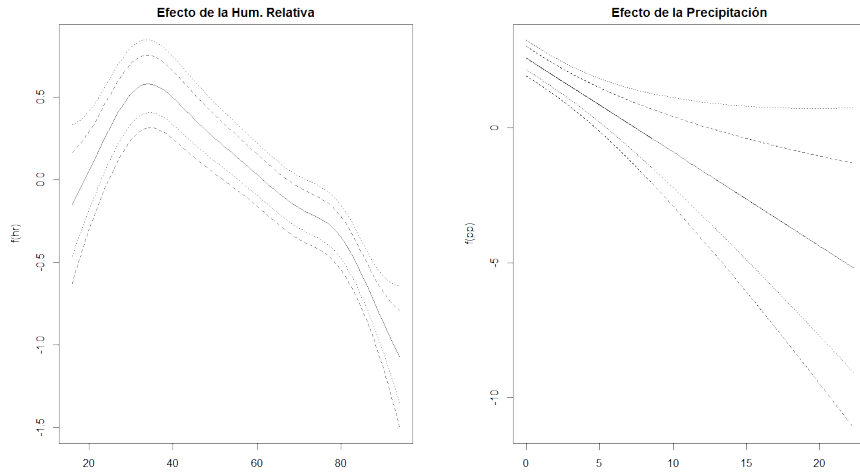


Figura 3: Estimación de los efectos suavizados de las covariables hr (figura izquierda) y pp (figura derecha) con sus correspondientes bandas de credibilidad al 80 y 95 % - Ocurrencia de incendios forestales totales

Los efectos de la covariable Sin asignar modelo (ver Figura 4, panel izquierdo) es no lineal y creciente. Según aumenta la cantidad de superficie sin conocimiento del tipo de estructura vegetal por la cual está formada, la ocurrencia de incendios forestales es mayor. De nuevo, las bandas de credibilidad se separan según aumenta la superficie de esta variable, por lo que no se debe de sobreinterpretar los efectos debido a la incertidumbre de estos resultados.

Los efectos de grupo de Pastos (ver Figura 4, panel derecho) muestran un efecto lineal creciente más acentuado para la variable matorral que en la variable pastos. Además, hay que destacar que la variable pasto, a medida que aumenta la superficie de este tipo de combustible por cuadrícula asociada a un concello, las bandas de credibilidad al 80 y 95 % se van separando del efecto de la variable, esto es debido a la escasa cantidad en m^2 de pasto. Por el contrario, el efecto de la variable

grupo de matorral, muestra que a medida que aumenta la cantidad de superficie en cada cuadrícula de esta agrupación de modelos de combustible, la ocurrencia de incendio forestal también es mayor.

La estimación de los grados de libertad de ambas covariables estudiadas se recogen en la Tabla 2.

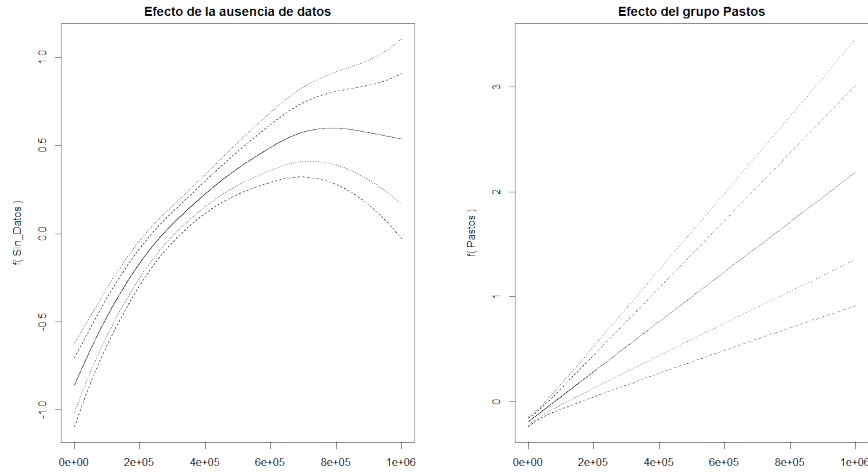


Figura 4: Estimación de los efectos suavizados de la variable Sin Asignar Modelo (figura izquierda) y grupo de Pastos (figura derecha) con sus correspondientes 80 y 95 % de bandas de credibilidad - Ocurrencia de incendios forestales totales

Los efectos de las variables grupo de matorral, hojarasca y de restos se muestran en la Figura 5 donde se observa que la primera de las variables, la variable grupo de matorral, muestra que a medida que aumenta la cantidad de superficie en cada cuadrícula de esta agrupación de modelos de combustible, la ocurrencia de incendio forestal también es mayor. Por su parte, el grupo de hojarasca, presenta un efecto no lineal suave y creciente hasta valores en torno a los 400.000 m^2 de superficie ocupada por hojarasca en donde a partir de aproximadamente de esa superficie, el efecto se estabiliza hasta el final (1 millón de m^2 de superficie). La última variable de estudio, variable grupo de restos, (ver Figura 5, panel derecho) presenta un efecto lineal y constante entorno al cero que según aumenta la cantidad restos de corta en superficie en cada cuadrícula, las bandas de credibilidad al 80 y 95 % se hacen más amplias, lo que de nuevo alerta de la escasa concentración en m^2 del grupo de modelos de combustible de restos de cortas y/o operaciones selvícolas.

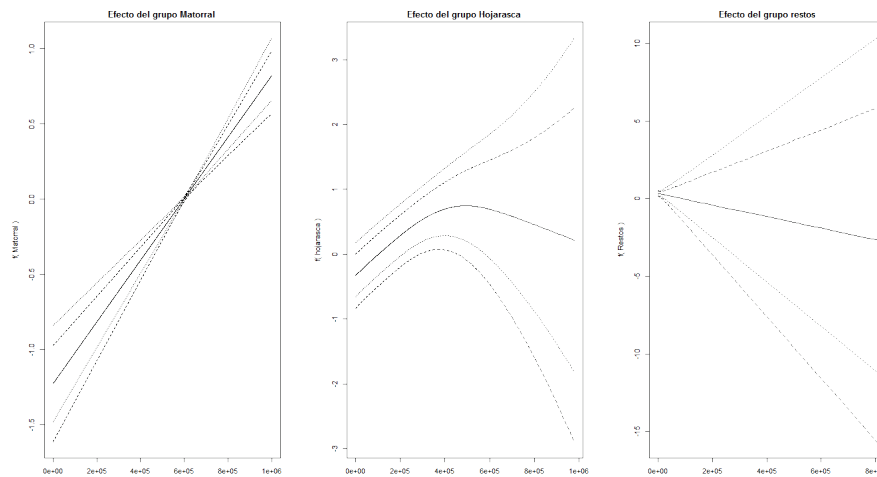


Figura 5: Estimación de los efectos suavizados de los grupos Matorral, Hojarasca y Restos con sus correspondientes 80 y 95 % de bandas de credibilidad - Ocurrencia de incendios forestales totales

Los efectos espaciales se muestran en las Figuras 6 y 7. Los efectos espaciales estructurados

(Figura 6) marcan una tendencia de mayor ocurrencia de incendios hacia el oeste y sur de Galicia frente a la menor ocurrencia de incendios cara el norte y este gallego. Los intervalos de credibilidad al 95 % (panel derecho) muestran concellos con gran probabilidad de ocurrencia de incendios (concellos en negro) y baja o ninguna probabilidad de ocurrencia de incendios (concellos en blanco). Los concellos en gris son concellos no significativos frente a la variable en estudio. Los resultados de las efectos espaciales no estructurados (ver Figura 7) no apuntan en ninguna dirección específica. Todos los concellos son no significativos (color gris, panel derecho) excepto uno, Chandrexa de Queixa, que obliga a un análisis más detallado de cual ha sido la causa de incendios de este municipio que los diferencia de los de su entorno y del resto de Galicia.

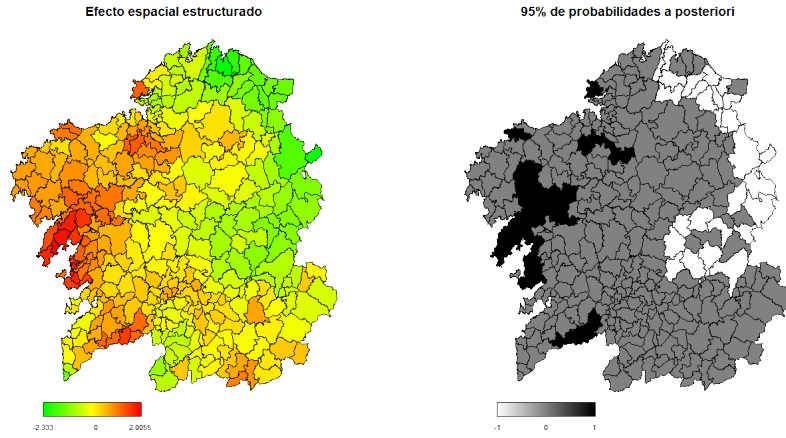


Figura 6: Efectos espaciales estructurados - Ocurrencia de incendios forestales totales

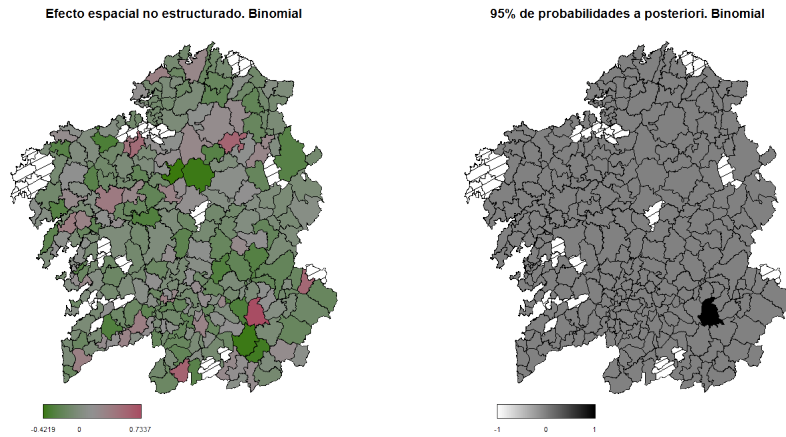


Figura 7: Efectos espaciales no estructurados - Ocurrencia de incendios forestales totales

El ajuste obtenido en la modelización se muestra en el Tabla 3.

Modelo	2xlog-likelihood	g.l.	AIC	BIC	GCV
Binomial (totales)	13993.4	164.997	14323.4	16065.3	0.049

Tabla 3: Criterio de información Akaike (AIC), criterio de información Bayesiano (BIC), y validación cruzada generalizada (GCV) para el modelo.

Los resultados obtenidos utilizando modelización STAR parecen indicar que las variables meteorológicas pueden ser usadas para la estimación de la probabilidad de incendio. En cambio, las

variables de grupos de modelos de combustibles deben de ser sustituidas por otro tipo de variables como por ejemplo, las formaciones vegetales, que aporten mayor información al modelo.

3. CONCLUSIONES

Los modelos STAR permiten superar limitaciones de otras metodologías aplicadas hasta la actualidad para comprender un fenómeno complejo como es la ignición de incendios forestales y que en su mayoría son provocados. Esto es debido a la capacidad de introducir la componente espacial como un factor más del modelo, lo hace de forma que se pueden analizar por separado de forma fácil y visual aspectos espaciales estructurados relacionados a una escala comarcal, que estan relacionados con problemas de conflictividad debidos al tipo de propiedad, la gestión forestal o conflictos de uso entre diferentes colectivos, como puede ser el caso de ganaderos, propietarios particulares o comuneros de montes vecinales, la gestión por parte de la Administración Forestal de convenios y consorcios o la presencia de otros colectivos como ganaderos, cazadores, ecologistas, etc. que tienen sus propios objetivos de uso del monte. Estos resultados pueden ser muy útiles a efectos de propuesta de medidas políticas y de planificación forestal que reduzcan la problemática.

A su vez la metodología STAR permite obtener una explicación del fenómeno estudiado, en este caso los incendios, de su comportamiento local. Este resultado es de gran importancia pues permite visualizar en que zonas existen aspectos especiales relacionados con la presencia de un pirómano o un conflicto específico y no trasladable a nivel comarcal, como puede ser el caso de problemas de herencias, parques eólicos, presencia de minas y canteras, etc. La solución a los incendios va a pasar siempre por este problema específico que ha sido detectado.

Unido a lo anterior esta la flexibilidad de los modelos STAR que permiten incorporar otras variables que van a permitir conocer con mayor detalle y ayudara a anticipar se en la prevención y el cálculo de riesgo de incendios con la consecuente mejora del principal riesgo ambiental pero también social y económico anual de regiones como Galicia.

REFERENCIAS

- Mermoz, M., Kitzberger, T., Veblen, T.T. (2005). Landscape influences on occurrence and spread of wildfires in Patagonian forests and shrublands. *Ecology* 86, 2705-2715.
- Moreira, F., Vaz, P., Catry, F., Silva, J.S. (2009). Regional variations in wildfire susceptibility of land-cover types in Portugal: implications for landscape management to minimize fire hazard. *International Journal of Wildland Fire* 18, 563-574.
- National Interagency Fire Center (2011). The science of wildland fire (available in: <http://www.nifc.gov/>).
- Pyne S.J., Andrews P.L., Laven R.D. (1996). *Introduction to wildland fire*, 2nd edn. Wiley.
- Badia A., Pallarès M., Piñol J. (2003). Incendis periurbans front a incendis rurals. Els casos de Àrea Metropolitana de Barcelona i la comarca del Bages. *Documents Anàlisi Geogràfica* 41, 17-42.
- Badia-Perpinya, A., Pallares-Barbera, M. (2006). Spatial distribution of ignitions in Mediterranean periurban and rural areas: the case of Catalonia. *International Journal of Wildland Fire* 15, 187-196.
- SanMiguel-Ayanz, J. (2008). Integration of socio-economic and environmental variables for modelling long-term fire danger in southern Europe. *European Journal of Forest Research* 127, 149-163.
- González, J.R., Pukkala, T., (2007). Characterization of forest fires in Catalonia (north-east Spain). *European Journal of Forest Research* 126, 421-429.
- González-Olabarria, J.R., Mola-Yudego, B., Pukkala, T., Palahi, M. (2011). Using multiscale spatial analysis to assess fire ignition density in Catalonia, Spain. *Annals of Forest Science* 68, 861-871.
- Juan, P., Mateu, J., Saenz, M. (2012). Pinpointing spatio-temporal interactions in wildfire patterns. *Stoch Environ Res Risk Assess* (in press).

- Romero-Calcerrada, R., Novillo, C.J., Millington, J.D.A., Gomez-Jimenez, I. (2008). GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (Central Spain). *Landscape Ecology* 23, 341-354.
- Badia-Perpinya, A., Pallares-Barbera, M. 2006. Spatial distribution of ignitions in Mediterranean periurban and rural areas: the case of Catalonia. *International Journal of Wildland Fire* 15, 187-196.
- Catry F.X., Rego F.C., Baao, F., Moreira, F. (2009). Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire* 18, 921-931.
- Chuvieco, E., Aguado, I., Yebra, M., Nieto, H., Salas, J., Martn, M.P., Vilar, L., Martnez, J., Martin, S., Ibarra, P., de la Riva, J., Baeza, J.,Rodrguez, F., Molina, J.R., Herrera M.A., Zamora, R. (2010). Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling* 221, 46-58.
- Fernandes, P. (2009). Combining forest structure data and fuel modelling to classify fire hazard in Portugal. *Annals of Forest Science* 66, 415-415.
- Kalabokidis, K.D., Koutsias, N., Konstantinidis, P., Vasilakos, C. (2007). Multivariate analysis of landscape wildfire dynamics in a Mediterranean ecosystems of Greece. *Area* 39, 392 - 402.
- Martnez, J., Chuvieco, E., Martn, P. (2008). Estimation of risk factors of human ignition of fires in Spain by means of logistic regression. *Proceedings of the Second International Symposium on Fire Economics, Planning, and Policy: A Global View*, 265-278.
- Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., Barbati, A., Corona, P., Vaz, P., Xanthopoulos, G., Mouillot, F. & Bilgili, E. (2011). Landscape wildfire interactions in southern Europe: implications for landscape management. *Journal of Environmental Management* 92, 2389-2402.
- Vzquez A., Moreno J.M. (1998). Patterns of lightning- and human-caused fires in peninsular Spain. *International Journal of Wildland Fire* 8, 103-115.
- Acosta, A., Carranza, M.L., Giancola, M. (2005). Landscape change and ecosystem classification in a municipal district of a small city (Isernia, central Italy). *Environmental Monitoring and Assessment* 108, 323-335.
- Aranzabal, I., Schmitz, M.F., Aguilera, P., Pineda, F.D., (2008). Modelling of landscape changes derived from the dynamics of socio-ecological systems: a case of study in a semiarid Mediterranean landscape. *Ecological Indicators* 8, 672-685.
- Faluccci, A., Maiorano, L., Boitani, L., (2007). Changes in land-use/land-cover patterns in Italy and their implications for biodiversity conservation. *Landscape Ecology* 22, 617-631.
- Bajocco, S., Ricotta, C. (2008). Evidence of selective burning in Sardinia (Italy): which land-cover classes do wildfires prefer? *Landscape Ecology* 23, 241-248. 323-335.
- Daz-Delgado, R., Lloret, F., Pons, X., (2004). Spatial patterns of fire occurrence in Catalonia, NE Spain. *Landscape Ecology* 19, 731-745.
- Gonzlez-Olabarria, J.R., Mola-Yudego, B., Pukkala, T., Palahi, M. (2011). Using multiscale spatial analysis to assess fire ignition density in Catalonia, Spain. *Annals of Forest Science* 68, 861-871.
- Gonzlez, J.R., Pukkala, T., (2007). Characterization of forest fires in Catalonia (north-east Spain). *European Journal of Forest Research* 126, 421-429.
- Koutsias, N., Arianoutsou, M., Kallimanis, A.S., Dimopoulos, P. (2009). Is there any special pattern of the extreme wildland fires occurred in Greece in the summer of 2007? In: 52 nd International Symposium of the International Association for Vegetation Science, Vegetation Processes and Human Impact in a Changing World, Chania, Greece.
- Mouillot, F., Ratte, J., Joffre, R., Mouillot, D., Rambal, S., (2005). Long-term forest dynamic after land abandonment in a fire prone Mediterranean landscape (central Corsica, France). *Landscape Ecology* 20, 101-112.
- Nunes, M.C.S., Vasconcelos, M.J., Pereira, J.M.C., Dasgupta, N., Alldredge, R.J., Rego, F.C., (2005). Land cover type and fire in Portugal: do fires burn land cover selectively. *Landscape Ecology* 20, 661-673.

- Sebastian-Lopez, A., Salvador-Civil, R., Gonzalo-Jimenez, J., SanMiguel-Ayanz, J. (2008). Integration of socio-economic and environmental variables for modelling long-term fire danger in southern Europe. *European Journal of Forest Research* 127, 149-163.
- Arianoutsou, M., (2001). Landscape changes in Mediterranean ecosystems of Greece: implications for fire and biodiversity issues. *Journal of Mediterranean Ecology* 2, 165-178.
- Cardille, J.A., Ventura, S.J., Turner, M.G. (2001). Environmental and social factors influencing wildfires in the Upper Midwest, United States. *Journal of Applied Ecology* 11, 111?127.
- Curt, T., Delcros, P. (2010). Managing road corridors to limit fire hazard. A simulation approach in southern France. *Ecological Engineering* 4, 1-12.
- Loboda T.V., Csizar I.A. (2007). Assessing the risk of ignition in the Russian Far East within a modeling framework of fire threat. *Journal of Applied Ecology* 17, 791? 805.
- Romero-Calcerrada, R., Novillo, C.J., Millington, J.D.A., Gomez-Jimenez, I. (2008). GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (Central Spain). *Landscape Ecology* 23, 341-354.
- Vasconcelos, M.J.P., Silva, S., Tome, M., Alvin, M., Pereira, J.M.C. (2001). Spatial prediction of FIRE ignition probabilities: comparing logistic regression and neural networks. *Photogrammetric Engineering & Remote Sensing* 67, 73?81
- Vilar, L., Woolford, D.G., Martell, D.L., Pilar-Martín, M. (2010). A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. *International Journal of Wildland Fire* 19, 325-337.
- Wittenberg, L., Malkinson, D., (2009). Spatio temporal perspectives of forest fires regimes in a maturing Mediterranean mixed pine landscape. *European Journal of Forest Research* 128, 297-304.
- Zea-Bermudez, P., Mendes, J., Pereira, J.M.C., Turkman, K.F., Vasconcelos, M.J.P. (2009). Spatial and temporal extremes of wildfire sizes in Portugal (1984?2004). *International Journal of Wildland Fire* 18, 983-991.
- Cardille, J.A., Ventura, S.J., Turner, M.G. (2001). Environmental and social factors influencing wildfires in the Upper Midwest, United States. *Journal of Applied Ecology* 11, 111?127.
- Catry F.X., Rego F.C., Baçao, F., Moreira, F. (2009). Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire* 18, 921-931.
- Vilar del Hoyo, L., Gómez Nieto, I., Martín Isabel, M.P., Martínez Vega, F.J. (2007). Análisis comparativo de diferentes métodos para la obtención de modelos de riesgo humano de incendios forestales. 4th International Wildland fire Conference-Wildfire 2007. Sevilla.
- Amatulli, G., Pérez-Cabello, F., de la Riva, J. (2007). Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. *Ecological Modelling* 200, 321-333.
- Romero-Calcerrada, R., Barrio-Parra, F., Millington, J.D.A., Novillo, C.J. (2010). Spatial modelling of socioeconomic data to understand patterns of human-caused wildfire ignition risk in the SW of Madrid (central Spain). *Ecological Modelling* 221, 34-45.
- Wood, S.N., Augustin, N.H. (2002). GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling* 157, 157-177.
- Verde, J.C., Zêzere, J.L. (2010). Assessment and validation of wildfire susceptibility and hazard in Portugal. *Natural Hazards and Earth System Sciences* 10, 485-497.
- Benavet-Corai, J., Rojo, C., Suárez-Torres, J., Velasco-García, L. (2007). 2017,86 *Ecological Modelling* 205, 336 - 342.
- Mandallaz, D., Ye, R., (1997). Prediction of forest fires with Poisson models. *Canadian Journal of Forest Research* 27, 1685-1694.
- Wotton B.M., Martell D.L., Logan K.A. (2003). Climate change and people-caused forest fire occurrence in Ontario. *Climate Change* 60, 275-295
- de la Riva, J., Pérez-Cabello, F., Lana-Renault, N., Koutsias, N. (2004). Mapping wildfire occurrence at regional scale. *Remote Sensing of Environment* 92, 288 - 294

- Fahrmeir, L., & Kneib, T., (2011). Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data. Oxford University Press, USA.
- Fahrmeir, L., Kneib, T. & Lang, S.,(2004). Penalized additive regression for spacetime data: a Bayesian perspective. *Statistica Sinica* 14, 731-761.
- Hastie, T.J. & Tibshirani, R.J.,(1990). Generalized Additive Models. Chapman and Hall.
- Rigby, R.A. & Stasinopoulos, D.M.,(2005). Generalized Additive Models for Location, Scale and Shape. *Applied Statistics* 54, 507-554.
- Wood, S.N.,(2006a). Generalized Additive Models: An Introduction with R. Chapman and Hall.
- Ruppert, D., Wand, M.P., & Carroll, R.J.,(2003). Semiparametric Regression. Cambridge University Press, Cambridge.
- Fahrmeir, L. & Lang, S.,(2001a). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society C* 50, 201-220.
- Kammann, E. & Wand, M.,(2003). Ge additive models. *Journal of the Royal Statistical Society C* 52, 1-18.
- Fotheringham, A., Brunsdon, C., & Charlton, M.,(2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester.
- Hastie, T.J. & Tibshirani, R.J.,(1990). Generalized Additive Models. Chapman and Hall.
- Chen, Z.,(1993). Fitting multivariate regression functions by interaction spline models. *Journal of the Royal Statistical Society B* 55 (2), 473-491.
- Green, P.J. & Silverman, B.W.,(1994). Nonparametric Regression and Generalized Linear Models. Chapman & Hall, London.
- Fan, J. & Gijbels, I.,(1984). Local Polynomial Modelling and Its Applications. Chapman & Hall, London. and Hall.
- Loader, C.,(1999). Local Regression and Likelihood. Springer, New York.
- De Boor, C.,(1978). A Practical Guide to Splines. Springer, New York.
- Dierckx, P.,(1995). Curve and Surface Fitting with Splines. Oxford: Clarendon Press.
- Biller, C.,(2000). Adaptive bayesian regression splines in semiparametric generalized linear models. *Journal of Computational and Graphical Statistics* 9, 122-140.
- Eilers, P.H. & Marx, B.D.,(1996). Flexible smoothing with B-splines and Penalties. *Statistical Science* 11, 89-121.
- Marx, B.D. & Eilers, P.,(1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics and Data Analysis* 28 (2), 193-209.
- Brezger, A. & Steiner, W.,(2006). Monotonic Regression based on Bayesian P-Splines: An application to estimating price response functions from store-level scanner data. *Journal of Business and Economic Statistics*, to appear.
- Besag, J., York, J., & Mollie, A.,(1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* 43 (1), 1-59.
- Kneib, T.,(2006). Mixed model based inference in structured additive regression. PhD thesis, Dr.Hut-Verlag.
- Kammann, E. & Wand, M.,(2003). Ge additive models. *Journal of the Royal Statistical Society C* 52, 1-18.
- Fahrmeir, L. & Tutz, G.,(2001). Multivariate Statistical Modelling Based on Generalized Linear Models. Springer.
- Conedera, M., Torriani, D., Neff, C., Ricotta, C., Bajocco, S., Pezzatti, G.B. (2011). Using Monte Carlo simulations to estimate relative fire ignition danger in a low-to-medium fire-prone region. *Forest Ecology and Management* 261, 2179-2187.