



MÁSTER EN TÉCNICAS ESTADÍSTICAS

TRABAJO FIN DE MÁSTER

Modelos de Regresión Aditivos Estructurados (STAR) con respuesta binaria para la predicción de incendios forestales en Galicia. Agosto 2006

Autora: Laura Ríos Pena

Codirectores: Carmen M^a Cadarso Suárez
Manuel Fco. Marey Pérez

JULIO DE 2013

Este trabajo se presenta para su evaluación como Trabajo Fin de Máster, en la modalidad de prácticas en empresas, de la alumna Laura Ríos Pena.

Ha sido realizado bajo la dirección de Carmen María Cadarso Suárez y Manuel Francisco Marey Pérez que autorizan la entrega del mismo.

A Coruña, a 1 de julio de 2013

Carmen M^a Cadarso Suárez

Manuel Fco. Marey Pérez

Agradecimientos

Debo agradecer de una manera especial y sincera a la Profesora Carmen Cadarso Suárez por aceptarme para realizar este trabajo fin de máster bajo su dirección. Su apoyo y confianza en mi trabajo ha sido un aporte invaluable, no solamente en el desarrollo del trabajo, sino también en mi formación como investigadora. Las ideas propias, siempre enmarcadas en su orientación y rigurosidad, han sido la clave del buen trabajo que hemos realizado juntas.

Quiero expresar también mi más sincero agradecimiento al profesor Manuel Marey Pérez por su importante aporte y participación activa en el desarrollo de este trabajo. Debo destacar, por encima de todo, su disponibilidad y paciencia que hizo que nuestras siempre acaloradas discusiones redundaran benéficamente tanto a nivel científico como personal. No cabe duda que su participación ha enriquecido el trabajo realizado.

A Luís Franco, por su ayuda computacional y todos los buenos consejos a lo largo del trabajo.

Gracias a Biostatech, Advice, Training & Innovation in Biostatistics, S.L. por aceptarme para realizar las prácticas en empresa y, especialmente a Vicente Lustres Pérez que ha ejercido de tutor a lo largo de mi práctica formativa, darle las gracias por su constante motivación y confianza brindada en el desarrollo del trabajo cada día.

Por último quiero dar las gracias al profesorado del Programa Interuniversitario del Tercer Ciclo de Estadística e Investigación Operativa de las Universidades de A Coruña, Santiago de Compostela y Vigo por la formación recibida.

Índice general

1. Introducción	5
2. Área de Estudio y Material de Partida	9
2.1. Área de Estudio	9
2.2. Origen y descripción de los datos utilizados	10
2.2.1. Mapa de concellos de Galicia	13
2.2.2. Base de Datos Climatológicos	14
2.2.3. Modelo Digital del Terreno (MDT)	15
2.2.4. Cuarto Inventario Forestal Nacional (IFN-4)	16
3. Metodología estadística	21
3.1. Modelo de Regresión Aditivo Estructurado (STAR)	21
3.2. Las componentes del modelo	23
3.2.1. Efectos Lineales	24
3.2.2. Variables Categóricas	25
3.2.3. Covariables Continuas	26
3.2.4. Escala de Tiempo	28
3.2.5. Covariables Espaciales	29
3.2.6. Heterogeneidad no observada	30
3.2.7. Coeficientes de Variación	31
3.3. Criterios de Selección del modelo	32
3.3.1. Criterio de Información de Akaike (AIC)	32

3.3.2.	Criterio de Información Bayesiano (BIC)	33
3.3.3.	Validación Cruzada Genaralizada (GCV)	34
4.	Explorando la incidencia de incendios forestales totales y provocados	37
4.1.	Modelización a partir de variables meteorológicas	38
4.1.1.	Corrección de los datos meteorológicos	38
4.1.2.	Modelización de los incendios forestales totales	39
4.1.3.	Modelización de los incendios forestales provocados	43
4.1.4.	Ajuste de modelos	47
4.2.	Modelización a partir de grupos de modelos de combustible	48
4.2.1.	Modelización de los incendios forestales totales	48
4.2.2.	Modelización de los incendios forestales provocados	53
4.2.3.	Ajuste de modelos	58
4.3.	Modelización a partir de variables climatológicas y grupos de modelos de combustible	58
4.3.1.	Modelización de los incendios forestales totales	59
4.3.2.	Modelización de los incendios forestales provocados	65
4.3.3.	Comparativa de modelos	71
5.	Conclusiones y Futuras Líneas de Investigación	73
A.	Base de datos	75
A.1.	Variables Meteorológicas	75
A.1.1.	Temperatura Media Diaria	76
A.1.2.	Humedad Relativa Media	77
A.1.3.	Precipitaciones	78
A.2.	Variables de grupos de modelos de combustible de Rothermel	79
A.2.1.	Sin Asignar Modelo	80
A.2.2.	Grupo de Pastos	81

A.2.3. Grupo de Matorral	81
A.2.4. Grupo de Hojarasca bajo arbolado	82
A.2.5. Grupo de Restos y/o Operaciones Selvícolas	83
B. Software	85
B.1. Un ejemplo de código R para modelos STAR	86
B.1.1. Script de la modelización de la ocurrencia de incendios totales con variables climatológicas	86
B.1.2. Script para la representación de los efectos de las covariables continuas	86
B.1.3. Script para dibujar los efectos espaciales estructurados y no estructurados	87
Bibliografía	87

Capítulo 1

Introducción

Los incendios ocurren en multitud de ecosistemas forestales (Moritz, 2003), con una gran influencia en las consecuencias ecológicas (Minnich & Bahre, 1995) y las funciones económicas del bosque (Hardy, 2005). Estos fenómenos, al igual que otros como avalanchas, terremotos, tormentas de arena, etc., tienen la propiedad de que cuando ocurren por encima de un determinado umbral producen una cascada de actividad ambiental, social y económica (Malamud et al., 1998). A su vez experimentan comportamientos muy distintos según los diferentes ámbitos en los que se puedan localizar (Reed & Mckelvey, 2002), variando en función de variables naturales y socioeconómicas ligadas a los paisajes forestales y presentando comportamientos difícilmente parametrizables.

Se sabe que los incendios naturales han sido esenciales para mantener la biodiversidad y también han sido una herramienta ampliamente utilizada para la gestión del territorio (Moreno 1989). Sin embargo, en las últimas décadas, los incendios causados de forma natural han experimentado alteraciones significativas (frecuencia de incendios, intensidad y gravedad) que han agravado la ecológica, consecuencias sociales y económicas (Westerling et al. 2006, FAO 2007). Los cambios en los patrones tradicionales del uso de la tierra derivados de la transformación socio-económica en las zonas rurales (abandono rural) han modificado recientemente la incidencia de incendios debido a una acumulación inusual de los combustibles forestales, que en particular aumenta el riesgo de incendio y de la severidad del fuego. Además, el uso cada vez mayor de los bosques como recurso recreativo implica una mayor incidencia de los incendios de origen humano, ya sea por negligencia o incendio (Chuvieco et al. 1999). La importancia de contar con información precisa y fiable sobre la incidencia de los incendios forestales, así como sobre las causas del fuego es evidente. El conocimiento de los mecanismos que hay detrás de la ocurrencia de incendios espera mejorar las actividades de prevención de incendios (estimación del riesgo de incendio) y reducir su impacto ecológico negativo, impactos económicos, sociológicos y ambientales.

En Galicia, la problemática de los incendios forestales se observa a partir de 1961, año en el que comienzan las estadísticas sobre incendios forestales, hasta octubre de 2011. En este periodo se han producido en Galicia 245.593 incendios forestales. Estos incendios arrasaron una superficie de 1.794.578 ha, lo que equivale a (62%) del área geográfica de la Comunidad Autónoma y casi todos (87%) de la superficie forestal total. Por ello los incendios forestales son percibidos como el principal problema medioambiental de la región (Alonso-Betanzos et al, 2003; Chas, 2007). En la Figura 1.1 se observa esta evolución.

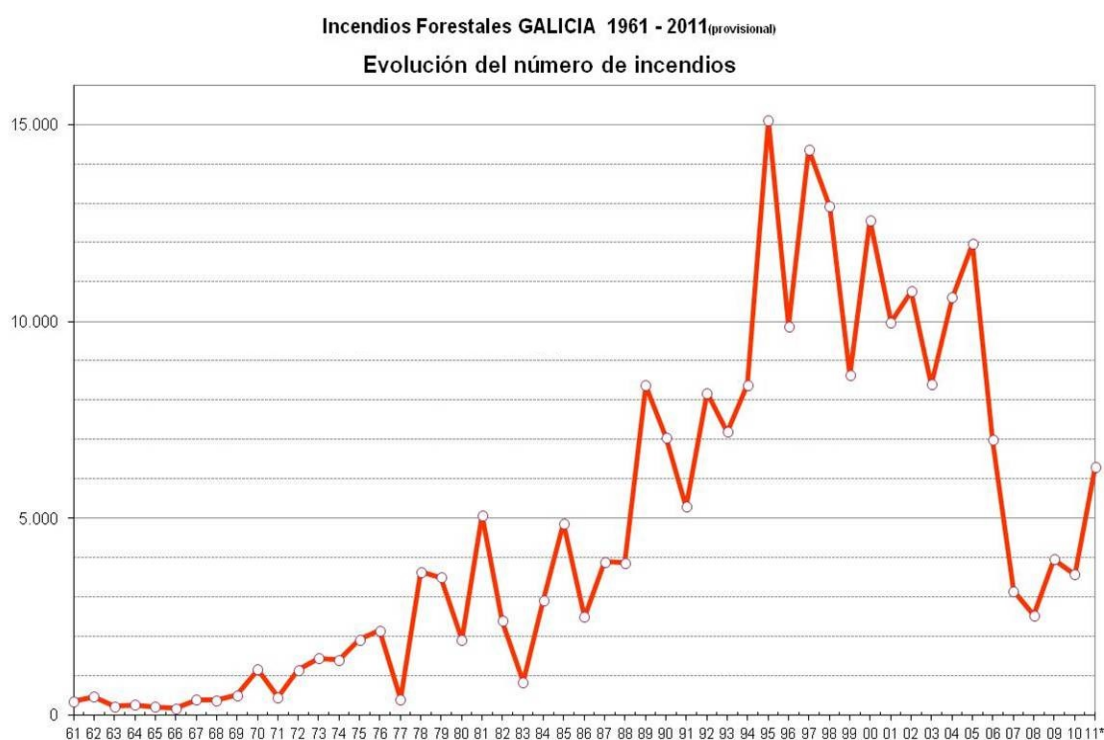


Figura 1.1: Evolución de los incendios forestales desde 1961-2011. Fuente: Consellería de Medio Rural e do Mar - Xunta de Galicia

Diferentes técnicas han sido desarrolladas para predecir el comportamiento del fuego. Técnicas de regresión logística han sido comúnmente utilizadas para referirse al riesgo y a la ocurrencia de incendios, y para determinar los días en los que se produzcan grandes incendios o varios incendios, mediante el uso de, por ejemplo, modelos logísticos clásicos (Brillinger de 2003, González, et al., 2006, González y Pukkala, 2007), modelos logísticos con funciones spline (Preisler y Westerling, 2007) y modelos de regresión logística con componentes espaciales y temporales (Preisler et al, 2004). Cunningham y Martell (1973), Martell et al. (1987, 1989) y Anderson

et al. (2000) han utilizado la regresión logística para estudiar la relación entre los índices de riesgo y la probabilidad de incendio por día. Chou et al. (1990, 1993) aplicó la regresión logística incluyendo el clima como variable explicativa para evaluar la autocorrelación espacial entre los incendios ubicados cerca. Preisler et al. (2004) también utiliza la regresión logística para estimar la probabilidad de la ocurrencia de incendios. Además de los modelos logísticos, otros modelos de regresión se han utilizado también, como los modelos lineales generalizados (Preisler y Westerling, 2007) y los modelos aditivos generalizados con efectos fijos, efectos aleatorios y de efectos mixtos (Breslow y Clayton, 1993; Brillinger, y col ., 2003, Brillinger et al. 2004, Brillinger et al. 2006) para evaluar las probabilidades asociadas con la ocurrencia de incendios forestales como por ejemplo, la probabilidad de que pueda ocurrir un incendio en una región determinada durante un día determinado, semana o mes del año. Diversas técnicas de aprendizaje se han desarrollado, incluyendo redes neuronales, que se utilizan para determinar la propagación del fuego (Alonso-Betanzos et al 2003;. Loehle, 2004) y para determinar la mejor posición para los equipos de extinción aéreos (Varela, 2002), así como los métodos de simulación (Reinhardt, et al. 2001).

Hay una serie de estudios que analizan la ocurrencia de incendios forestales, que se define como el número de incendios por unidad espacial, como se trata en este trabajo, de manera que la variable respuesta en esos modelos se refiere a la ocurrencia, comúnmente obtenida a partir de información disponible en bases de datos de incendios forestales, sin embargo, la precisión de la ubicación espacial de los incendios varía significativamente dependiendo de la fuente de datos. Como resultado, una amplia gama de posibilidades se pueden encontrar en la literatura. Algunos autores han utilizado como variable de respuesta en sus modelos estadísticos ubicaciones exactas a partir de las coordenadas geográficas (coordenadas x , y) de los puntos de ignición (Cardille et al 2001; Pew y Larsen 2001; Vasconcelos et al. 2001; Koutsias et al. 2004; Robin et al. 2006; Kalabokidis et al. 2007; Yang et al. 2007). Otros autores han utilizado como variable respuesta en sus modelos menos exactos información sobre los incendios, espacialmente colocados en una rejilla y/o diferentes unidades administrativas (municipio, provincia, etc.).

El presente trabajo persigue la obtención de modelos predictivos 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) para generar modelos predictivos especializados a una resolución de $1 \text{ Km} \times 1 \text{ Km}$ asociando las cuadrículas de la rejilla al concello al que pertenecen y tomando como variables respuesta información para los incendios provocados y otras causas. 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 dos tipos de variable respuesta para un mismo modelo de variables explicativas; la ocurrencia de incendios forestales provocados y totales por concello.
- Generar modelos de regresión aditivos estructurados (STAR) de respuesta binaria para ambas variables respuesta.
- Validar los resultados obtenidos en cada modelo de regresión aditivo estructurado (STAR) con respuesta binaria.

El propósito final de la línea de trabajo es desarrollar modelos consistentes espacio-temporalmente 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.

El documento se estructura en un primer apartado que describe el área de estudio, junto con material de partida (bases de datos) y una breve descripción de las variables. Posteriormente en el capítulo 3 se presenta la metodología en el estudio. Los resultados se discuten en el capítulo 4, y finalmente las conclusiones y futuras líneas de investigación se describen en el capítulo 5.

Capítulo 2

Área de Estudio y Material de Partida

2.1. Área de Estudio

Los modelos de ocurrencia de incendio debidos a condiciones climáticas y de tipo de combustible para cualquier causa de incendio forestal han sido elaborados para la Comunidad Autónoma de Galicia (ver Figura 2.1). 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.

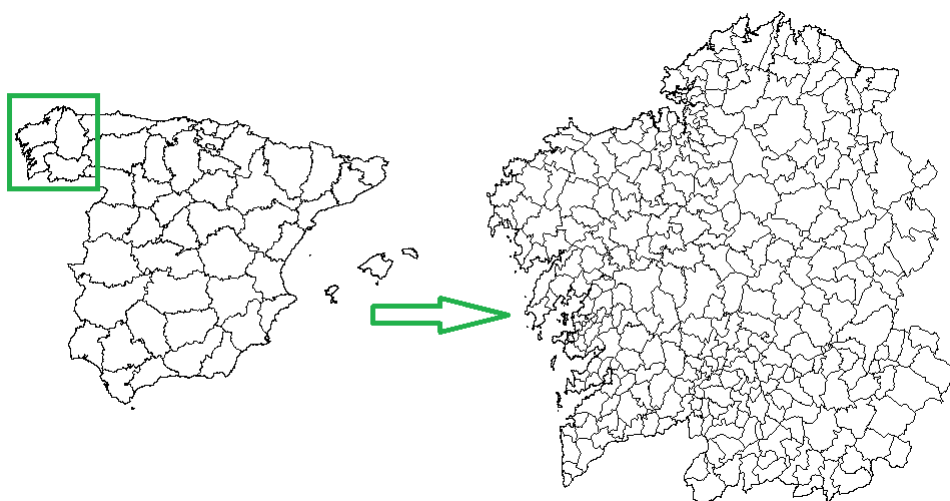


Figura 2.1: Mapa de España - Área de estudio por concellos

2.2. Origen y descripción de los datos utilizados

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. El apéndice A es una descripción estadística y espacial detallada de las 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, de los cuales 1553 han sido incendios intencionados, 222 por causas desconocidas, 198 por reproducción, 58 debido a otras causas, 9 por líneas eléctricas, 5 de fumadores, 5 por hogueras, 3 de escapes de vertederos, 3 por quemas agrícolas, 3 de motores y máquinas y 1 debido a quema de matorral.

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 2.2 aporta una descripción de las variables respuesta (las dos primeras), 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)
numero_incendios_provocados	Ocurrencia de incendio forestal provocado con categorías “no ocurrencia”(=0) y “ocurrencia” (=1)
altitud	Altitud media de cada cuadrícula en m.
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

Cuadro 2.1: Variables generadas a partir del material de partida.

La Figura 2.2 representa la distribución espacial de la variable respuesta `numero_incendios = 1`. Se observa una mayor concentración de incendios en las provincias de A Coruña y Pontevedra, en especial en la zona de las rías baixas y en la zona sur de la provincia de Ourense.

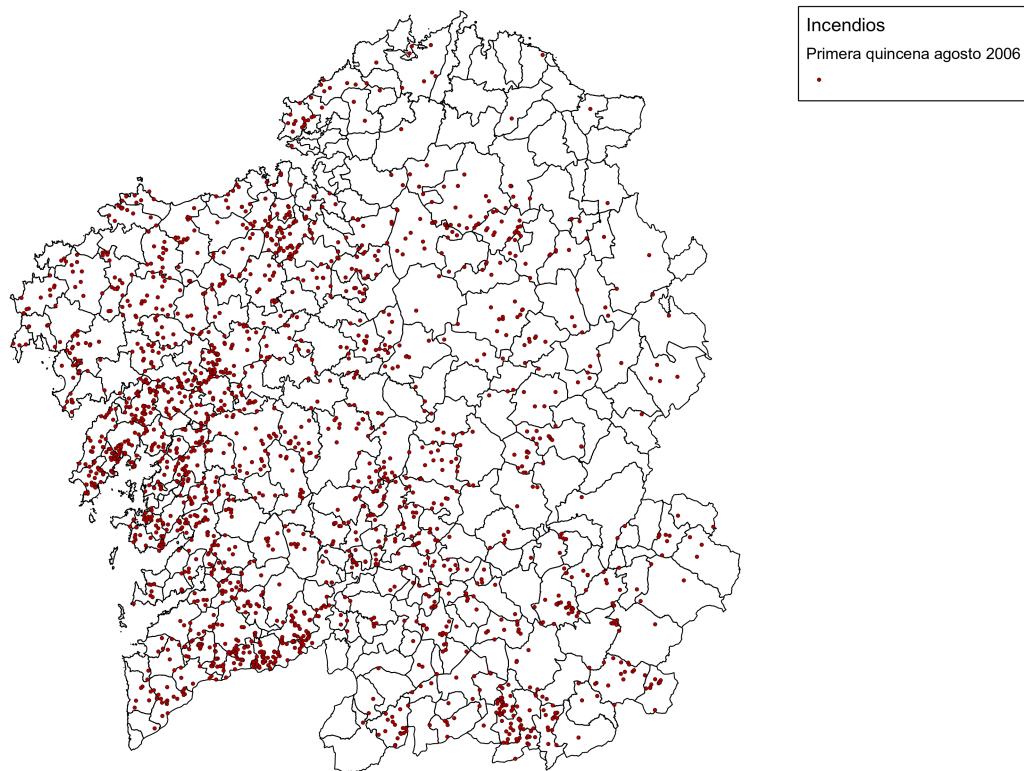


Figura 2.2: Incendios forestales registrados durante la primera quincena de agosto de 2006

La Figura 2.3 representa la distribución espacial de la variable respuesta `numero_incendios_provocados` \bar{I} con un total de 1553. La distribución espacial sigue un patrón espacial similar a la variable anterior.

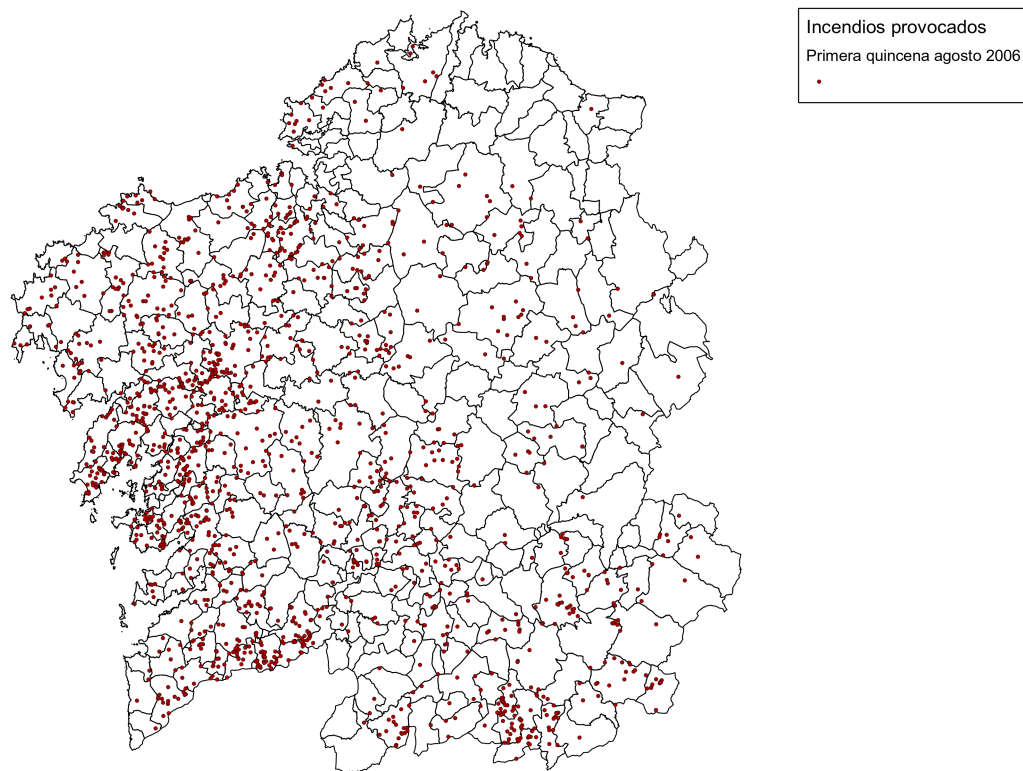


Figura 2.3: Incendios forestales provocados que fueron registrados en la primera quincena de agosto de 2006

2.2.1. Mapa de concellos de Galicia

Ha sido necesario incorporar un mapa de Galicia dividido en sus correspondientes concellos (Instituto Geográfico Nacional). Consta de un total de 315 concellos, de ellos, 94 pertenecen a la provincia de A Coruña, 67 a la de Lugo, 92 a la de Orense, y 62 municipios están en Pontevedra. La Figura 2.7 muestra el mapa de localización del área de estudio dividido por concellos.

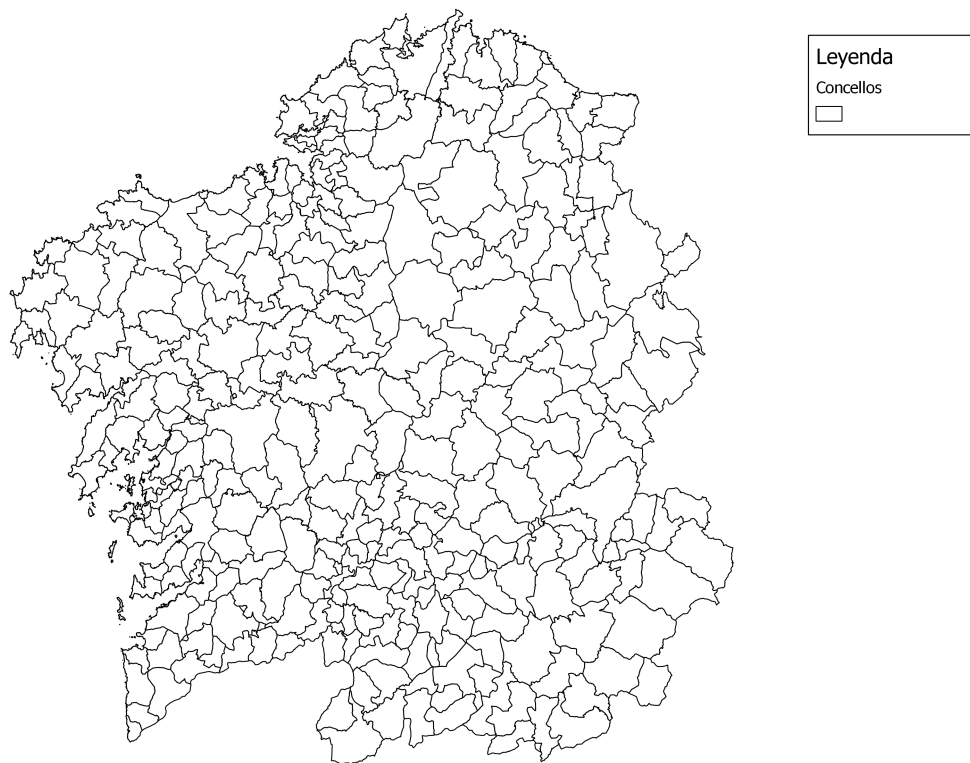


Figura 2.4: Mapa de Galicia dividido por concellos

Este mapa permitirá interpretar el efecto de la correlación espacial de los concellos en relación a la ocurrencia de incendios.

2.2.2. Base de Datos Climatológicos

Meteogalicia cuenta con un total de 134 estaciones meteorológicas. Para el periodo de análisis de la primera quincena de agosto de 2006, solamente 71 de ellas se encontraban operativas. Las variables obtenidas fueron:

- Precipitaciones (l/m^2)
- Humedad relativa media (%)
- Temperatura media del aire ($^{\circ}C$)

En la Figura 2.5 se muestran las estaciones meteorológicas utilizadas en el análisis.

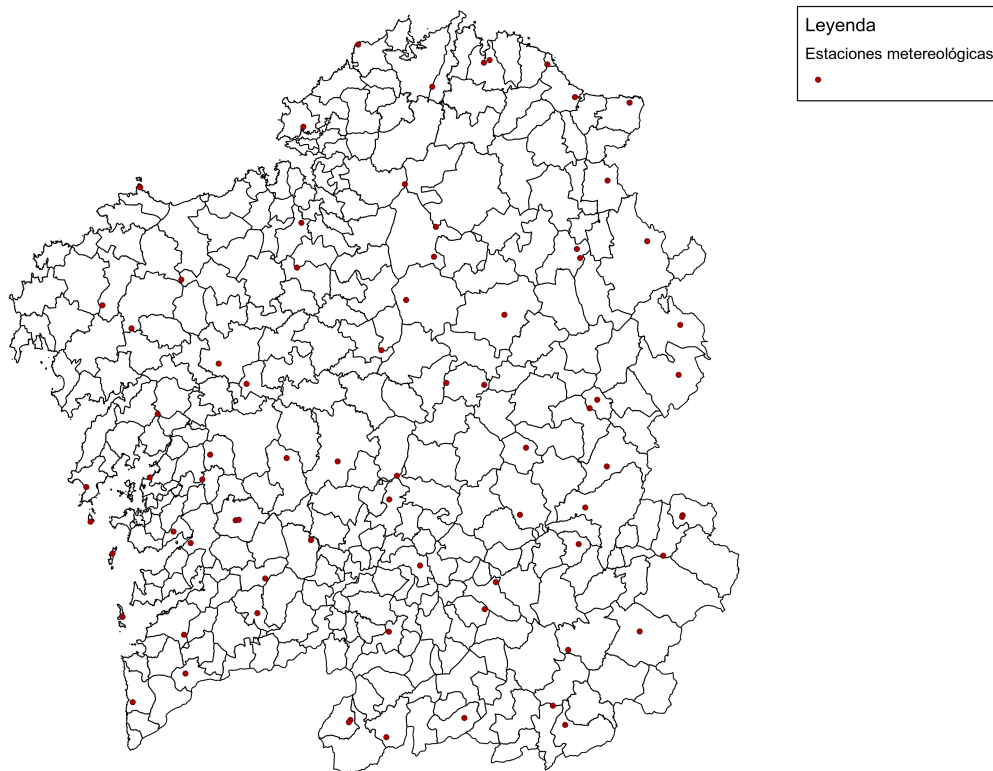


Figura 2.5: Estaciones climatológicas de Galicia

2.2.3. Modelo Digital del Terreno (MDT)

El MDT es una estructura numérica de datos que representa la distribución espacial de una variable cuantitativa y continua (Felicísimo, 1994), en nuestro caso se trata de una elevación del terreno.

MDT200

El MDT200 se ha obtenido por interpolación de modelos digitales del terreno de 5 m de paso de malla procedentes del Plan Nacional de Ortofotografía Aérea (PNOA). El modelo digital del terreno con paso de malla de 200 m, se distribuye por provincias (rectángulo envolvente de cada provincia). Sistema geodésico de referencia ETRS89. Proyección UTM en el huso correspondiente a cada provincia. En la Figura 2.6 se muestra el resultado.

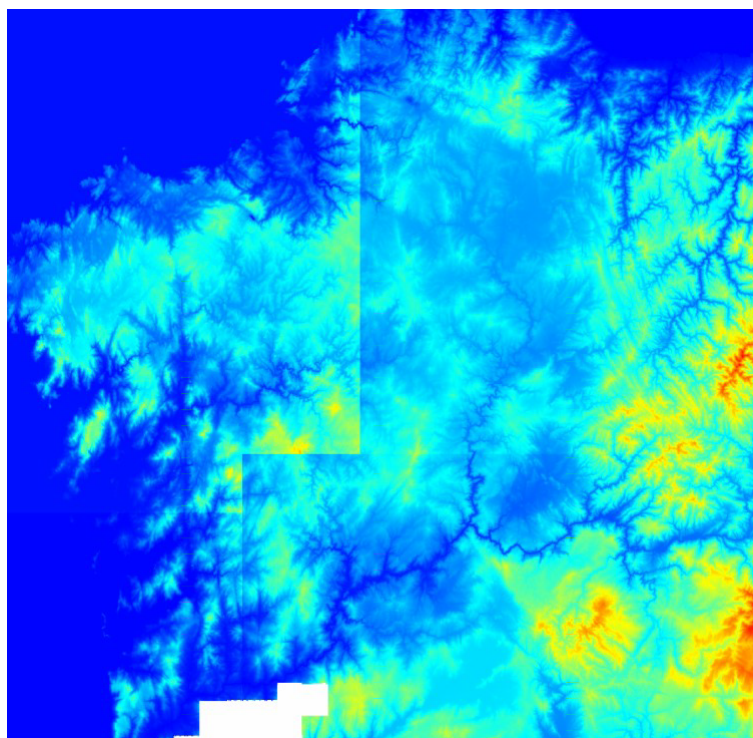


Figura 2.6: MDT200 de Galicia

2.2.4. Cuarto Inventario Forestal Nacional (IFN-4)

La serie de inventarios forestales nacionales (IFN) comienzan en el año 1966 y se encuentra en la actualidad realizándose el cuarto ciclo (MAGRAMA 2013). En el año 1966 se pone en marcha con carácter previo al IFN la realización del mapa forestal de España (MFE). Esta cartografía tenía como objetivo principal servir de base para el cálculo de existencias de madera de los bosques y consistía en la teselación (tesela, superficie de terreno con el mismo uso) de todo el territorio nacional. Este objetivo inicial se ha ido desarrollando y en el tercer inventario forestal (1997) se incorporaron nuevos campos a la base de datos asociada a la cartografía, entre ellos, el mapa de modelos de combustibles de Rothermel, 1983. En el año 2008 se inicia el cuarto ciclo, y al igual que los anteriores se comienza por la Comunidad Autónoma de Galicia, por lo que aunque los datos no están publicados, ya se dispone de ellos por parte de los grupos de investigación de la universidad.

Los modelos de combustible de Rothermel, son una clasificación de los diferentes tipos de vegetación y su comportamiento en el fuego. La combustibilidad de un sistema forestal se define como la capacidad del mismo para arder, desprendiendo la energía suficiente para consumirse y provocar la inflamación de la vegetación vecina, propagando de esta forma el fuego. A cada tipo de vegetación, le corresponde una

inflamabilidad y combustibilidad determinadas, que varían en función del tipo y cantidad de biomasa, y su distribución espacial o estratificación (Galiana et al. 2009). La clasificación de modelos de combustible considera 13 tipos de modelos de combustible, divididos en 4 grandes grupos en función de cuál sea el principal medio de propagación del fuego: pasto, matorral, hojarasca bajo arbolado y restos de corta o de tratamientos selvícolas. En este trabajo se ha incluido un quinto grupo que es un grupo denominado sin asignar modelo. Se ha hecho uso del IFN-4 para conocer el tipo de combustible existente en cada punto del territorio gallego.

Un modelo útil de riesgo para la planificación forestal debe tener en cuenta la estructura y composición de las especies (Díaz Delgado et al., 2004) que son obtenidas en los inventarios forestales (González, et al., 2007). Los modelos, basados en datos empíricos, deben ser sensibles a los efectos de tratamientos selvícolas y tipos de gestión para establecer el riesgo de incendio en una determinada situación. Los modelos existentes no parecen cumplir todos estos requisitos. En algunos modelos el riesgo de incendio depende de la edad (Reed, 1994; Moritz, 2003), y otros al igual que la clasificación de riesgo de incendio siguiendo el modelo utilizado por Thompson et al. (2000), se basan en evaluaciones cualitativas de expertos (González, et al. 2007) o en encuestas a ciudadanos (Tábara et al., 2003; Riera & Mogas, 2004). También resulta interesante conocer que vegetación resulta de una serie continuada de fuegos y si esta puede estar formando parte del ciclo natural del mismo (Viedma, et al. 2006).

A continuación se muestra la clasificación y descripción de los diferentes tipos de modelos de combustible.

1. Sin Asignar Modelo

- Modelo de combustible 0: no se dispone de información, se recogerá la información en visita a campo.

2. Pastos

- Modelo de Combustible 1:
 - Pasto fino, seco y bajo, que recubre completamente el suelo.
 - Pueden aparecer algunas plantas leñosas dispersas ocupando menos de 1/3 de la superficie.
 - Cantidad de combustible (materia seca): 1-2 t/ha.
- Modelo de Combustible 2:
 - Pasto fino, seco y bajo, que recubre completamente el suelo.
 - Las plantas leñosas dispersas cubren de 1/3 a 2/3 de la superficie, pero la propagación del fuego se realiza por el pasto.

- Cantidad de combustible (materia seca): 5-10 t/ha.

- Modelo de Combustible 3:

- Pasto grueso, denso, seco y alto (>1m).
- Puede haber algunas plantas leñosas dispersas.
- Cantidad de combustible (materia seca): 4-6 t/ha.

3. Matorral

- Modelo de Combustible 4:

- Matorral o plantación joven muy densa, de más de 2 m de altura; con ramas muertas en su interior
- Propagación del fuego por la copas de las plantas.
- Cantidad de combustible (materia seca): 25-35 t/ha.

- Modelo de Combustible 5:

- Matorral denso y verde, de menos de 1 m de altura.
- Propagación del fuego por hojarasca y el pasto.
- Cantidad de combustible (materia seca): 5-8 t/ha.

- Modelo de Combustible 6:

- Parecido al modelo 5, pero con especies más inflamables o con restos de corta y plantas de mayor talla.
- Propagación del viento con vientos moderados a fuertes.
- Cantidad de combustible (materia seca): 10-15 t/ha.

- Modelo de Combustible 7:

- Matorral de especies muy inflamables; de 0.5 a 2 m de altura, situado como sotobosque en masas de coníferas.
- Cantidad de combustible (materia seca): 10-12 t/ha.

4. Hojarasca bajo arbolado

- Modelo de Combustible 8:

- Bosque denso, sin matorral
- Propagación del fuego por la hojarasca muy compacta.

- Cantidad de combustible (materia seca): 10-12 t/ha.

■ Modelo de Combustible 9:

- Parecido al modelo 8, pero con hojarasca menos compacta, formada por acículas largas y rígidas o follaje de frondosas de hojas grandes.

- Cantidad de combustible (materia seca): 7-9 t/ha.

■ Modelo de Combustible 10:

- Bosque con gran cantidad de leña y árboles caídos, como consecuencia de vendavales, plagas intensas, etc.

- Cantidad de combustible (materia seca): 30-35 t/ha.

5. Restos de cortas y operaciones selvícolas

■ Modelo de Combustible 12:

- Predominio de los restos sobre arbolado

- Restos de poda o aclareo cubriendo todo el suelo.

- Cantidad de combustible (materia seca): 50-80 t/ha.

■ Modelo de Combustible 13:

- Grandes acumulaciones de restos gruesos y pesados, cubriendo todo el suelo.

- Cantidad de combustible (materia seca): 100-150 t/ha.

Capítulo 3

Metodología estadística

Este capítulo describe los modelos de regresión basados en un predictor aditivo estructurado (modelos STAR). Estos modelos de regresión son muy generales y pueden hacer frente a diferentes tipos de variables dependientes y también con diferentes tipos de covariables. En la primera sección 3.1 de este capítulo, se da una breve introducción en modelos de regresión incluyendo la generalización a modelos STAR. La sección 3.2 trata los efectos de las distintas covariables y la sección 3.3 aborda los distintos criterios de selección de los modelos STAR.

3.1. Modelo de Regresión Aditivo Estructurado (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:

$$E(y|x_1, \dots, x_q) = \gamma_0 + \gamma_1 x_1 + \dots + \gamma_q x_q = \gamma' x =: \eta \quad (3.1)$$

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 3.1 es lineal en los coeficientes de regresión y la relación entre covariables y los valores esperados también son lineales.

Si la variable respuesta ya 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:

$$E(y|x_1, \dots, x_q) = h(\gamma_0 + \gamma_1 x_1 + \dots + \gamma_q x_q) = h(\gamma' x) = h(\eta) \quad (3.2)$$

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

$$\eta := \gamma_0 + \gamma_1 x_1 + \dots + \gamma_q x_q = \gamma' x \quad (3.3)$$

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

$$\eta := \gamma_0 + f_1(x_1) + \dots + f_q(x_q) + \gamma_1 u_1 + \dots + \gamma_p u_p = f_1(x_1) + \dots + f_q(x_q) + \gamma' u \quad (3.4)$$

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 endite-
mize

El predictor aditivo estructurado (3.4) supera las dificultades sustituyendo 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 3.4, 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 aditivo estructurado 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)), modelos de coeficientes diferentes (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

Como ya mencionamos en la sección anterior, vamos a tratar con diferentes tipos de variables en el contexto de los modelos STAR. 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,

$$\mathbf{f} = \mathbf{X}\boldsymbol{\beta} \quad (3.5)$$

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

$$p(\boldsymbol{\beta}) \propto \exp\left(\frac{-1}{2\tau^2}\boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta}\right) \quad (3.6)$$

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

$$p(\boldsymbol{\beta}|y) \propto L(y|\boldsymbol{\beta}) \cdot p(\boldsymbol{\beta}) \quad (3.7)$$

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

$$\log(p(\boldsymbol{\beta}|y)) \propto l(y|\boldsymbol{\beta}) + \log(p(\boldsymbol{\beta})) \propto l(y|\boldsymbol{\beta}) - \frac{1}{2\tau^2}\boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta} \quad (3.8)$$

La Fórmula 3.8 es equivalente al logaritmo de verosimilitud penalizado 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

$$l_{pen}(y|\boldsymbol{\beta}) = \phi \cdot l(y|\boldsymbol{\beta}) - \frac{1}{2}\lambda\boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta} \quad (3.9)$$

- Todas las distribuciones a priori (3.6) 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.

3.2.1. Efectos Lineales

Como se mencionó en la sección anterior, 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 $\gamma = (\gamma_0, \dots, \gamma_q)'$ 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.

3.2.2. 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

$$u_j = \begin{cases} 1, & \text{si } u = j \\ 0, & \text{en otro caso} \end{cases} \quad (3.10)$$

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

$$u_j = \begin{cases} 1, & \text{si } u = j \\ -1, & \text{si } u = k \\ 0, & \text{en otro caso} \end{cases} \quad (3.11)$$

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 .

3.2.3. 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.

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 $\mathbf{f} = \mathbf{X}\boldsymbol{\beta}$.

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:

$$l_{pen}(y|\beta_1, \dots, \beta_p) = \phi \cdot l(y|\beta_1, \dots, \beta_p) - \frac{1}{2} \cdot penalty(\lambda), \quad (3.12)$$

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,

$$penalty(\lambda) = \lambda \cdot \sum_{j=k+1}^p (\Delta^k \beta_j)^2 = \lambda \cdot \beta' \mathbf{P} \beta, \quad (3.13)$$

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:

$$\Delta^1 \beta_j = \beta_j - \beta_{j-1} \quad y \quad \Delta^2 \beta_j = \beta_j - 2\beta_{j-1} + \beta_{j-2} \quad (3.14)$$

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 $\mathbf{D}_k\boldsymbol{\beta}$. 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 la 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).

3.2.4. Escala de Tiempo

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

$$f_{tiempo}(t) = f_{tend. suave}(t) + f_{comp. estacional}(t) \quad (3.15)$$

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)).

3.2.5. 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.

$$f_{\text{espacial}}(s) = f_{\text{estructurado}}(s) + f_{\text{no estructurado}}(s) \quad (3.16)$$

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 $f_{\text{espacial}}(s) = \beta_s$ toma la forma

$$\beta_s | \beta_{s'}, s' \neq s \sim N \left(\frac{1}{N_s} \sum_{s' \in \delta_s} \beta_{s'}, \frac{\tau^2}{N_s} \right) \quad (3.17)$$

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,

$$\beta_s | \beta_{s'}, s' \neq s \sim N \left(\sum_{s' \in \delta_s} \frac{w'_{ss}}{w_{s+}} \beta_{s'}, \frac{\tau^2}{w_{s+}} \right) \quad (3.18)$$

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 3.17 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.

3.2.6. 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,

$$\beta_i \sim N(0, \tau^2) \quad (3.19)$$

Aquí la distribución conjunta para β es una distribución normal adecuada. Sin embargo, se puede escribir en la misma forma general 3.6 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.

3.2.7. Coeficientes de Variación

En los apartados anteriores se han introducido diversos métodos de modelización de diferentes tipos de efectos de una dimensión. Ahora vamos a describir extensiones que nos permiten modelar las interacciones de dos dimensiones. Distintos coeficientes fueron primero popularizado por Hastie y Tibshirani (1993) en el contexto de splines suavizados. Aquí, la pendiente de una variable z varía suavemente sobre el alcance de otra variable v definiendo el término,

$$f(v, z) = g(v)z \quad (3.20)$$

A menudo, la variable de interacción z es categórica, pero puede ser continua también. El efecto de modificar v puede ser una variable continua, una localización espacial o un grupo indicador. El vector de la función de evaluaciones \mathbf{f} puede ser escrito como una combinación lineal.

$$f = \mathbf{X}\boldsymbol{\beta}, \quad (3.21)$$

Usando una matriz de diseño \mathbf{X} y un vector de coeficientes $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$. La matriz de diseño para $f(v, z)$ se basa tanto en las observaciones de z y de v y se calcula como

$$\mathbf{X} = \text{diag}(z_1, \dots, z_n)\mathbf{V}, \quad (3.22)$$

donde \mathbf{V} es la matriz de diseño correspondiente a $g(v)$.

Algunos casos especiales de diferentes coeficientes a veces aparecen bajo un nombre diferente: si el efecto de modificar v es un indicador de grupo, la función $f(v, z)$ de dos dimensiones (v, z) se llama la pendiente aleatoria. Los modelos que incluyen un coeficiente diferente con un efecto espacial como función modificada se conoce como regresión geográficamente ponderada (ver Fotheringham, Brunson y Charlton (2002)). Los modelos dinámicos se basan en coeficientes de variaciones de tiempo (véase Fahrmeir y Tutz (2001)).

Por último, echamos un vistazo en el límite de $f(v, z)$ para $\lambda \rightarrow \infty$ o equivalente $\tau^2 \rightarrow 0$. Esto depende de la distribución a priori impuesta a la función univariante $g(v)$. Las funciones de límite $g^{(\infty)}(v)z$ se describen en las secciones respectivas de todas las funciones univariantes g . La función de límite del coeficiente de variación es

$$f^{(\infty)}(v, z) = g^{(\infty)}(v)z \quad (3.23)$$

Esto significa que, $f^{(\infty)}(v, z)$ es igual a cero si $g(v)$ es un efecto aleatorio de z para un paseo aleatorio a priori de primer orden (MRF o P-spline de primer orden). Para un paseo aleatorio a priori de segundo orden se obtiene una interacción de la forma $f^{(\infty)}(v, z) = c_1 \cdot z + c_2 \cdot v \cdot z$.

3.3. Criterios de Selección del modelo

La selección de variables y parámetros de suavizado se basa en los criterios de selección. Hay una amplia variedad de criterios de selección disponibles. En este trabajo se restringe a algunos de los criterios más utilizados que pueden ser usados en combinación con nuestros algoritmos de selección. Una descripción detallada de este tema se puede encontrar en Miller (2002), por ejemplo.

3.3.1. Criterio de Información de Akaike (AIC)

El Criterio de Información de Akaike o AIC fue introducido originalmente por Akaike (1973). Burnham y Anderson (1998) o Cavanaugh (1997) derivaron el AIC desde la distancia de Kullback-Leibler

$$I(f, g) = \int f(y) \ln \left(\frac{f(y)}{g(y|\theta)} \right) dy = \int f(y) \ln(f(y)) dy - \int f(y) \ln(g(y|\theta)) dy, \quad (3.24)$$

A menudo, el modelo g presenta una familia de modelos que dependen del parámetro θ . Cuanto menor sea el valor de $I(f, g)$ mejor es el modelo g asumido. El AIC es una estimación de la expectativa del segundo término, multiplicado por dos. Por lo tanto, el AIC no tiene cero natural. Eso significa que, el AIC puede ser utilizado para comparar los modelos, pero no da evidencia de la calidad real de un modelo. La fórmula para AIC es

$$AIC = -2 \cdot l(\theta|y) + 2 \cdot p, \quad (3.25)$$

donde $l(\theta|y) = \ln(g(y|\theta))$ es el logaritmo de verosimilitud del modelo y p es el número de parámetros estimados en θ . Para la selección en modelos de regresión lineal, el vector θ incluye todos los coeficientes de regresión y, posiblemente, un parámetro de escala (en función del tipo de distribución de la variable respuesta).

En fórmula 3.25 los modelos g asumidos son funciones de verosimilitud. En los modelos aditivos estructurados realizamos inferencia de máxima verosimilitud penalizada, por lo que los modelos g asumidos ahora son probabilidades penalizadas. En la forma general el AIC tiene la fórmula (comparar Hastie y Tibshirani (1990))

$$AIC = -2l(\theta|y) + 2tr(\mathbf{H}) = -2l(\theta|y) + 2df_{total}, \quad (3.26)$$

donde la matriz \mathbf{H} es la matriz que proyecta la información y de los valores ajustados, es decir, $\hat{y} = \mathbf{H}y$. En el caso de respuesta no gaussiana, \mathbf{H} es la matriz de evaluación de la última iteración del algoritmo de puntuación, es decir, $\hat{\eta} = \mathbf{H}\bar{y}$. Además, nos referimos a la cantidad $df_{total} := tr(\mathbf{H})$ como los grados de libertad del modelo. En inferencia de máxima verosimilitud la cantidad $tr(\mathbf{H})$ es igual al número de parámetros de regresión.

3.3.2. Criterio de Información Bayesiano (BIC)

El Criterio de Información Bayesiano (BIC) se deriva de Schwarz (1978). El BIC se origina a partir de un contexto bayesiano. Supongamos que tenemos dos modelos diferentes $M_i, i = 1, 2$, los cuales asumen unas probabilidades a priori $p(M_1)$ y $p(M_2)$. Los coeficientes de regresión a priori son definidos condicionalmente en el modelo por $p(\theta_i|M_i)$. Con el teorema de Bayes se obtiene la probabilidad a posteriori para cada modelo por

$$p(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y|M_1)p(M_1) + p(y|M_2)p(M_2)}, \quad (3.27)$$

donde la expresión $p(y|M_i)$ es la verosimilitud marginal del modelo y se calcula como

$$p(M_i|y) = \int p(y|\theta_i, M_i)p(\theta_i|M_i)d\theta_i. \quad (3.28)$$

El término $p(y|\theta_i, M_i)$ es la función de verosimilitud para los parámetros θ_i . Para responder a la cuestión de cuál de los dos modelos es superior, se puede utilizar el factor de Bayes (ver Kass y Raftery (1995))

$$B_{12} = \frac{p(y|M_1)}{p(y|M_2)}, \quad (3.29)$$

El cual apoya que M_1 si $B_{12} > 1$. El BIC es una aproximación al factor de Bayes y permite evitar la especificación de probabilidades a priori. En algunos lugares, la selección del modelo de BIC es incluso igual a la selección basada en los factores de Bayes (ver Kass y Raftery (1995) para más detalles). La fórmula de BIC es

$$BIC = -2l(\theta|y) + \log(n) \cdot p, \quad (3.30)$$

donde p es el número de parámetros y n el número de observaciones.

3.3.3. Validación Cruzada Genaralizada (GCV)

Supongamos que tenemos un modelo con una variable respuesta con distribución normal y que solo contiene una función no lineal, es decir,

$$y_i = f(x_i) + \epsilon_i, \quad (3.31)$$

donde $\epsilon_i \sim N(0, \sigma^2)$. En este caso, la proyección de la matriz \mathbf{H} y los datos y sobre los valores ajustados vienen dados por

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{P})^{-1}\mathbf{X}' \quad (3.32)$$

donde \mathbf{X} es el diseño respectivo y \mathbf{P} es la matriz penalizada. Con el fin de determinar un valor adecuado para el parámetro de suavizado λ se puede utilizar la validación cruzada dejando fuera una observación a la vez. Es decir, el criterio

$$CV = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_\lambda^{-i}(x_i))^2, \quad (3.33)$$

reduce al mínimo λ , donde $\hat{f}_\lambda^{-i}(x_i)$ se estimó sin observación (y_i, x_i) . La función $\hat{f}_\lambda^{-i}(x_i)$ puede calcularse directamente a partir de la matriz \mathbf{H} sin estimar los diferentes modelos n a través de

$$\hat{f}_\lambda^{-i}(x_i) = \sum_{j=1, j \neq i}^n \frac{\mathbf{H}_{ij}}{1 - \mathbf{H}_{ii}} y_j. \quad (3.34)$$

Usando esta relación, la fórmula 3.33 se puede escribir como

$$CV = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{f}_\lambda^{-i}(x_i)}{1 - \mathbf{H}_{ii}} \right)^2. \quad (3.35)$$

Mediante la sustitución de los elementos diagonales \mathbf{H}_{ii} por su valor medio $tr(\mathbf{H})/n$ se obtiene la validación cruzada generalizada

$$GCV = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{f}_\lambda(x_i)}{1 - tr(\mathbf{H})/n} \right)^2. \quad (3.36)$$

En un modelo aditivo estructurado con varios términos y con posiblemente errores heterocedásticos la fórmula más general es

$$GCV = \frac{1}{n} \sum_{i=1}^n w_i \left(\frac{y_i - \hat{\eta}_i}{1 - tr(\mathbf{H})/n} \right)^2, \quad (3.37)$$

donde \mathbf{H} representa de nuevo la matriz hat para todo el modelo y, además, incluye la matriz de pesos \mathbf{W} . Para el caso de una variable respuesta que siga un distribución no normal, el criterio GCV se puede adaptar mediante el uso de la suma de cuadrados residual basado en el último paso del algoritmo de puntuación (ver Wood (2006a)):

$$GCV = \frac{1}{n} \sum_{i=1}^n d_i \left(\frac{(\bar{y}_i - \hat{\eta}_i)^2}{1 - tr(\mathbf{H})/n} \right)^2, \quad (3.38)$$

Capítulo 4

Explorando la incidencia de incendios forestales totales y provocados

La probabilidad de incendio depende de dos factores: 1) Las condiciones climáticas (Andrews, 1986; Van Wagner, 1987, Dickson, et al. (2006) y 2) El tipo y estructura de la vegetación (Mermoz, et al. 2005) que en algunos casos favorece y en otros casos parece frenar su avance (Nunes, et al. 2005) y que en zonas consideradas de referencia están totalmente relacionadas con las condiciones socioeconómicas (Chas, 2007; Marey-Pérez, et al. 2006; Marey- Pérez y Rodríguez-Vicente, 2008). Las condiciones meteorológicas futuras no se conocen con exactitud, lo que invalida los modelos basados en factores meteorológicos en la planificación forestal, pero si pueden ser utilizados para la simulación de previsibles escenarios. Por otra parte, la gran cantidad de datos de valores medios de las diferentes variables climáticas son conocidos, lo que significa que puede utilizarse en los modelos predictores. La existencia de correlaciones de estas variables climáticas con parámetros de sitio tales como longitud, latitud, altitud, pendiente y orientación implica que estos parámetros pueden mejorar las predicciones en la elaboración de modelos.

En este capítulo se presentan aplicaciones de los modelos de regresión STAR de respuesta binomial. El objetivo principal es evaluar factores de riesgo naturales que serán incluidos como covariables en los modelos de regresión, los cuales son analizados con más detalle a lo largo de este capítulo. Para estimar dichos modelos, se empleó la función bayesx del paquete estadístico R2BayesX para R (R Development Core Team 2007).

4.1. Modelización a partir de variables meteorológicas

Como punto de partida en el análisis de los datos, en esta sección se relaciona la ocurrencia de incendios forestales totales y provocados con las variables meteorológicas ta_media , hr y pp junto con el MDT200 que permite conocer la altitud en cada cuadrícula.

4.1.1. Corrección de los datos meteorológicos

Un primer paso para el modelado de la ocurrencia de incendios forestales a partir de variables meteorológicas, es conocer que estación mide las variables necesarias para su cálculo y que además sea la estación más cercana que midió esa variable en cada cuadrícula. Debido a la diferencia de altitud entre la estación más cercana a la cuadrícula y la cuadrícula, es necesario hacer una corrección de las variables de temperatura y precipitaciones con respecto a la diferencia de altimetría, tal y como se explica a continuación.

1. Corrección de los datos de temperatura media:

La variación de la temperatura con la altura es de $0,65\text{ }^{\circ}\text{C}/100\text{ m}$ hasta los 11000 m (Núñez y Colhoun 1986), es decir:

$$tH = 15\text{ }^{\circ}\text{C} - 0,0065\text{ }^{\circ}\text{C}/\text{m} \cdot H \quad (4.1)$$

donde tH es la temperatura del aire a una altura H y H es la diferencia de altitud entre dos puntos.

2. Corrección de los datos de precipitación:

Un método simple de corrección de la precipitación (P) en función de la altitud (h), del que es preciso saber que no contempla efectos de factores como la ubicación a barlovento o sotavento en montañas próximas al mar, sería el siguiente:

$$P_{\text{corregida}}(\text{mensual}) = P_{\text{mensual}} \pm 0,08 \quad (4.2)$$

Por lo tanto, el gradiente de precipitaciones es de un 8% de variación por cada 100 m de diferencia de cota entre la cuadrícula y el observatorio meteorológico.

El principal objetivo es investigar el posible efecto espacial del número de incendios e incendios provocados a partir de las variables meteorológicas y la altitud por cuadrícula asociada a un concello. A continuación se presentan los resultados obtenidos en la modelización.

4.1.2. Modelización de los incendios forestales totales

El primer análisis en este trabajo explora la evolución de los incendios totales de acuerdo a las variables meteorológicas, la altitud y el posible efecto espacial del concello. En este análisis se consideran un total de 362625 registros, incluyendo todos los incendios ocurridos en el periodo de estudio que han sido registrados en cada una de las cuadrículas en las que se dividió Galicia.

El modelo se representa por:

$$\begin{aligned}
 \text{logit}(P_k) &= \\
 &= \log\left(\frac{P_k}{1 - P_k}\right) = \\
 &= \beta_0 + f_1(\text{altitud}) + f_2(\text{ta_media}) + f_3(\text{hr}) + f_4(\text{pp}) + \\
 &+ f_{\text{estr}}(\text{cdconc}) + f_{\text{no_estr}}(\text{cdconc})
 \end{aligned} \tag{4.3}$$

Para simplificar la notación consideramos k como el índice del conjunto de vóxeles espacio-tiempo 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 4.1 y 4.2 ilustran los efectos suavizados de las covariables altitud, ta_media, hr y pp en la ocurrencia de incendio forestal para los modelos STAR con respuesta binomial. Además, la estimación de los grados de libertad del intercepto y las variables explicativas junto con los intervalos de credibilidad al 95 % se muestran en el Cuadro 4.1.

Variables	grados libertad	IC al 95 %
cte	1	(-9.2360, -7.9079)
altitud	2.502	Ver Figura 4.1
ta_media	4.069	Ver Figura 4.1
hr	5.271	Ver Figura 4.2
pp	1.053	Ver Figura 4.2

Cuadro 4.1: Estimación de los grados de libertad e intervalos de credibilidad al 80 y 95 %

Los grados de libertad de cada función suavizada dan una clara indicación de la necesidad de incorporar la flexibilidad a los modelos y además permite detectar linealidad (estimación de los grados de libertad = 1).

En la Figura 4.1, el efecto de la altitud (ver panel izquierdo) muestra un efecto decreciente moderado de los incendios hasta aproximadamente los 300 m de altitud, con un efecto decreciente más acusado a partir de esta altitud. La estimación de los grados de libertad de esta covariable es de 2.502 lo que nos confirma el descenso suave y decreciente a altitudes bajas junto con un descenso más acusado a medida que las altitudes aumentan. Se puede concluir que, la ocurrencia de incendios parece estar relacionada con altitudes bajas para el periodo de estudio considerado. Hay que señalar también que las bandas de credibilidad son más grandes a partir de los 1000 m de altitud aproximadamente, por lo que no debe sobreinterpretarse debido a la incertidumbre de estos resultados, ya que existen pocos datos en el área de estudio que muestren altitudes por encima de los 1000 metros.

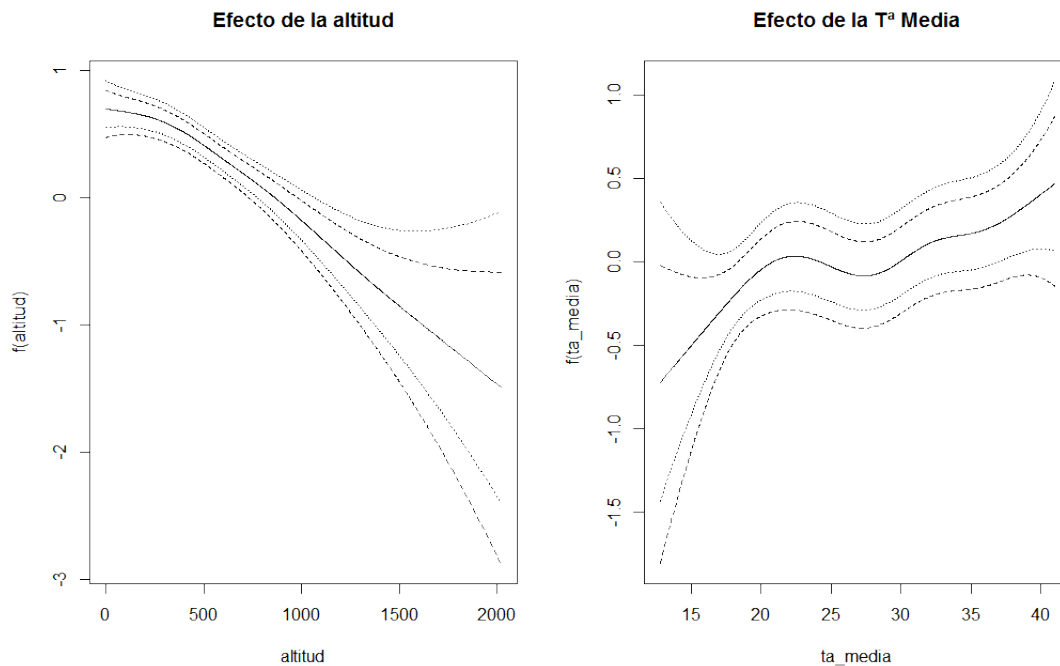


Figura 4.1: Estimación de los efectos suavizados de la altitud (panel izquierdo) y ta_media (panel derecho) con sus correspondientes bandas de credibilidad punteadas al 80 y 95 % - Ocurrencia de incendios forestales totales

La covariable ta_media (temperatura media diaria) parece tener un efecto creciente acusado en la ocurrencia de incendios forestales hasta cierto nivel (alrededor de 20°C), tal y como se muestra en el panel derecho de la Figura 4.1. A partir de esta temperatura, la ocurrencia de incendios para temperaturas comprendidas en-

entre 20 y 30°C aproximadamente se estabiliza, mostrando a partir de los 30°C, un efecto creciente no tan drástico como a temperaturas bajas ($< 20^\circ\text{C}$). Los grados de libertad de la covariable ta_{media} toman el valor de 4.069 (ver cuadro 4.1).

Los efectos de las covariables hr y pp se muestra en la figura 4.2 en donde se observa que los efectos de la hr (panel izquierdo) son crecientes para valores de humedad relativa bajos hasta el 35% aproximadamente, a partir de ahí el efecto de la humedad relativa es decreciente siendo más acusado para humedades superiores al 75 - 80%. Por lo tanto, la ocurrencia de incendios se asocia a porcentajes bajos de humedad relativa. Esto se debe a que la mayoría de los incendios se sitúan en zonas costeras donde las humedades relativas por debajo del 50% no suelen ser comunes. A partir de este punto, unos incrementos en la humedad relativa se asocian a un descenso del número de incendios.

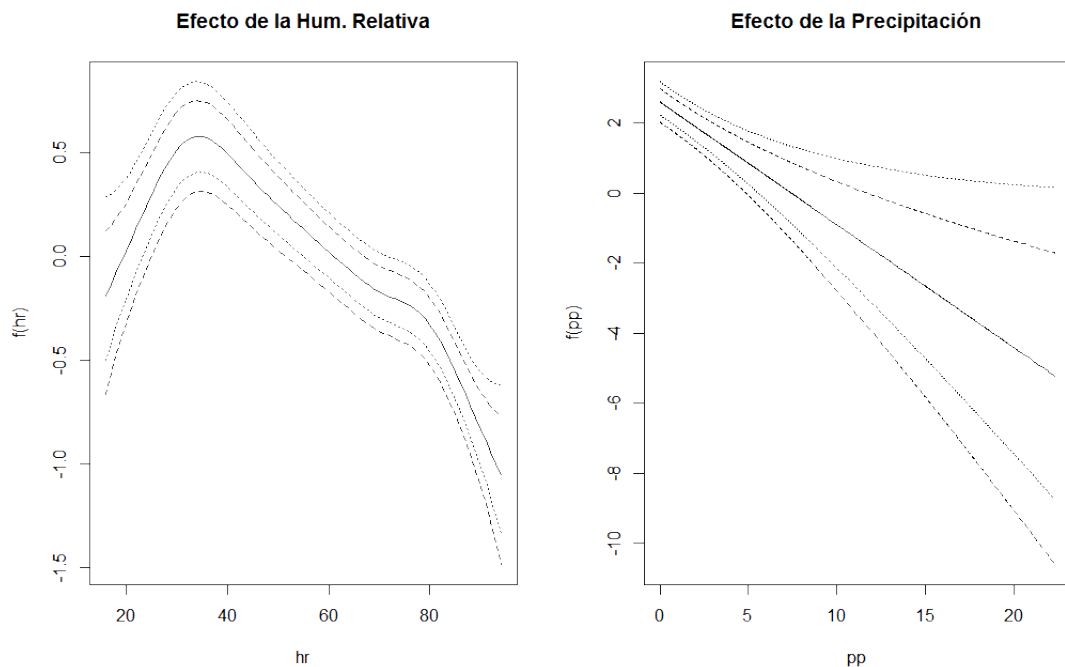


Figura 4.2: Estimación de los efectos suavizados de la hr (panel izquierdo) y pp (panel derecho) con sus correspondientes bandas de credibilidad al 80 y 95% - Ocurrencia de incendios forestales totales

El efecto de la precipitación (panel derecho) tiene un efecto prácticamente lineal y decreciente, lógicamente, 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 el Cuadro 4.1.

Los efectos espaciales se dividen en efectos estructurados que presentan una fuerte estructura espacial y efectos no estructurados que son los que están presentes a nivel local. El mapa con los efectos espaciales estructurados (ver Figura 4.3, panel izquierdo) muestra concellos que tienen claramente menor ocurrencia de incendios y aparecen principalmente en el Norte y Noreste de Galicia (en verde) frente a una gran ocurrencia de incendios (en rojo) y alberga prácticamente toda la costa Oeste de Galicia excepto los concellos de A Guardia y O Rosal, Suroeste de Galicia. Todo ello de acuerdo a los datos del periodo analizado.

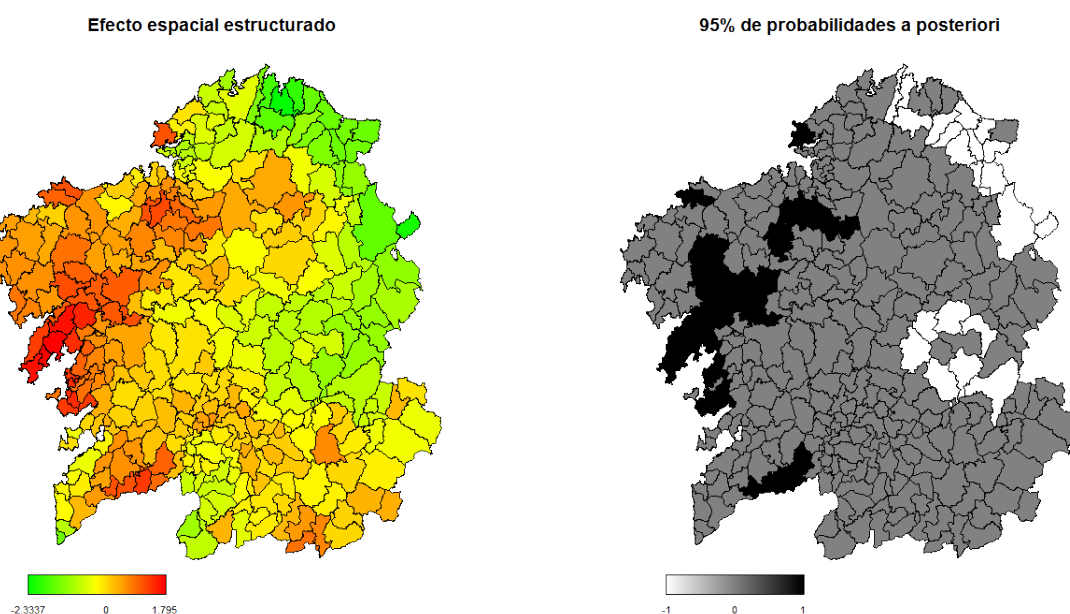


Figura 4.3: Efectos espaciales estructurados - Ocurrencia de incendios forestales totales

Las probabilidades a posteriori al 95 % de credibilidad se representa en la Figura 4.3 (panel derecho). Existen concellos con intervalos de credibilidad estrictamente positivos (regiones en negro) o negativos (regiones en blanco). Los concellos dibujados en gris son concellos no significativos de la ocurrencia de incendios forestales. Como se observa, existe una gran zona de indefinición, lo que nos sugiere la necesidad de mejorar el modelo.

Los resultados de los efectos espaciales no estructurados para los modelos STAR (ver Figura 4.4) no apuntan en una dirección determinada. Los concellos coloreados en rosa apuntan a una mayor posibilidad de efecto local frente a los concellos coloreados en verde que representan una menor posibilidad de efecto local. Todos

los concellos en gris (ver Figura 4.4, panel derecho) son concellos no significativos de la ocurrencia de incendio forestal a escala local. Los concellos blancos rallados no registran dato y solo un concello aparece como significativo que es el concello de Chandrexa de Queixa, lo que obliga a un análisis más detallado de cual ha sido la causa de los incendios de este municipio que los diferencia de los de su entorno y del resto de Galicia.

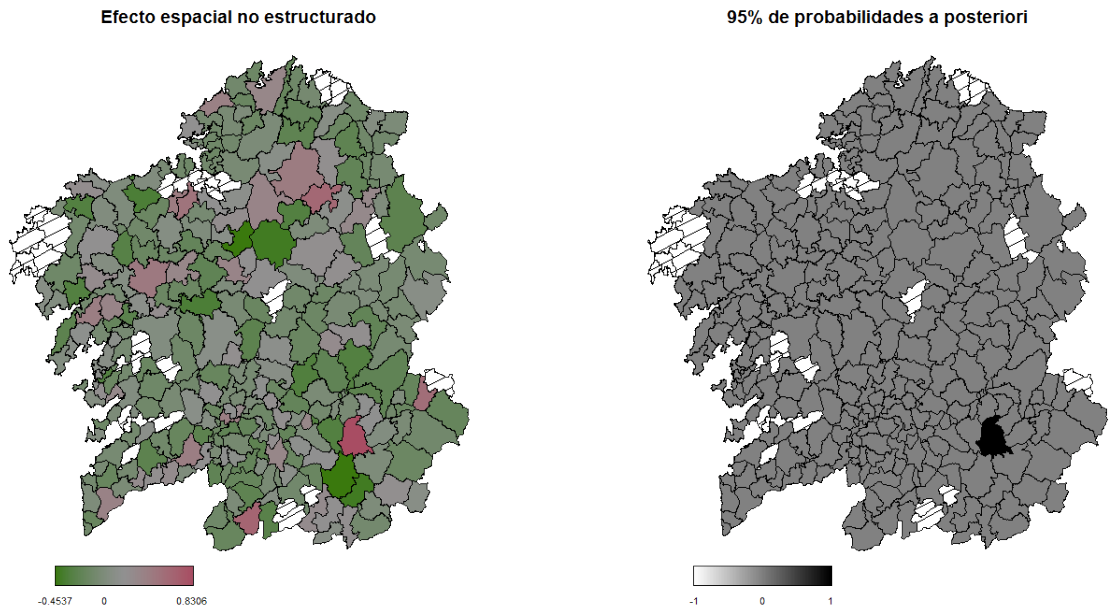


Figura 4.4: Efectos espaciales no estructurados - Ocurrencia de incendios forestales totales

4.1.3. Modelización de los incendios forestales provocados

Como ya se comentó anteriormente, los incendios forestales provocados en Galicia suponen más del 80 % de los incendios forestales totales, por esa razón, ha sido necesario tenerlos en consideración y conocer su evolución en el área de estudio. Para esta nueva modelización se usan los mismos registros que en la subsección anterior (362625 registros) pero ahora se tiene en cuenta la causa asociada a cada incendio, es decir, todos los incendios causados de otra forma que no fuera intencionada, se considera como “no ocurrencia” de incendio forestal y los incendios causados de forma intencionada se considera como “ocurrencia” de incendio.

El modelo se representa de la misma forma que en la sección anterior, la única diferencia está en la variable respuesta, cuya interpretación es ahora la ocurrencia o no ocurrencia de incendios forestales provocados y se escribe como sigue:

$$\begin{aligned}
\text{logit}(P_k) &= \\
&= \log\left(\frac{P_k}{1 - P_k}\right) = \\
&= \beta_0 + f_1(\text{altitud}) + f_2(\text{ta_media}) + f_3(\text{hr}) + f_4(\text{pp}) + \\
&+ f_{\text{estr}}(\text{cdconc}) + f_{\text{estr}}(\text{cdconc})
\end{aligned} \tag{4.4}$$

Las Figuras 4.5 y 4.6 ilustran los efectos suavizados de las covariables altitud, ta_media, hr y pp en la ocurrencia de incendio forestal provocado. La estimación de los grados de libertad del intercepto y las variables explicativas junto con los intervalos de credibilidad al 95 % se muestran en el Cuadro 4.2.

VARIABLES	grados libertad	IC al 95 %
cte	1	(-21.4290, 0.9813)
altitud	1.004	Ver Figura 4.5
ta_media	4.485	Ver Figura 4.5
hr	3.726	Ver Figura 4.6
pp	2.897	Ver Figura 4.6

Cuadro 4.2: Estimación de los grados de libertad e intervalos de credibilidad al 95 %

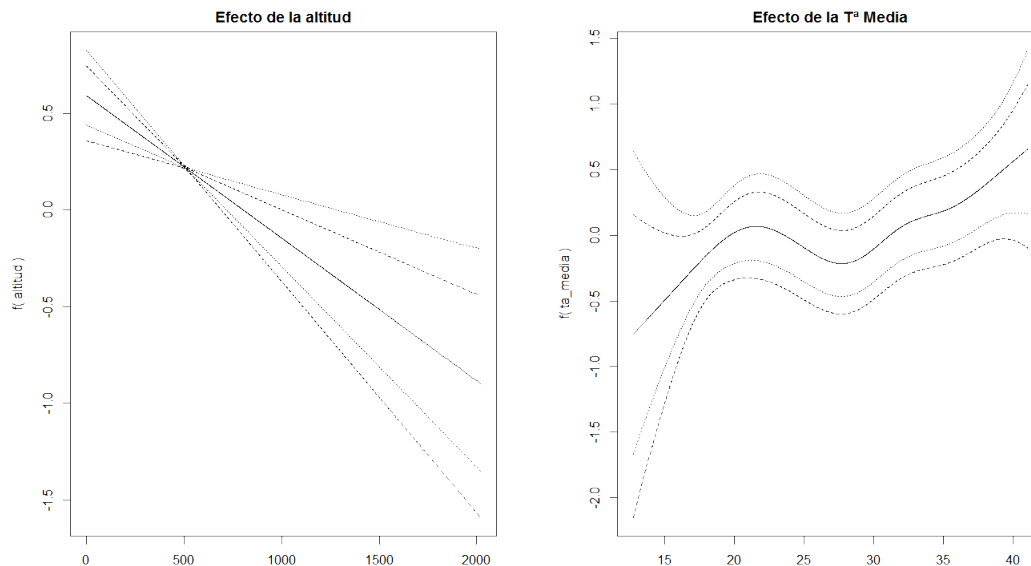


Figura 4.5: Estimación de los efectos suavizados de la altitud (panel izquierdo) y ta_media (panel derecho) con sus correspondientes bandas de credibilidad punteadas al 80 y 95 % - Ocurrencia de incendios forestales provocados

El efecto observado de la altitud con respecto a la ocurrencia de incendios forestales (ver Figura 4.5, panel izquierdo) es prácticamente lineal, produciendo un efecto decreciente a medida que aumenta la altitud. Aproximadamente, en altitudes entorno a los 1000 m las bandas de credibilidad al 80 y 95 % 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 ha sido de 1.004, cifra que responden a un efecto lineal de la covariable en el modelo.

En la Figura 4.5 también se puede observar el efecto de la covariable *ta_media* (panel derecho), apreciándose un efecto no lineal de la covariable *ta_media* sobre la ocurrencia de incendios forestales provocados, que aumenta ligeramente para los incendios a medida que aumentan las temperaturas, pero en donde a temperaturas bajas el aumento creciente de los incendios es más acusado, volviendo de nuevo a acusarse a temperaturas superiores a los aproximadamente 28°C. La estimación de los grados de libertad de esta covariable para ambos modelos se recoge en el Cuadro 4.2.

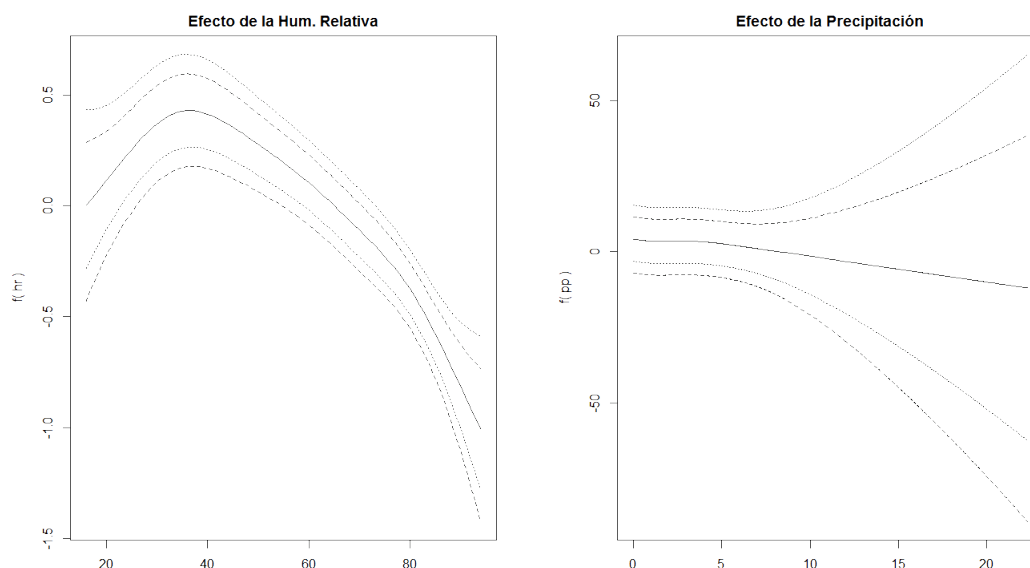


Figura 4.6: Estimación de los efectos suavizados de la *hr* (panel izquierdo) y *pp* (panel derecho) con sus correspondientes bandas de credibilidad punteadas al 80 y 95 % - Ocurrencia de incendios forestales provocados

Los efectos observados de la covariable *hr* (humedad relativa media) en la ocurrencia de incendios forestales provocados (ver Figura 4.6, paneles izquierdo) muestra un efecto no lineal. Esta covariable es ligeramente creciente hasta valores de aproximadamente un 40 % donde, a partir del 40 % de humedad relativa, vuelve a decrecer. El efecto decreciente en la ocurrencia de incendios forestales provocados es moderado hasta aproximadamente el 75 % de humedad relativa, siendo fuertemente decreciente después de esto. La ocurrencia de incendios forestales provocados parece

ser decreciente para humedades relativas superiores al aproximadamente 40 %.

Para la covariable pp (precipitaciones) (ver Figura 4.6, panel derecho) se observa una función prácticamente lineal donde a medida que aumenta la cantidad de l/m^2 de precipitación, las bandas de credibilidad al 80 y 95 % se hacen más amplias, por lo que no debe de sobreinterpretarse los efectos de la precipitación en la ocurrencia de incendios forestales provocados debido a la incertidumbre de estos resultados. Los incendios provocados tienen una menor sensibilidad a las precipitaciones.

La estimación de los grados de libertad para estas covariables (hr y pp) se recogen en el Cuadro 4.2.

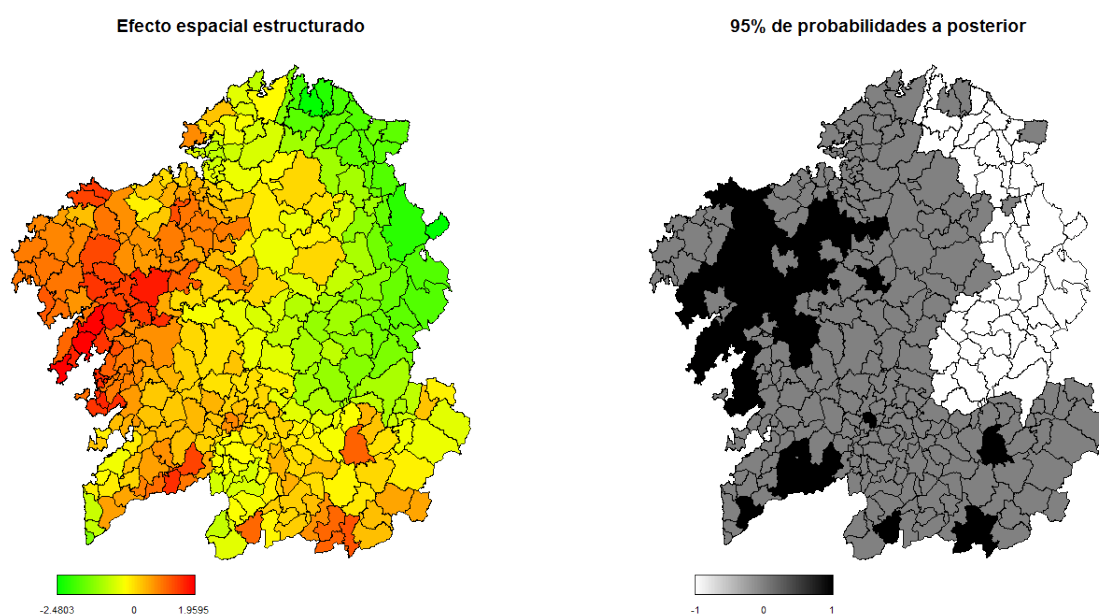


Figura 4.7: Efectos espaciales estructurados del modelo STAR con respuesta binomial - Ocurrencia de incendios forestales provocados

Los efectos espaciales estructurados para la ocurrencia de incendios forestales provocados (ver Figura 4.7, panel derecho) muestran que los concellos situados al noroeste, oeste y suroeste (excepto la punta más saliente del sur) presentan mayor ocurrencia de incendios que los concellos situados al noreste y este de Galicia, que se sitúan los concellos con los valores más bajos para la variable respuesta. Los intervalos de credibilidad al 95 % se muestran en la Figura 4.7 (panel derecho) donde los concellos en negro representan mucha probabilidad de ocurrencia de incendios frente a los concellos en blanco que representan una baja probabilidad de ocurrencia de incendios forestales provocados. Los concellos en gris son concellos no significativos frente a la variable en estudio. Aumenta el número de concellos significativos

donde las zonas núcleo parecen ser más grandes y donde el mapa parece más claro en relación a los resultados obtenidos para todos los incendios.

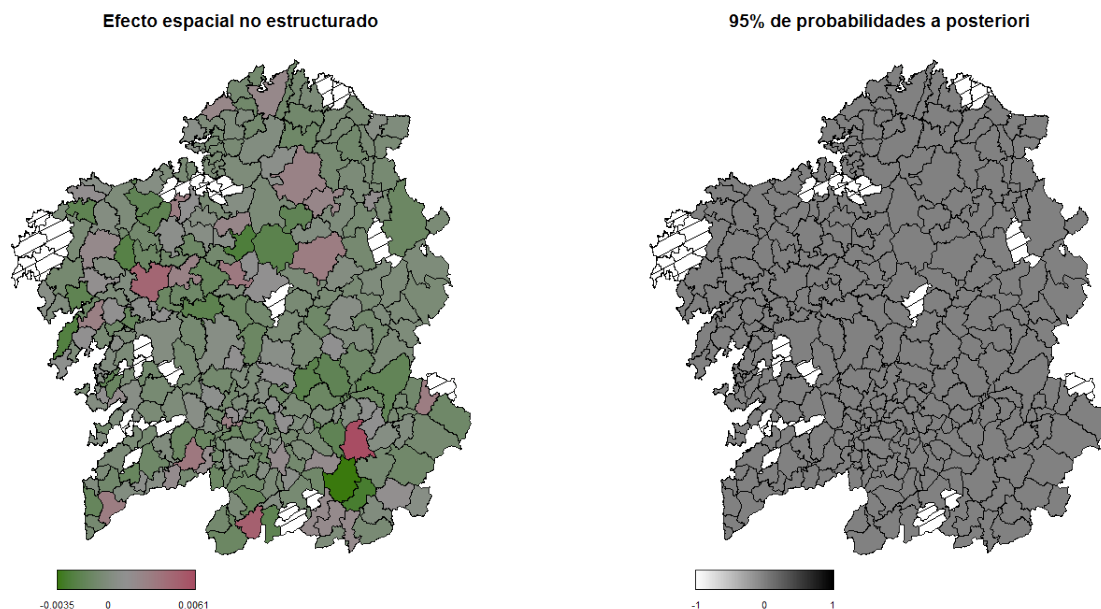


Figura 4.8: Efectos espaciales no estructurados del modelo STAR con respuesta binomial - Ocurrencia de incendios forestales provocados

Los resultados de las efectos espaciales no estructurados (ver Figura 4.8) no apuntan en ninguna dirección específica, sin embargo se observan concellos con efectos espaciales no estructurados mas positivos (concellos de color rosa), lo que contribuye a una mayor posibilidad de efecto local, pero que no son municipios significativos (concellos de color gris) como se aprecia en en la Figura 4.8 (panel derecho). Los concellos pintados de blanco y rayados no aportan información de ningún tipo, ya que no se obtuvo dato para poder estimar este efecto.

4.1.4. Ajuste de modelos

El enfoque a seguir en esta sección es el ajuste de los modelos anteriores a partir de criterios AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) y GCV (Generalized Cross-Validation). En el Cuadro 4.3 se presentan los 2 modelos ajustados para las variables meteorológicas, los cuales quedan definidos por el predictor. Los valores de AIC, BIC y GCV se exponen para los dos modelos. También son incluidos los grados de libertad de cada modelo estimado.

Modelos	2xlog-likelihood	g.l.	AIC	BIC	GCV
Binomial (totales)	13632.5	160.732	13954	15623.6	0.057
Binomial (provocados)	10830.9	119.137	11069.2	12306.7	0.045

Cuadro 4.3: Criterio de información Akaike (AIC), criterio de información Bayesiano (BIC), y validación cruzada generalizada (GCV) para los dos modelos.

4.2. Modelización a partir de grupos de modelos de combustible

Como ya se comentó anteriormente, la probabilidad de incendio depende de factores complejos como el tipo de estructura de la vegetación entre otros, que en algunos casos favorece la ignición de un incendio y en otros puede frenar su avance (Nunes, et al. 2005).

En esta sección se presentan aplicaciones de los modelos de regresión STAR para para la ourencia de incendios forestales totales y provocados a partir de variables de tipo de vegetación en cada cuadrícula que posteriormente se agrega por concello.

4.2.1. Modelización de los incendios forestales totales

Partiendo del análisis de los datos, en este apartado se trata de modelizar la ocurrencia de incendios forestales totales con las variables de tipos de modelos de combustible agrupados en sin asignar modelo, pastos, matorral, hojarasca bajo arbolado y por último el grupo de restos de cota y operaciones selvícolas.

El modelo se representa por:

$$\begin{aligned}
\text{logit}(P_k) &= \\
&= \log\left(\frac{P_k}{1 - P_k}\right) = \\
&= \beta_0 + f_1(\text{altitud}) + f_2(\text{Sin_Asignar_Modelo}) + f_3(\text{Pastos}) + \\
&+ f_4(\text{Matorral}) + f_5(\text{Hojarasca}) + f_6(\text{Restos}) + \\
&+ f_{\text{estr}}(\text{cdconc}) + f_{\text{estr}}(\text{cdconc})
\end{aligned} \tag{4.5}$$

En donde se considera k como el índice del conjunto de voxels espacio-tiempo 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 4.9, 4.10 y 4.11 ilustran los efectos suavizados de las covariables altitud, sin asignar modelo, pastos, matorral, hojarasca y restos en la ocurrencia de incendio forestal para los modelos STAR con respuesta binomial. Además, la estimación de los grados de libertad del intercepto y los efectos de las covariables junto con los intervalos de credibilidad al 95 % se muestran en el Cuadro 4.4.

Variab les	grados libertad	IC al 95 %
cte	1	(-4.0702, -2.8204)
altitud	1.008	Ver Figura 4.9
Sin_Datos	3.596	Ver Figura 4.9
Pasto	1.165	Ver Figura 4.10
Matorral	1.013	Ver Figura 4.10
Hojarasca	2.134	Ver Figura 4.11
Restos	1.002	Ver Figura 4.11

Cuadro 4.4: Estimación de los grados de libertad e intervalos de credibilidad al 95 %

Los grados de libertad de cada función suavizada dan una clara indicación de la necesidad de incorporar la flexibilidad a los modelos y además permite detectar linealidad (estimación de los grados de libertad (edf) = 1).

Los efectos de las covariables altitud y grupo de ausencia de datos se muestran en la Figura 4.9 en donde se observa que los efectos de la covariable altitud (panel izquierdo) es lineal y a su vez decrece la ocurrencia de incendios según aumenta el valor de la altitud para el periodo considerado. A partir de los aproximadamente 800 m de altitud las bandas de credibilidad tanto al 80 como al 95 % se hacen cada vez más amplias por lo tanto, no se debe de sobreinterpretar los efectos de la altitud debido a la incertidumbre de estos resultados.

Los efectos de la covariable Sin asignar modelo (ver Figura 4.9, panel derecho) 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.

La estimación de los grados de libertad de ambas covariables presentadas para los dos modelos STAR de estudio se recogen en el Cuadro 4.4.

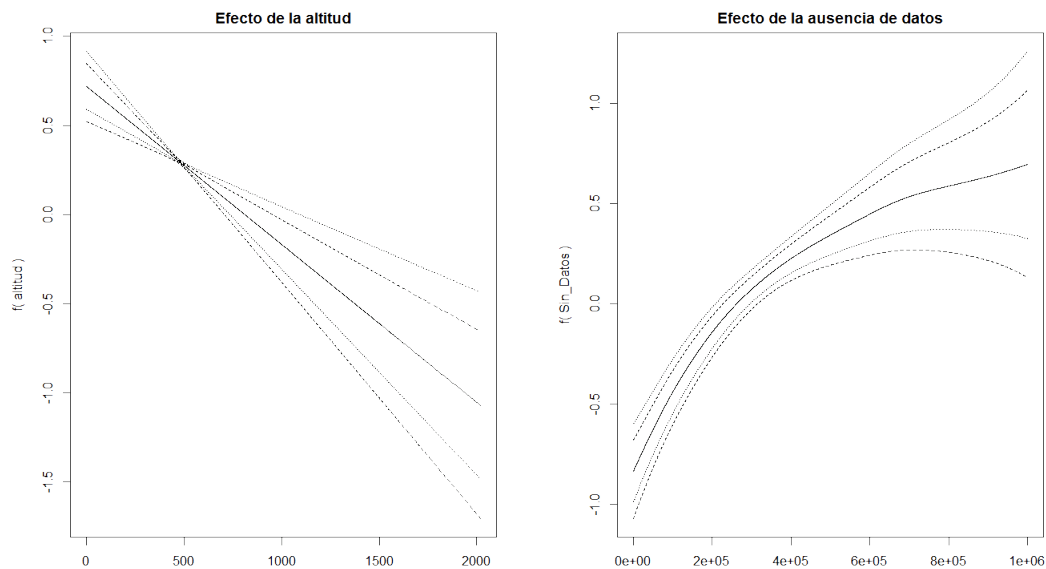


Figura 4.9: Estimación de los efectos suavizados de altitud (figura izquierda) y grupo asociado a la variable Sin Asignar Modelo de tipo de combustible (figura derecha) con sus correspondientes bandas de credibilidad punteadas al 80 y 95 % - Ocurrencia de incendios forestales totales

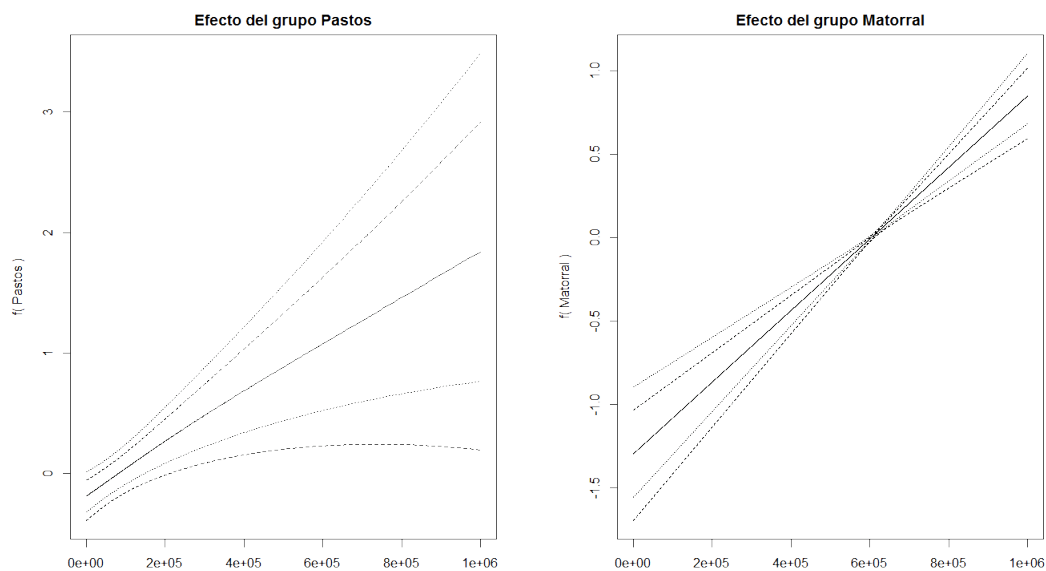


Figura 4.10: Estimación de los efectos suavizados del grupo Pastos (figura izquierda) y grupo Matorral (figura derecha) con sus correspondientes 80 y 95 % de bandas de credibilidad - Ocurrencia de incendios forestales totales

Los efectos de grupo de Pastos y grupo de Matorral (ver Figura 4.17) 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. También cabe destacar (ver apéndice A) que la variable con mayor representación por vóxeles y por lo tanto mayor representación en las zonas de ocurrencia de incendios es el grupo de matorral.

La estimación de los grados de libertad de ambas covariables estudiadas se recogen en el Cuadro 4.4.

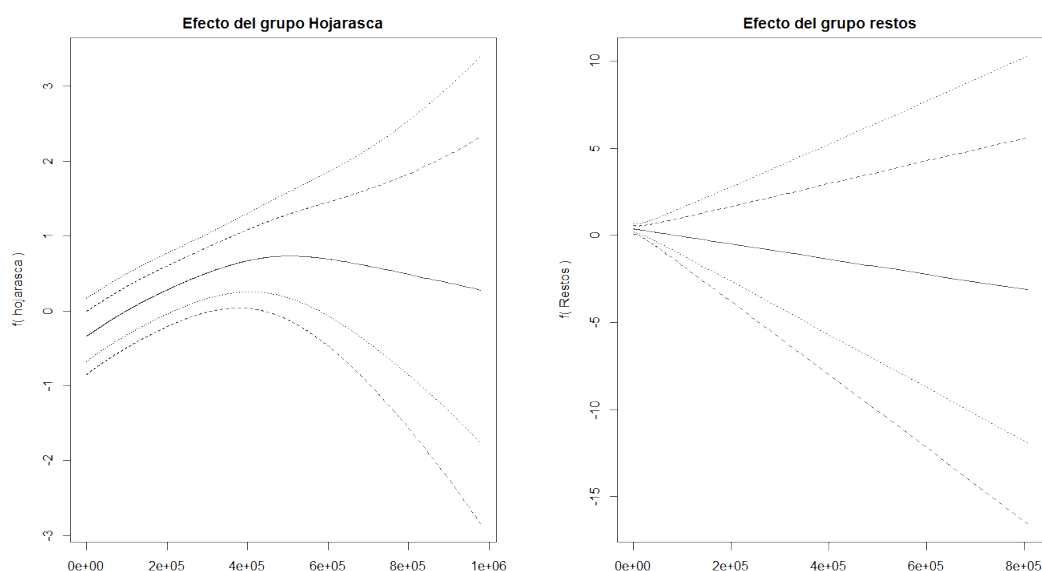


Figura 4.11: Estimación de los efectos suavizados del grupo de Hojarasca (figura izquierda) y grupo Restos (figura derecha) con sus correspondientes bandas de credibilidad punteadas al 80 y 95 % - Ocurrencia de incendios forestales totales

Los efectos de las variables grupo de hojarasca y grupo de restos se muestran en la Figura 4.11 donde se observa que la primera de las variables, grupo de hojarasca (panel de la izquierdo), 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). Por su parte, la segunda variable de estudio en la Figura 4.11, variable grupo de restos (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.

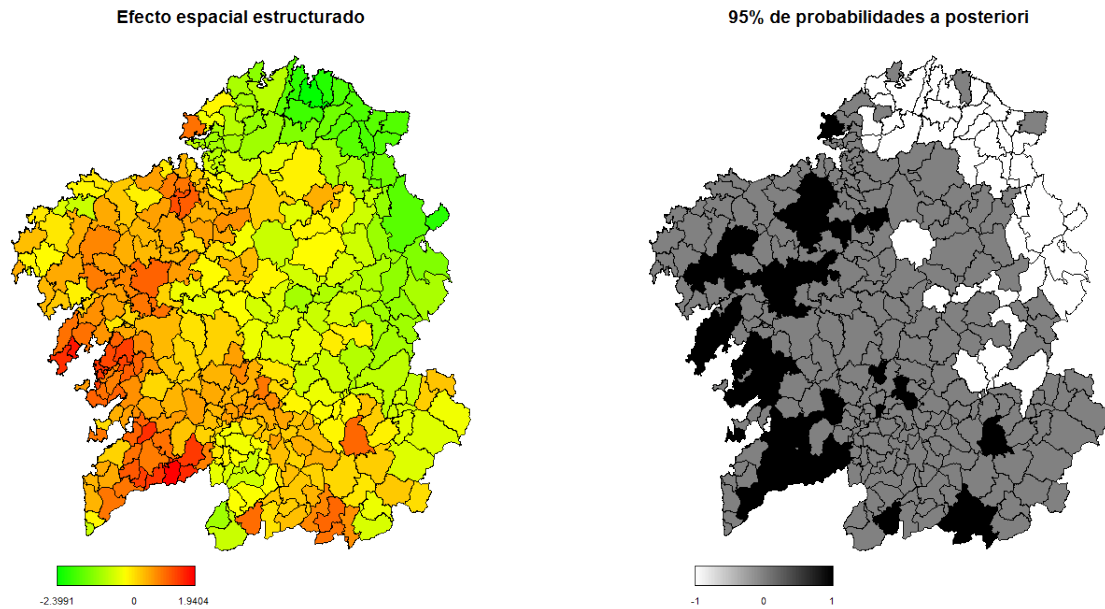


Figura 4.12: Efectos espaciales estructurados - Ocurrencia de incendios forestales totales

Los efectos espaciales estructurados para la ocurrencia de incendios forestales (ver Figura 4.12, panel izquierdo) muestran que los concellos situados al norte y noreste presentan la menor ocurrencia de incendios forestales que en los concellos situados al oeste y sur de Galicia, que se sitúan los concellos con los valores más altos para la variable de estudio. Los intervalos de credibilidad al 95 % se muestran en la Figura 4.12 (panel derecho) donde los concellos en negro representan mucha probabilidad de ocurrencia de incendios frente a los concellos en blanco que representan una baja probabilidad de ocurrencia de incendios forestales provocados. Los concellos en gris son concellos no significativos frente a la variable en estudio.

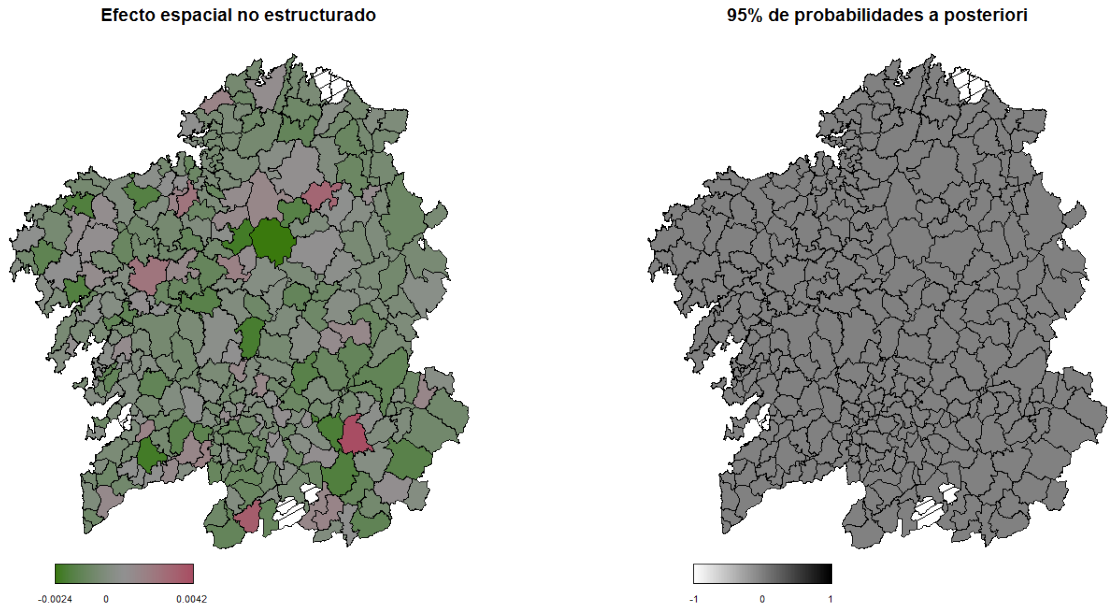


Figura 4.13: Efectos espaciales no estructurados - Ocurrencia de incendios forestales totales

Los resultados de los efectos espaciales no estructurados (ver Figura 4.12) no apuntan en ninguna dirección específica, sin embargo se observan concellos con efectos espaciales no estructurados mas positivos (color rosa), lo que contribuye a una mayor posibilidad de efecto local, pero que son municipios no significativos (color gris) como se aprecia en en la Figura 4.12 (panel derecho). Los concellos en blanco y rayados no aportan información de ningún tipo, ya que no se obtuvo dato para poder estimar este efecto.

4.2.2. Modelización de los incendios forestales provocados

En este apartado se trata de modelizar la ocurrencia de incendios forestales provocados con las variables de tipos de modelos de combustible agrupados en sin asignar modelo, pastos, matorral, hojarasca bajo arbolado y por último el grupo de restos de cota y operaciones selvícolas.

El modelo se representa por:

$$\begin{aligned}
\text{logit}(P_k) &= \\
&= \log\left(\frac{P_k}{1 - P_k}\right) = \\
&= \beta_0 + f_1(\text{altitud}) + f_2(\text{Sin_Asignar_Modelo}) + f_3(\text{Pastos}) + \\
&+ f_4(\text{Matorral}) + f_5(\text{Hojarasca}) + f_6(\text{Restos}) + \\
&+ f_{\text{estr}}(\text{cdconc}) + f_{\text{estr}}(\text{cdconc})
\end{aligned} \tag{4.6}$$

Las Figuras 4.14, 4.15 y 4.16 muestran los efectos suavizados de las covariables de estudio y la Tabla 4.5 se da la estimación de los grados de libertad e intervalos de credibilidad al 95 % como se muestra a continuación:

Variabes	grados libertad	IC al 95 %
cte	1	(-4.0455, -3.3717)
altitud	1.058	Ver Figura 4.14
Sin_Datos	3.352	Ver Figura 4.14
Pasto	1.037	Ver Figura 4.15
Matorral	2.631	Ver Figura 4.15
Hojarasca	1.012	Ver Figura 4.16
Restos	1.003	Ver Figura 4.16

Cuadro 4.5: Estimación de los grados de libertad e intervalos de credibilidad al 95 %

Las funciones de suavización de las covariables de estudio presenta unos resultados muy similares a los obtenidos para los incendios totales de la sección anterior. Cabe destacar que la covariable grupo de Matorral pasa a tener ahora un efecto no lineal con un número de grados de libertad de 2.631. Esta covariable junto con la covariable Sin_Asignar_Modelo son las que mejor explican la ocurrencia de incendios forestales, ya que como se estudió en el apéndice A, son las covariables con mayor representación por vóxeles y por lo tanto en zonas ocupadas por incendios. Ambas son funciones crecientes, según aumenta la superficie forestal, aumenta el riesgo de incendios en Galicia.

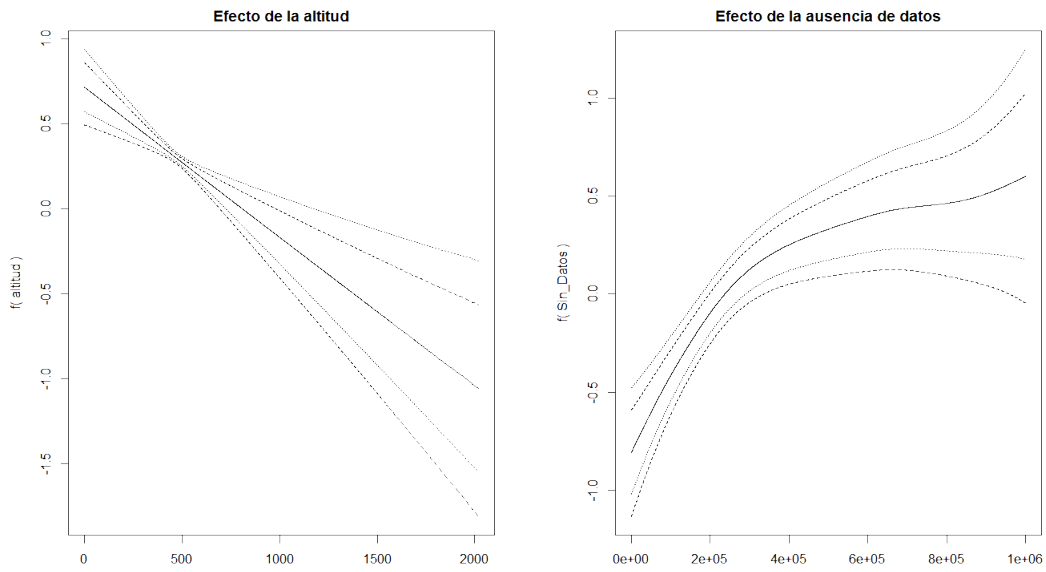


Figura 4.14: Estimación de los efectos suavizados de la altitud (figura izquierda) y Sin Asignar Modelo (figura derecha) con sus correspondientes 80 y 95 % de bandas de credibilidad - Ocurrencia de incendios forestales provocados

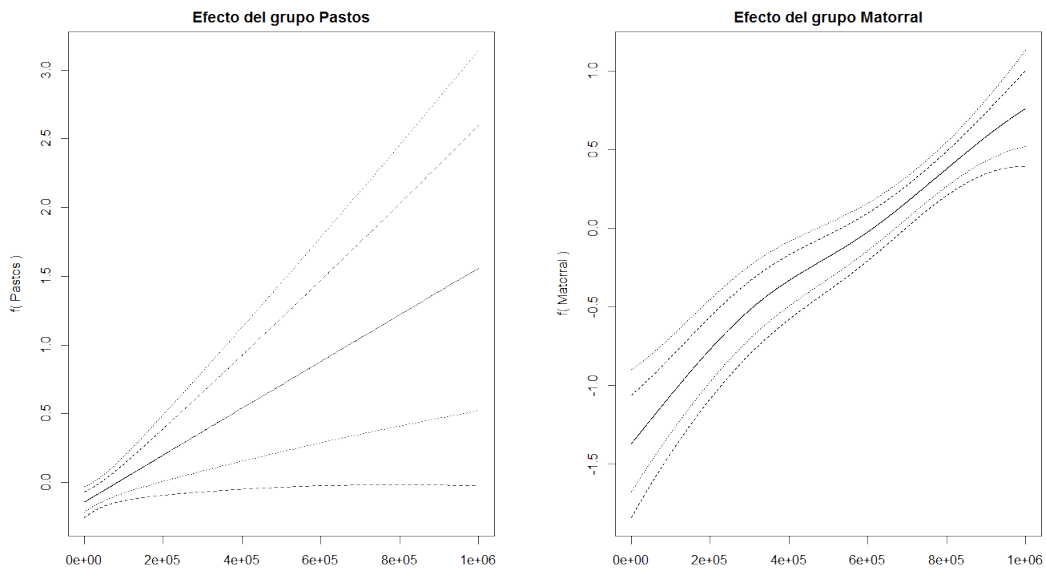


Figura 4.15: Estimación de los efectos suavizados del grupo Pastos (figura izquierda) y del grupo Matorral (figura derecha) con sus correspondientes 80 y 95 % de bandas de credibilidad - Ocurrencia de incendios forestales provocados

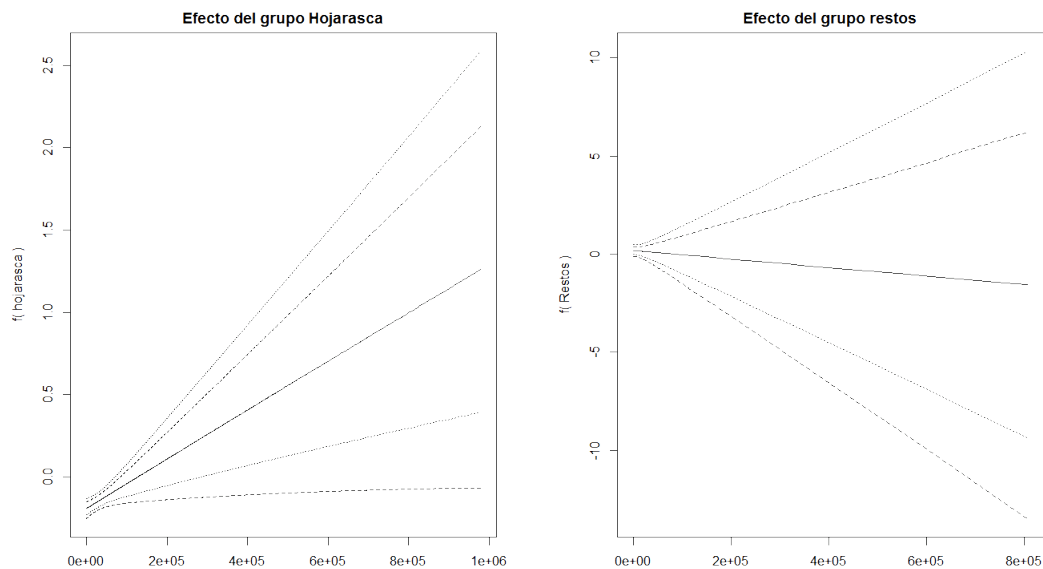


Figura 4.16: Estimación de los efectos suavizados del grupo Hojarasca (figura izquierda) y grupo Restos (figura derecha) con sus correspondientes 80 y 95% de bandas de credibilidad - Ocurrencia de incendios forestales provocados

El efecto espacial estructurado (ver Figura 4.17) sigue marcando la tendencia de mayor concentración 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.

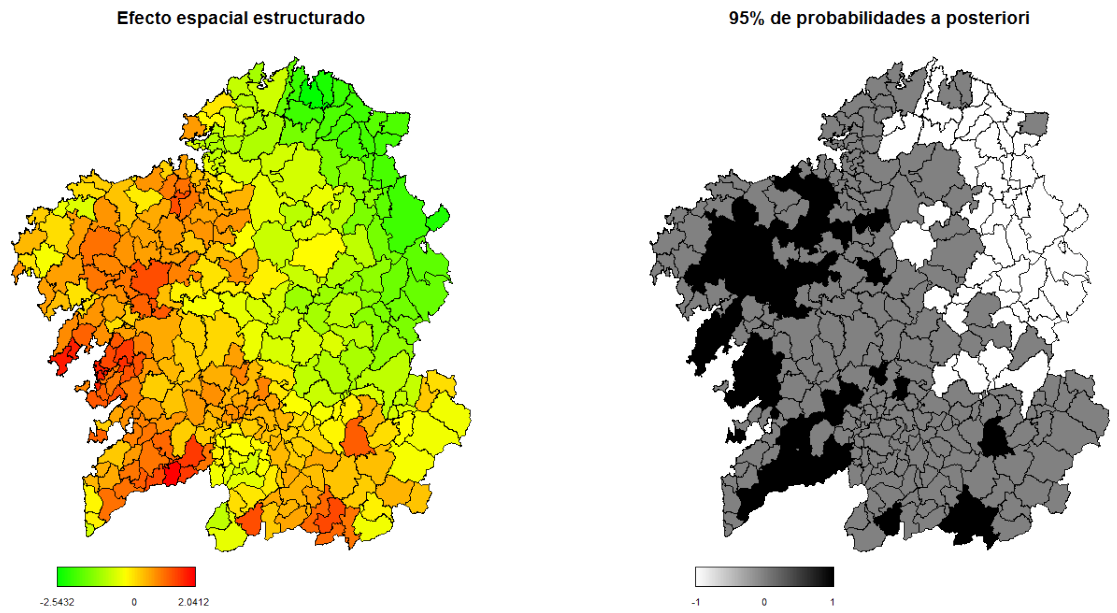


Figura 4.17: Efectos espaciales estructurados - Ocurrencia de incendios forestales provocados

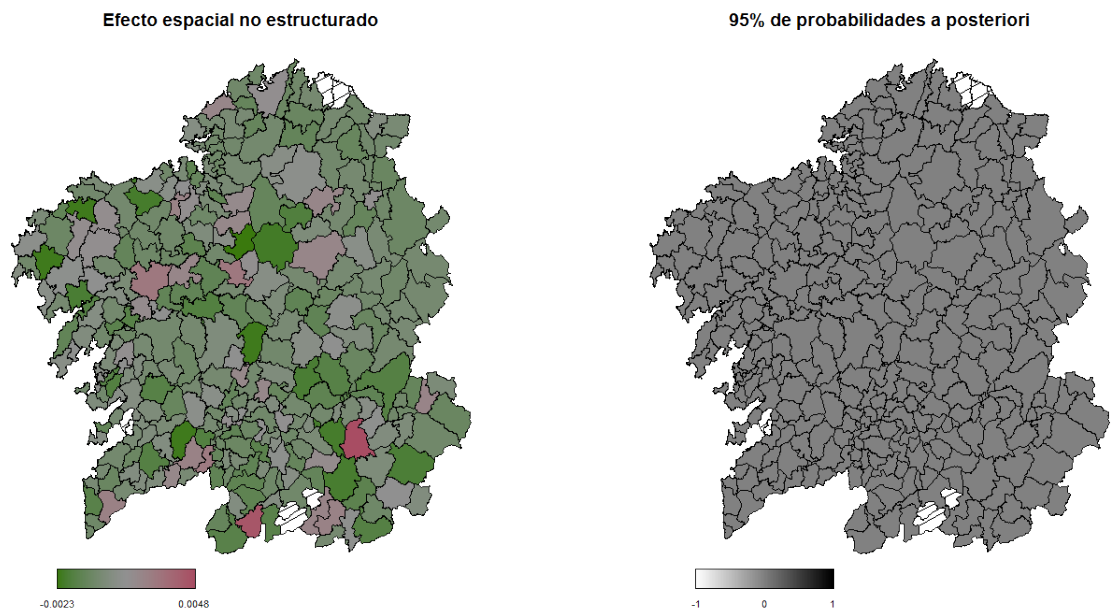


Figura 4.18: Efectos espaciales no estructurados - Ocurrencia de incendios forestales provocados

Los resultados de los efectos espaciales no estructurados (ver Figura 4.18) no apuntan en ninguna dirección específica. Todos los concellos son no significativos (color gris, panel derecho). Los concellos en blanco y rayados no aportan información de ningún tipo, ya que no se obtuvo dato para poder estimar este efecto.

4.2.3. Ajuste de modelos

En esta sección se muestra el ajuste de los modelos STAR para la ocurrencia de incendios forestales provocados y totales a partir de las variables grupos de tipos de combustible mediante los criterios AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) y GCV (Generalized Cross-Validation). En el Cuadro 4.6 se presentan los 2 modelos ajustados. También son incluidos los grados de libertad de cada modelo estimado.

Modelos	2xlog-likelihood	g.l.	AIC	BIC	GCV
Binomial (totales)	9565.87	142.573	9851.02	11026.2	0.344
Binomial (provocados)	7907.5	138.982	8185.46	9329.64	0.287

Cuadro 4.6: Criterio de información Akaike (AIC), criterio de información Bayesiano (BIC), y validación cruzada generalizada (GCV) para los dos modelos seleccionados.

Como se fue observando a lo largo de esta sección, las variables que aportan mayor información por voxel al modelo son las variables “grupo de matorral” y “grupo Sin Asignar Modelo”, ya que son las variables que mejor definen el tipo de estructura predominante en Galicia. Se puede concluir de esta sección diciendo que los modelos de combustibles de Rothermel para la estimación de la ocurrencia de incendios no explican suficientemente la presencia de incendios. En el futuro como posible mejora, se debería trabajar con formaciones vegetales en vez de modelos de combustible.

4.3. Modelización a partir de variables climatológicas y grupos de modelos de combustible

Después de haber estudiado por separado en secciones anteriores las variables meteorológicas y de tipo de combustible, el siguiente paso es establecer un modelo que englobe todas estas variables y así observar si el modelo permite estimar de forma más precisa la ocurrencia de incendios forestales totales y provocados. A continuación se describen los modelos de regresión STAR de respuesta binomial para la ocurrencia de incendios forestales totales y provocados que engloban todas las covariables estudiadas anteriormente (altitud, ta_media, hr, pp, Sin Asignar Modelo, Pastos, Matorral, Hojarasca y Restos).

4.3.1. Modelización de los incendios forestales totales

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

El modelo se representa por:

$$\begin{aligned} \text{logit}(P_k) &= \\ &= \log\left(\frac{P_k}{1-P_k}\right) = \\ &= \beta_0 + f_1(\text{altitud}) + f_2(\text{ta_media}) + f_3(\text{hr}) + f_4(\text{pp}) + \\ &+ f_5(\text{Sin_Asignar_Modelo}) + f_6(\text{Pastos}) + f_7(\text{Matorral}) + \\ &+ f_8(\text{Hojarasca}) + f_9(\text{Restos}) + \\ &+ f_{\text{estr}}(\text{cdconc}) + f_{\text{no_estr}}(\text{cdconc}) \end{aligned} \quad (4.7)$$

Para simplificar la notación consideramos k como el índice del conjunto de voxels espacio-tiempo 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 4.19, 4.20, 4.21 y 4.22 ilustran los efectos de las covariables para la ocurrencia de incendios forestal totales con modelos 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 el Cuadro 4.7.

Variabes	grados libertad	IC al 95 %
cte	1	(-9.3666, -7.5362)
altitud	1.089	Ver Figura 4.19
ta_media	4.075	Ver Figura 4.19
hr	5.183	Ver Figura 4.20
pp	1.071	Ver Figura 4.20
Sin Asignar Modelo	3.643	Ver Figura 4.21
Pasto	1.009	Ver Figura 4.21
Matorral	1.014	Ver Figura 4.22
Hojarasca	2.149	Ver Figura 4.22
Restos	1.001	Ver Figura 4.22

Cuadro 4.7: 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 4.19. 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, como ya se comentó en secciones anteriores a que en el periodo de estudio considerado en este trabajo, la mayoría de los incendios ocurrieron en zonas costeras en donde las altitudes tienden 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 ($< 20^{\circ}C$).

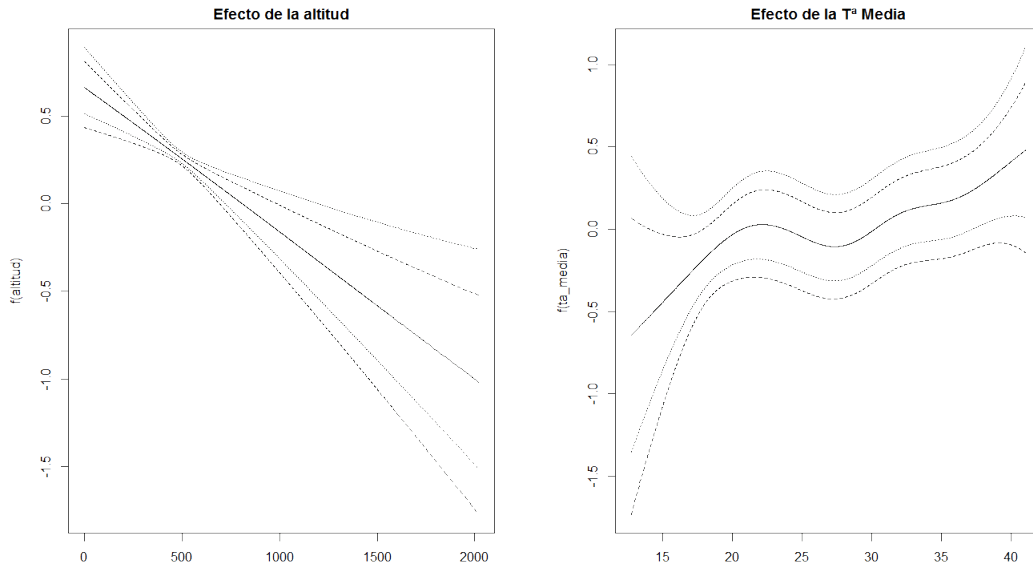


Figura 4.19: Estimación de los efectos suavizados de la altitud (figura izquierda) y ta_media (figura derecha) con sus correspondientes bandas de credibilidad punteadas al 80 y 95 % - Ocurrencia de incendios forestales totales

El efecto de la covariable hr (ver Figura 4.20, panel derecho) 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 (entorna al 75%). Incrementos de humedad relativa están asociados a un descenso de ocurrencia de incendios. El efecto de la covariable pp (ver Figura 4.20, panel derecho) 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 el Cuadro 4.7.

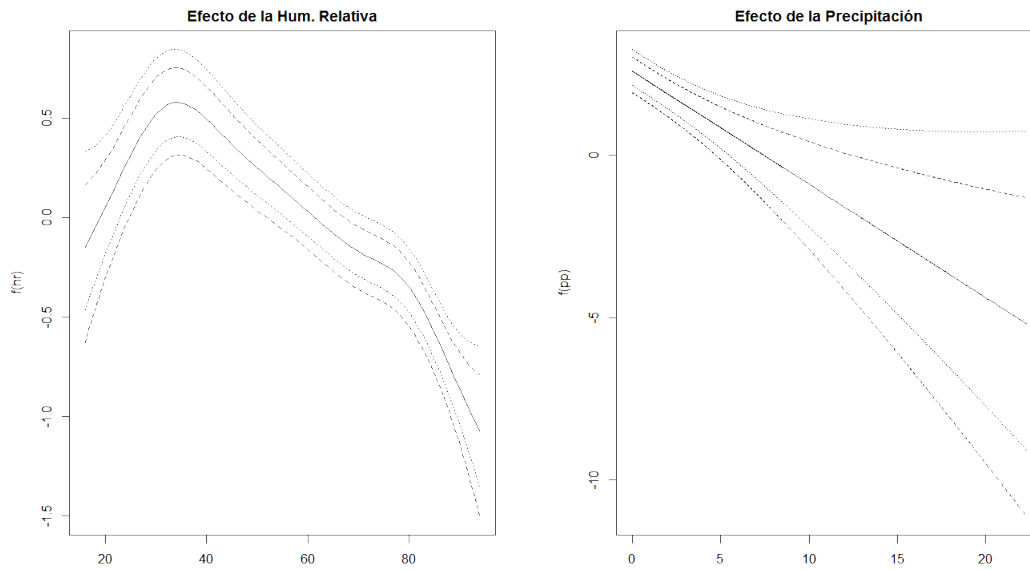


Figura 4.20: Estimación de los efectos suavizados de las covariables hr (figura izquierda) y pp (figura derecha) con sus correspondientes bandas de credibilidad al 80 y 95 % - Ocurrencia de incendios forestales totales

Las Figuras 4.21 y 4.22 muestran el efecto de las variables de grupos de modelos de combustibles introducidas en el modelo global. De ellas, cabe destacar de nuevo, que las variables Sin Asignar Modelo y grupo de Matorral son las variables con mayor representación por superficie de vóxel y por lo tanto las que mejor explican la ocurrencia de incendios forestales totales en la superficie gallega. A medida que aumenta la superficie por m^2 de vóxel, la ocurrencia de incendio es mayor.

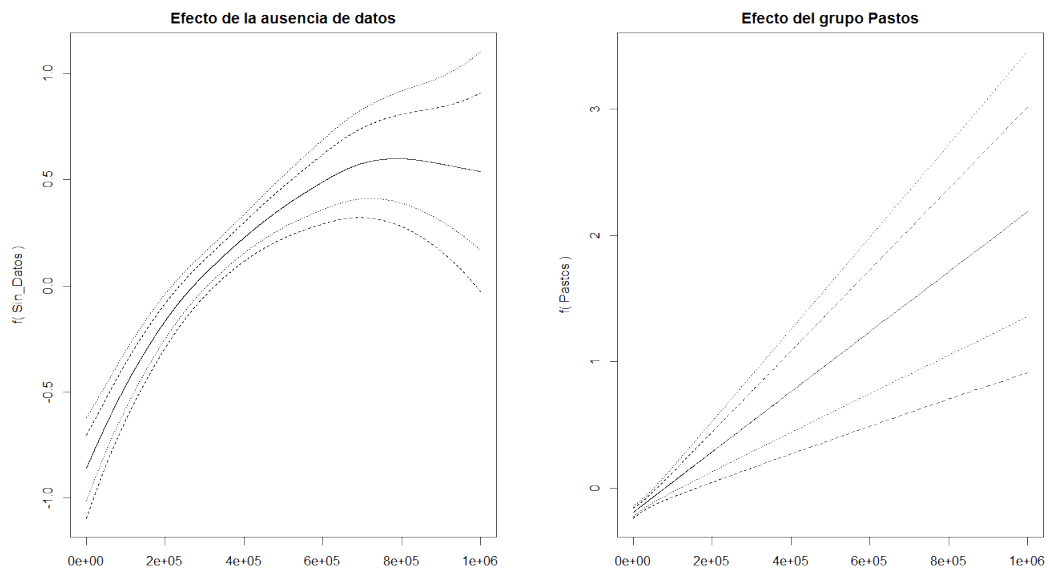


Figura 4.21: Estimación de los efectos suavizados de la variable Sin Asignar Modelo (figura izquierda) y grupo de Pastos (figura derecha) con sus correspondientes 80 y 95 % de bandas de credibilidad - Ocurrencia de incendios forestales totales

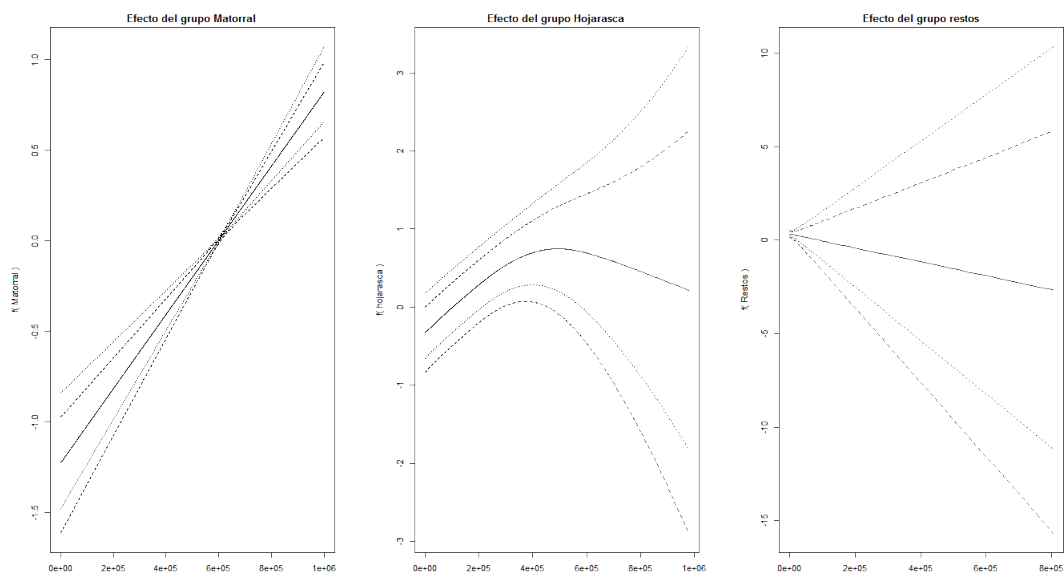


Figura 4.22: Estimación de los efectos suavizados de los grupos Matorral, Hojarasca y Restos con sus correspondientes 80 y 95 % de bandas de credibilidad - Ocurrencia de incendios forestales totales

Los efectos espaciales se muestran en las Figuras 4.23 y 4.24. Los efectos espaciales estructurados (Figura 4.23) siguen marcando 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 4.24) 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.

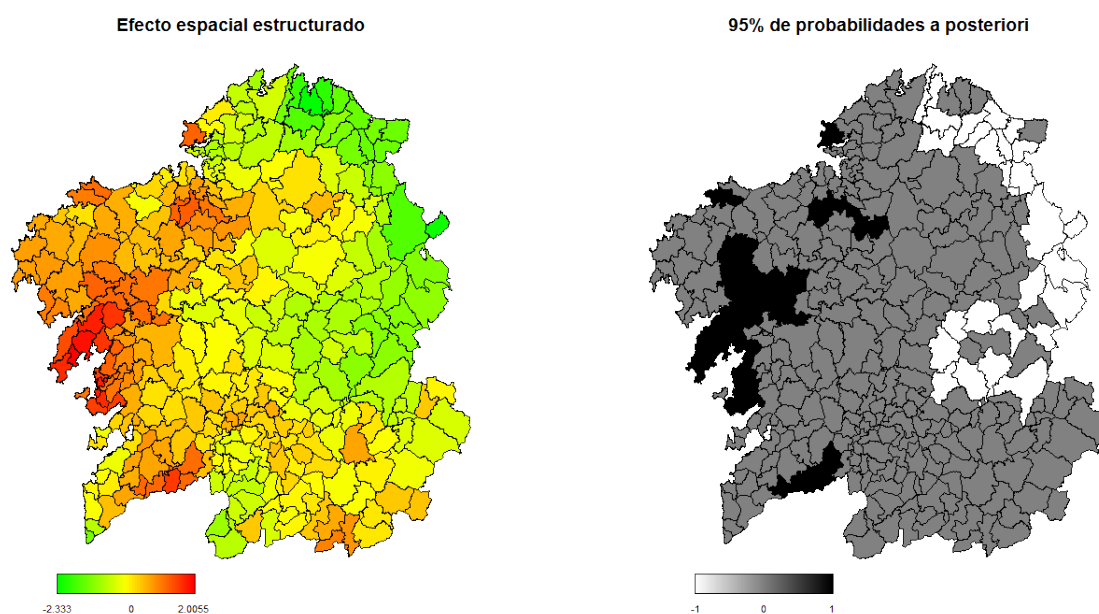


Figura 4.23: Efectos espaciales estructurados - Ocurrencia de incendios forestales totales

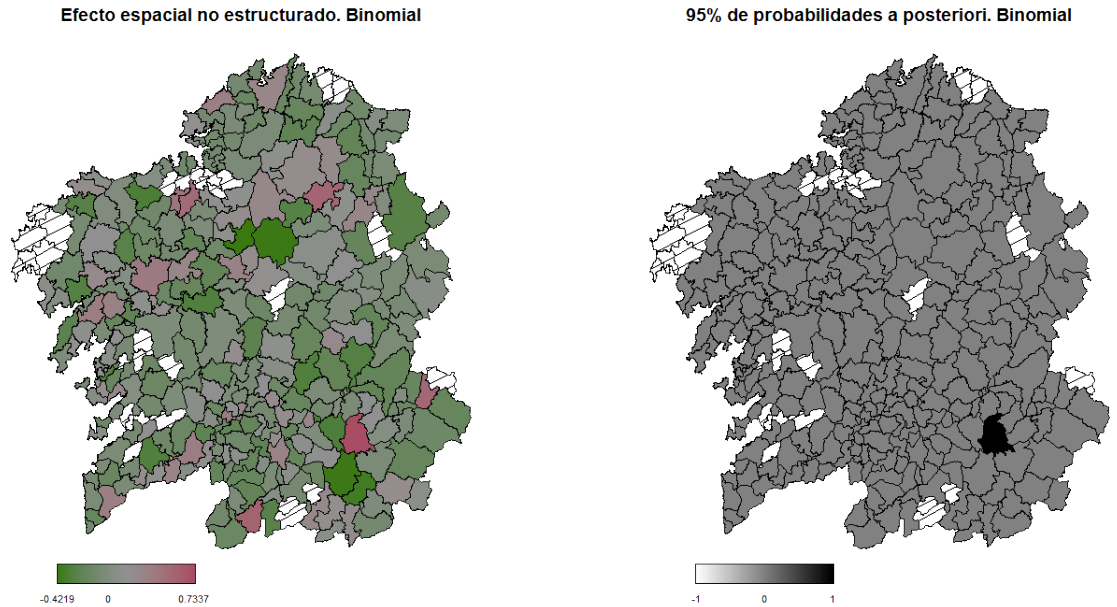


Figura 4.24: Efectos espaciales no estructurados - Ocurrencia de incendios forestales totales

4.3.2. Modelización de los incendios forestales provocados

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

El modelo se representa por:

$$\begin{aligned}
 \text{logit}(P_k) &= \\
 &= \log \left(\frac{P_k}{1 - P_k} \right) = \\
 &= \beta_0 + f_1(\text{altitud}) + f_2(\text{ta_media}) + f_3(\text{hr}) + f_4(\text{pp}) + \\
 &+ f_5(\text{Sin_Asignar_Modelo}) + f_6(\text{Pastos}) + f_7(\text{Matorral}) + \\
 &+ f_8(\text{Hojarasca}) + f_9(\text{Restos}) + \\
 &+ f_{\text{estr}}(\text{cdconc}) + f_{\text{no estr}}(\text{cdconc})
 \end{aligned} \tag{4.8}$$

Las Figuras 4.25, 4.26, 4.27 y 4.28 ilustran los efectos de las covariables de estudio en la ocurrencia de incendios forestales provocados con modelos STAR de respuesta binomial. Además, la estimación de los grados de libertad del intercepto y

los efectos de las covariables junto con los intervalos de credibilidad al al 80 y 95 % se muestran en el Cuadro 4.8.

Variabes	grados libertad	IC al 95 %
cte	1	(-21.6312, 1.0767)
altitud	1.002	Ver Figura 4.25
ta_media	4.362	Ver Figura 4.25
hr	3.591	Ver Figura 4.26
pp	2.912	Ver Figura 4.26
Sin Asignar Modelo	2.783	Ver Figura 4.27
Pasto	1.011	Ver Figura 4.27
Matorral	3.266	Ver Figura 4.28
Hojarasca	1.001	Ver Figura 4.28
Restos	1.003	Ver Figura 4.28

Cuadro 4.8: Estimación de los grados de libertad e intervalos de credibilidad al 95 %

El efecto de la altitud con respecto a la ocurrencia de incendios forestales (ver Figura 4.25, panel izquierdo) es prácticamente lineal, produciendo un efecto decreciente a medida que aumenta la altitud. Altitudes entorno a los 1000 m, las bandas de credibilidad al 80 y 95 % se hacen más amplias. No se debe sobreinterpretar los efectos debido a la incertidumbre de estos resultados. El efecto de la covariable ta_media (Figura 4.25, panel derecho) es no lineal, aumentando la ocurrencia de incendios a medida que las temperaturas aumentan, pero en donde a temperaturas bajas el aumento creciente de los incendios es más acusado, volviendo de nuevo a acusarse a temperaturas superiores a los aproximadamente 28°C. La estimación de los grados de libertad para ambas covariables explicadas se recoge en el Cuadro 4.8.

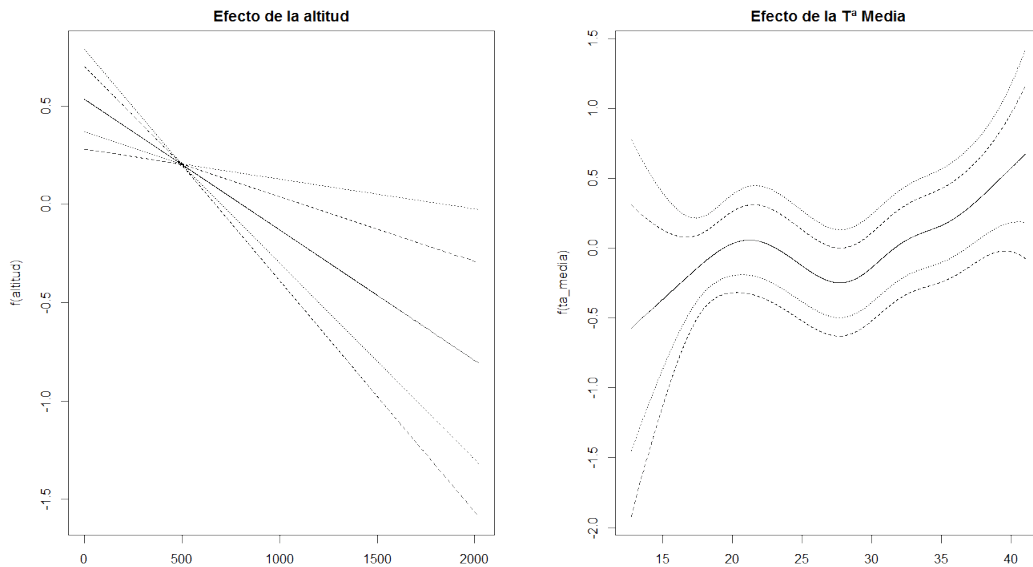


Figura 4.25: Estimación de los efectos suavizados de las covariables altitud (figura izquierda) y ta_media (figura derecha) con sus correspondientes bandas de credibilidad punteadas al 80 y 95 % - Ocurrencia de incendios forestales provocados

Los efectos observados de la covariable hr (humedad relativa media) en la ocurrencia de incendios forestales provocados (ver Figura 4.26, paneles izquierdo) muestra un efecto no lineal. Esta covariable es ligeramente creciente hasta valores de aproximadamente un 35 % de humedad relativa, y que a partir del 35 % de humedad relativa decrece considerablemente. Por lo tanto, la ocurrencia de incendios forestales provocados parece ser creciente a humedades bajas y decrece para humedades relativas superiores al 35 % aproximadamente.

La covariable pp (precipitaciones) (ver Figura 4.26, panel derecho) se observa una función prácticamente lineal y constante donde, a medida que aumenta la cantidad de l/m^2 de precipitación, las bandas de credibilidad al 80 y 95 % se hacen más amplias, por lo que no debe de sobreinterpretarse los efectos de la precipitación en la ocurrencia de incendios forestales provocados. Los incendios provocados tienen una menor sensibilidad a las precipitaciones.

La estimación de los grados de libertad para estas covariables (hr y pp) se recogen en el Cuadro 4.8.

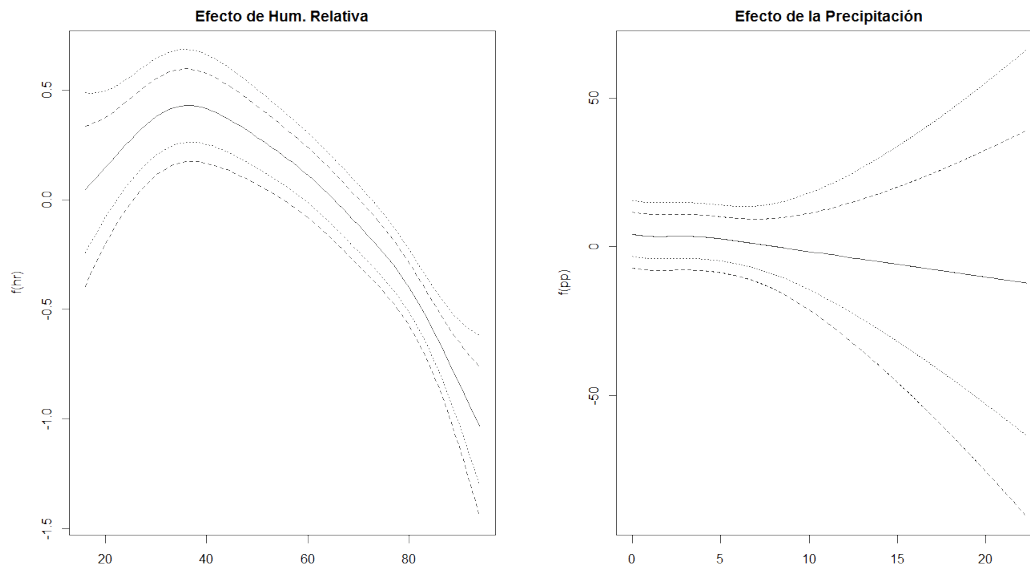


Figura 4.26: Estimación de los efectos suavizados de las covariables hr(figura izquierda) y pp (figura derecha) con sus correspondientes bandas de credibilidad al 80 y 95 % - Ocurrencia de incendios forestales provocados

Las efectos de las funciones de suavización de las variables grupos de modelos de combustible se muestran en las Figuras 4.27 y 4.28, donde de nuevo se observa que la variable Sin asignar modelo y grupo de matorral son las variables que mejor explican la ocurrencia de incendios forestales, ya que como se estudió en el apéndice A, son las variables con mayor representación por vóxeles y por lo tanto en zonas ocupadas por incendios. Ambas son funciones crecientes, según aumenta la superficie forestal, aumenta el riesgo de incendios en Galicia. El resto de variables de tipo de combustible son funciones que a medida que aumenta la superficie por m^2 de voxel se separan cada vez más las banda de credibilidad al 80 y 95 %, por lo que no debe de sobreinterpretarse los efectos de estas variables en la ocurrencia de incendios forestales provocados.

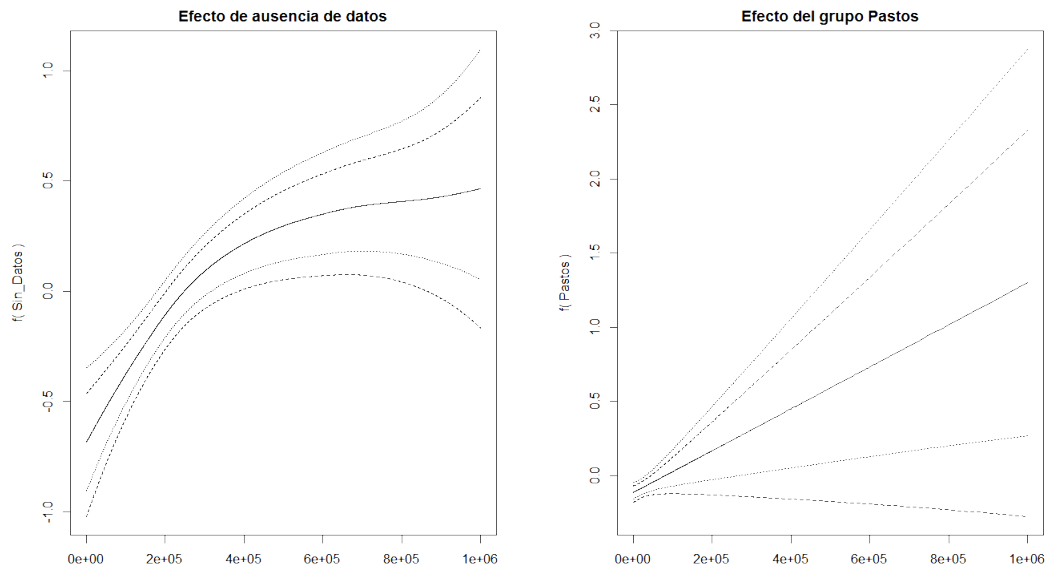


Figura 4.27: Estimación de los efectos suavizados de las covariables Sin Asignar Modelo (figura izquierda) y grupo de Pastos (figura derecha) con sus correspondientes 80 y 95% de bandas de credibilidad - Ocurrencia de incendios forestales provocados

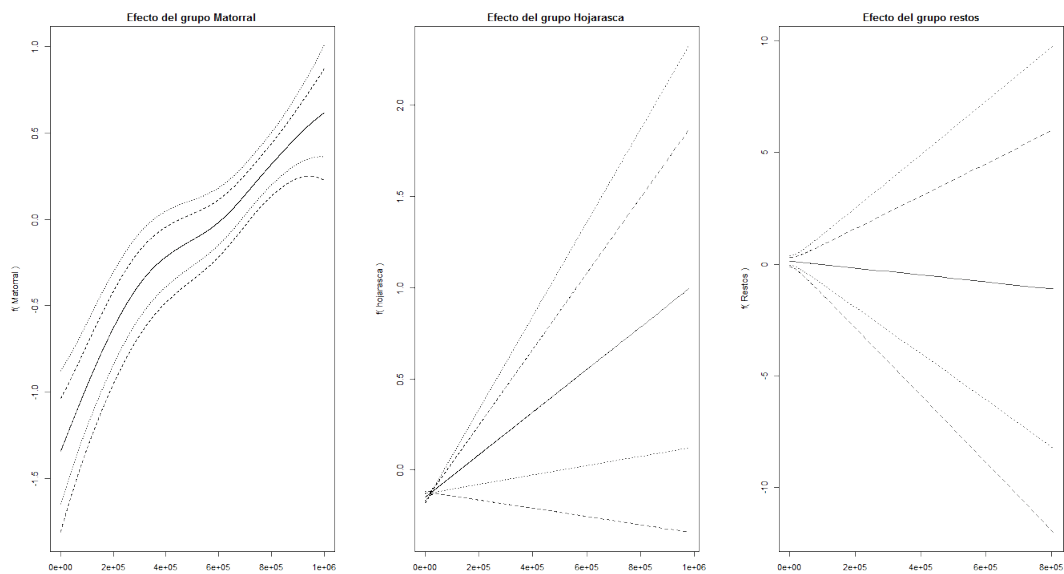


Figura 4.28: Estimación de los efectos suavizados de las covariables Matorral, Hojarasca y Restos con sus correspondientes 80 y 95% de bandas de credibilidad - Ocurrencia de incendios forestales provocados

Los efectos espaciales estructurados (ver Figura 4.29) muestran que los concejos situados al oeste presentan mayor ocurrencia de incendios forestales provocados

que los concellos situados al norte y este de Galicia (ver panel izquierdo). Los intervalos de credibilidad al 95 % se muestran en la Figura 4.29 (panel derecho) donde los concellos de color negro representan mucha probabilidad de ocurrencia de incendios frente a los concellos en blanco que representan una baja probabilidad de ocurrencia de incendios forestales provocados. Los concellos en gris son concellos no significativos frente a la variable en estudio. Aumenta el número de concellos en Galicia con alta (oeste) y baja (norete y este) probabilidad de ocurrencia de incendios provocados con respecto a los modelos estudiados en secciones anteriores.

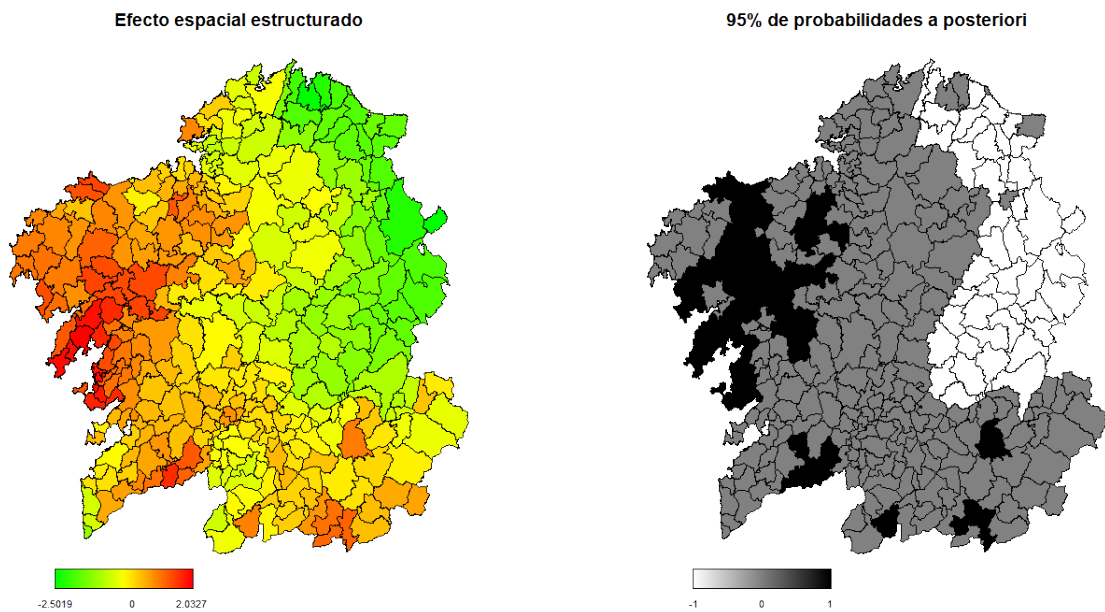


Figura 4.29: Efectos espaciales estructurados - Ocurrencia de incendios forestales provocados

Los resultados de los efectos espaciales no estructurados (Figura 4.30) no apuntan en ninguna dirección específica, Todos los concellos son no significativos (color gris, panel derecho). Los concellos pintados de blanco y rayados no aportan información de ningún tipo, ya que no se obtuvo dato para poder estimar este efecto.

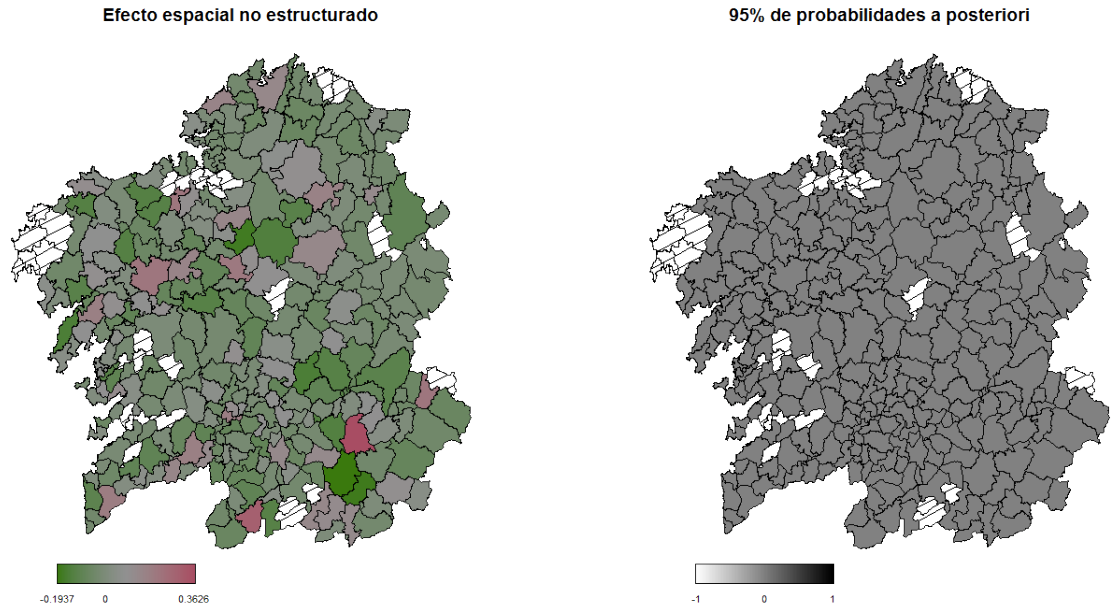


Figura 4.30: Efectos espaciales no estructurados - Ocurrencia de incendios forestales provocados

4.3.3. Comparativa de modelos

Se han estudiado los incendios forestales para dos variables respuesta, incendios totales e incendios provocados, ahora es necesario conocer el ajuste obtenido en ambas modelizaciones tal y como se muestra en el Cuadro 4.9.

Modelos	2xlog-likelihood	g.l.	AIC	BIC	GCV
Binomial (totales)	13993.4	164.997	14323.4	16065.3	0.049
Binomial (provocados)	11059	141.54	11342	12836.4	0.039

Cuadro 4.9: Criterio de información Akaike (AIC), criterio de información Bayesiano (BIC), y validación cruzada generalizada (GCV) para los dos modelos seleccionados.

Se concluye afirmando que los modelos STAR con respuesta binomial para la ocurrencia de incendios provocados y totales funcionan bien. Las variables meteorológicas pueden ser usadas para tal estimación mientras que las variables de grupos de modelos de combustibles deben de ser cambiadas por otro tipo de variables como por ejemplo, las formaciones vegetales, que aporten mayor información al modelo.

Capítulo 5

Conclusiones y Futuras Líneas de Investigación

Como resultado de la aplicación de las distintas técnicas de análisis exploratorio e inferencia de los modelos STAR en la ocurrencia de incendios forestales provocados y totales para Galicia durante la primera quincena de agosto de 2006, se han extraído las siguientes conclusiones:

1. Los modelos de Regresión Aditivos Estructurados (STAR) son adecuados para el modelado de los incendios forestales. Constituyen una aportación metodológica con ventajas sobre los clásicos modelos de regresión logística y los más actuales GLM y GAM. A pesar de ello, como toda innovación, exige un periodo de mejora para incrementar su eficiencia predictiva. Su mayor aportación 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 mapificar las predicciones en un rango de color, hace útil el modelo a efectos de planificación y gestión de incendios.
2. El problema modelizado en este trabajo es complejo. Los incendios forestales dependen de gran cantidad de variables: meteorológicas, fisiográficas, uso del suelo y relación de este con la probabilidad de arder (todas estas incorporadas en nuestros modelos), pero también de otras de tipo social, socioeconómico, histórico, conflictividad, etc., que presentan un comportamiento variable en el espacio y el tiempo, y que determina en gran medida la ignición de incendios o no incendios. Los desafíos futuros para la mejora de los modelos STAR y su aplicación a predicción de incendios está en incorporar estas variables a la escala de vóxel aquí utilizada.

3. Trabajamos con máximo rigor, pero somos conscientes de las limitaciones que presentan nuestros datos. Se observó que en algunos casos los efectos estimados de algunas covariables como altitud, precipitaciones, grupos de pastos, grupos de hojarasca y restos de cortas selvícolas en el tipo de combustible forestal en superficie eran perfectamente lineales.
4. Mejoras en la computación de los datos y en la metodología constituyen nuestra línea de investigación futura. La colaboración existente entre el grupo de investigación GRIDECMB de la USC y Thomas Kneib, nos ayudará a mejorar y poder conseguir modelos espacio-temporales flexibles que puedan proporcionar información esencial para poder predecir de forma más satisfactoria los incendios forestales.

Apéndice A

Base de datos

La base de datos analizada en este trabajo recoge los incendios registrados en Galicia, en el periodo que engloba la primera quincena de agosto de 2006. El total de incendios registrados en estos 15 días es de 2060, de los cuales 1553 son incendios provocados. Se estudian 9 variables, clasificadas en dos tipos:

A.1. Variables Meteorológicas

Las variables meteorológicas juegan un papel muy importante en la ocurrencia y transcurso de los incendios forestales, por ello es de vital importancia la consideración de ellas a la hora de estudiar los incendios. La obtención de estos datos se realiza a través de las estaciones meteorológicas ubicadas por todo el territorio gallego. Durante la primera quincena de agosto de 2006, se contó con un total de 71 estaciones meteorológicas operativas. La Figura A.1 representa en el mapa de Galicia la distribución de los lugares en los que se obtuvo y no se obtuvo datos de las para todas las variables meteorológicas de estudio.

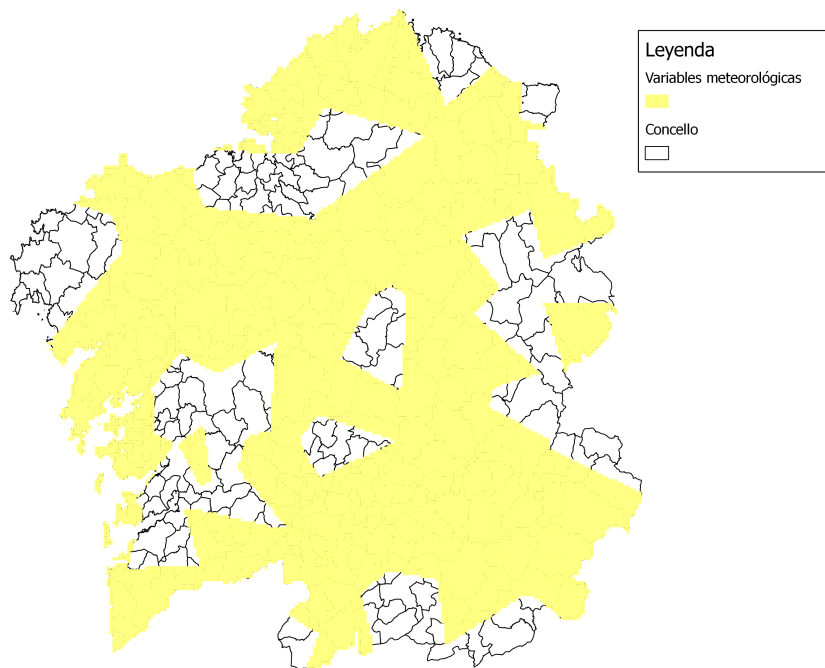


Figura A.1: Distribución de todas las variables meteorológicas en las que se obtuvo dato para la primera quincena de agosto de 2006.

Como era de esperar, la Figura A.1 muestra que no tenemos datos registrados para todas las variables. Esto se debe a que solo hay 71 estaciones meteorológicas operativas y además, la distancia máxima considerada entre cuadrícula y estación meteorológica no será nunca superior a 30 Km.

Las variables meteorológicas a estudiar son:

A.1.1. Temperatura Media Diaria

Se estudia la Temperatura Media diaria en $^{\circ}\text{C}$ por voxel a partir del dato obtenido en cada una de las estaciones meteorológicas. En la Figura A.2 se representa la distribución de temperaturas medias en los concellos en los que se ha registrado dato y en la Figura A.3 se representa un mapa en donde se observa la media de las temperaturas a partir de los datos registrados en las estaciones meteorológicas durante los primeros 15 días de agosto de 2006.

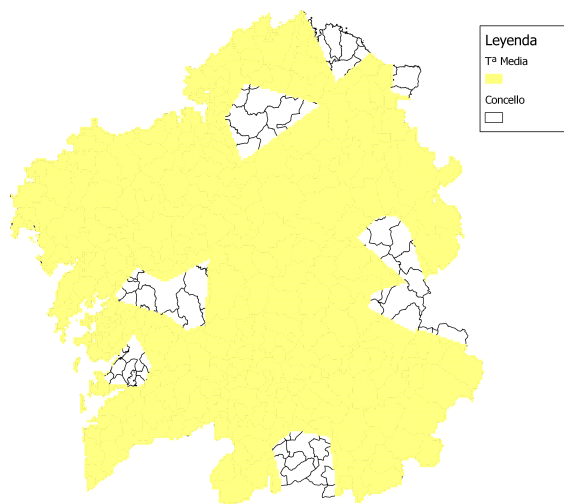


Figura A.2: Distribución de obtención de Tª Media Diaria.

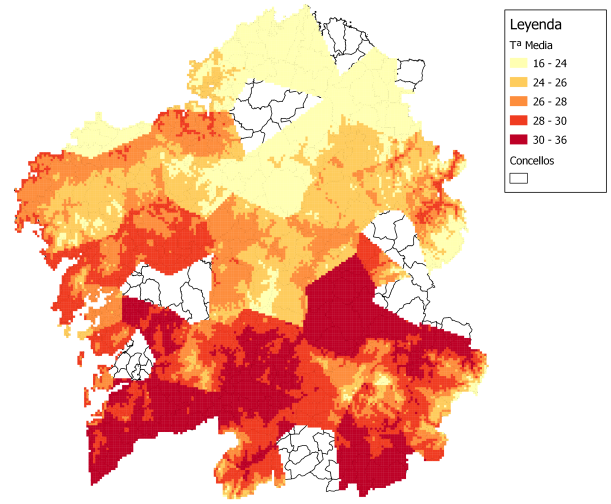


Figura A.3: Media de la Tª media diaria. Primera quincena de Agosto de 2006.

Se puede afirmar que la temperatura media registrada durante toda la primera quincena de agosto de 2006 es elevada y alcanzando valores muy altos en la mitad sur de Galicia. Los intervalos de temperaturas medias se han obtenido a partir del método de los cuantiles.

A.1.2. Humedad Relativa Media

Para la obtención de los datos de humedad relativa media medido en % se utiliza el mismo procedimiento que para las Tª medias. Uso de estaciones meteorológicas activas para la época de estudio y considerando distancias iguales o inferiores a 30 Km para registrar dato de la variable. En la Figura A.4 se representa la distribución de la humedad relativa media en los concellos en los que se ha obtenido dato y en la Figura A.5 se muestran las humedades relativas medias a partir de los datos registrados en las estaciones meteorológicas en la primera quincena de agosto de 2006.

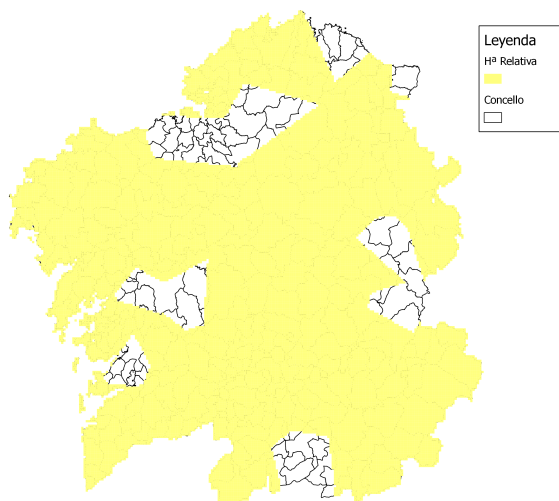


Figura A.4: Distribución de obtención de H^a Relativa Media Diaria.

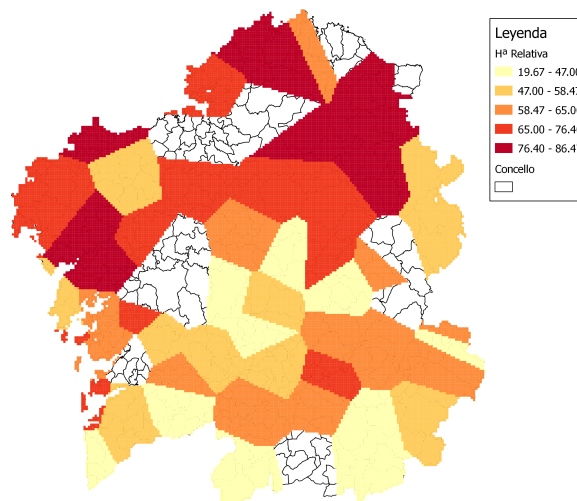


Figura A.5: Media de la H^a Relativa relativa durante la primera quincena de Agosto de 2006.

De nuevo no se obtiene dato de la humedad relativa para todo el área de estudio durante la primera quincena de agosto de 2006. Los intervalos de humedad relativa obtenidos por el método de los cuantiles en la Figura A.5 muestran humedades relativas bajas en prácticamente todo el territorio gallego, siendo algo más elevadas en la zona norte y noroeste de Galicia.

A.1.3. Precipitaciones

La última de las variables meteorológicas a estudiar en este trabajo es la precipitación registrada en l/m^2 . El procedimiento llevado a cabo para la obtención de los datos de precipitaciones a sido igual que para las variables meteorológicas anteriores. A continuación se muestra en la Figura A.6 la distribución de precipitaciones en los concellos en los que se ha registrado dato y en la Figura A.7 se muestra una media de las precipitaciones registradas en las estaciones meteorológicas en la primera quincena de agosto de 2006.

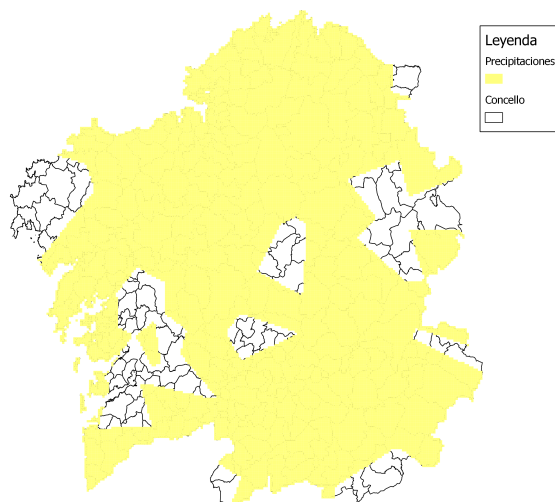


Figura A.6: Distribución de las Precipitaciones diarias.

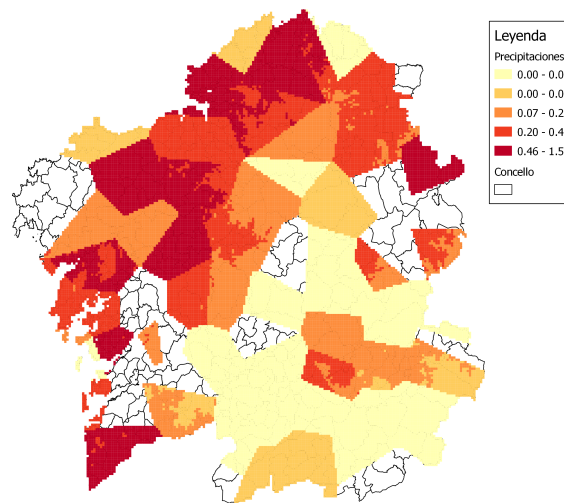


Figura A.7: Media de las precipitaciones. Primera quincena de Agosto de 2006.

Se observa en los gráficos anteriores que no se obtiene dato de la precipitación para todo el área de estudio. Cabe destacar en la Figura A.7 que los intervalos de precipitaciones considerados por el método de los cuantiles son relativamente bajos, lo que quiere decir que prácticamente hubo ausencia de precipitaciones durante toda la primera quincena de agosto de 2006.

A.2. Variables de grupos de modelos de combustible de Rothermel

La correcta caracterización del tipo de combustible forestal frente al comportamiento del fuego es un primer paso de cara a alcanzar una eficaz gestión de los incendios. Los modelos de combustible de Rothermel son una clasificación de los diferentes tipos de vegetación y su comportamiento del fuego. La clasificación de modelos de combustible considera 13 tipos de modelos de combustible, divididos en 4 grandes grupos en función de cuál sea el principal medio de propagación del fuego: pasto, matorral, hojarasca bajo arbolado y restos de corta o de tratamientos selvícolas. En este trabajo se ha incluido un quinto grupo que es la sin asignar modelo ya que existe mucha heterogeneidad en las masas gallegas y en ocasiones resulta muy difícil asignar una superficie forestal a uno de los 13 modelos de combustible propuestos.

Para la distribución de la superficie forestal gallega asociada a los grupos de modelos de combustible de Rothermel, se ha hecho uso del Cuarto Inventario Forestal Nacional (IFN-4) y de esta forma se puede conocer el tipo de combustible existente

en cada punto del territorio gallego. Los grupos de variables de estudio con su correspondiente distribución en 5 intervalos iguales de menor a mayor superficie por cuadrícula en Galicia se describen a continuación:

A.2.1. Sin Asignar Modelo

Esta variable, como ya se comentó anteriormente, se corresponde con todas aquellas áreas forestales a las que no se le asignan un modelo de combustible concreto y por consiguiente, a uno de los 4 grupos de combustibles establecidos, bien por su inmensa heterogeneidad o bien porque no se ha medido el tipo de combustible en esa zona en concreto. La Figura B.1 muestra al distribución de esta variable en Galicia.

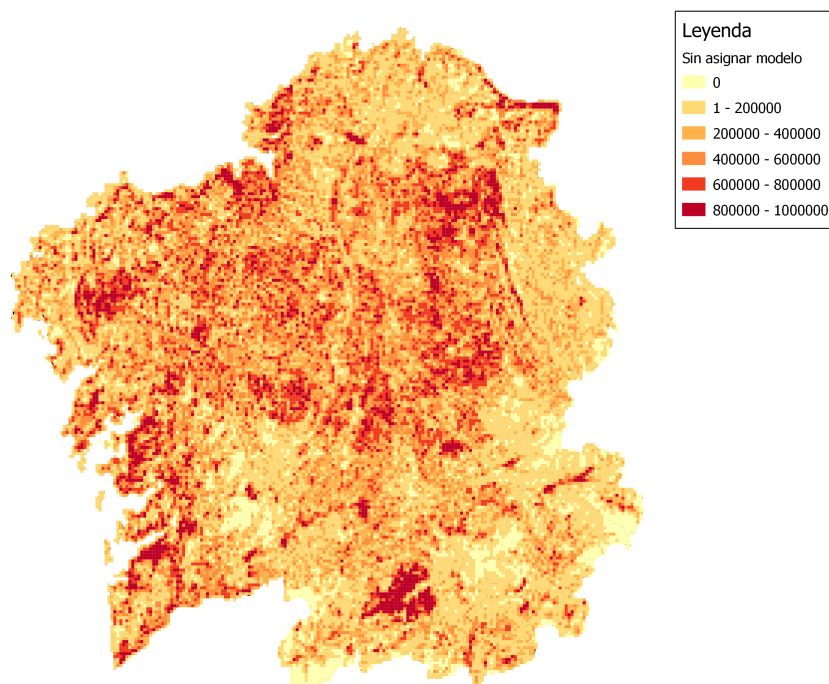


Figura A.8: Distribución de la variable Sin Asignar Modelo en Galicia

Desde el punto de vista del estudio de la ocurrencia de incendios forestales, es necesario tener en cuenta esta variable ya que como se aprecia en la Figura anterior, existen muchas zonas de superficie que no se asocian a ningún modelo de combustible y que a su vez, la heterogeneidad de esta variable nos puede ayudar a explicar un poco más la ocurrencia de los incendios en Galicia.

A.2.2. Grupo de Pastos

Los pastos son tierras utilizadas para la producción de plantas herbáceas y otros forrajes herbáceos que se distribuyen en Galicia de la siguiente manera:

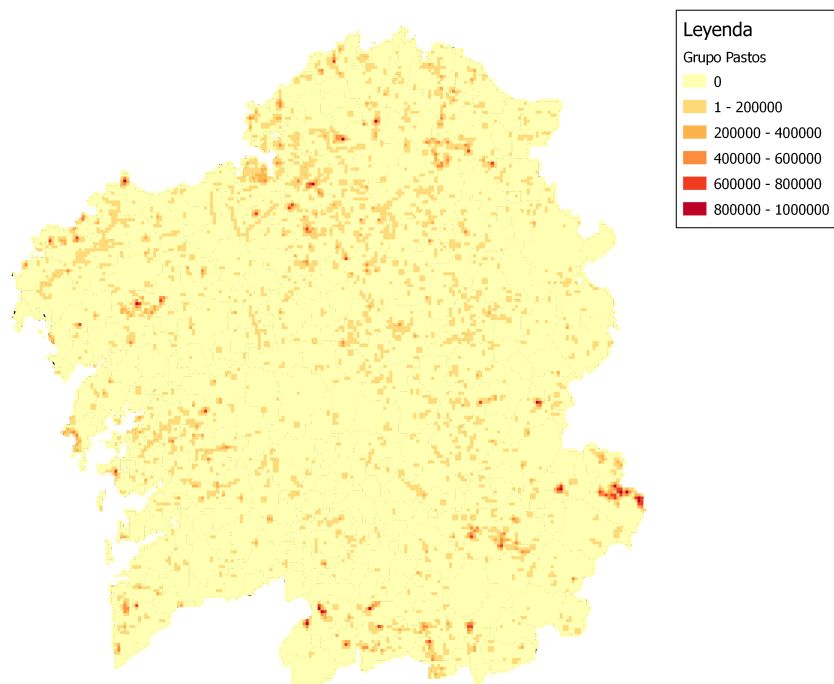


Figura A.9: Distribución de los Pastos en Galicia

Un pequeño porcentaje de la superficie forestal gallega está ocupada por pasto, algo que habrá que tener en consideración a la hora de modelizar está variable.

A.2.3. Grupo de Matorral

El matorral es un campo caracterizado por una vegetación dominada por arbustos. Su distribución en Galicia se muestra a continuación:

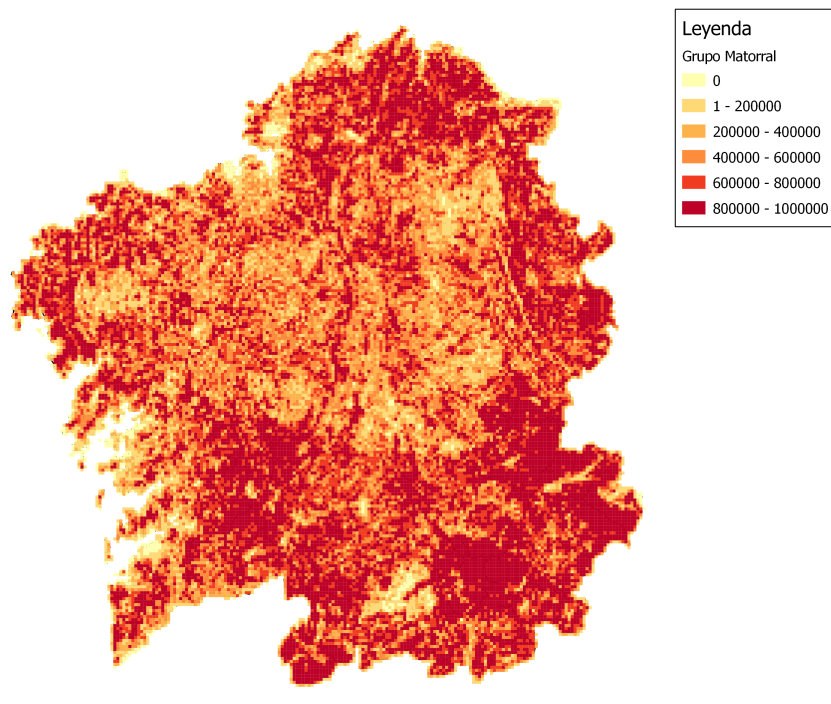


Figura A.10: Distribución del Matorral en Galicia

La mayor parte de la superficie forestal gallega está ocupada por matorral, que acumula necromasa con la edad y presenta alto riesgo de incendio forestal, aspecto muy importante a considerar a la hora de estudiar la ocurrencia de incendios. La gestión de este tipo de formaciones vegetales es necesaria con el objetivo de disminuir la posibilidad y los efectos de los incendios forestales.

A.2.4. Grupo de Hojarasca bajo arbolado

Este grupo engloba a los bosques densos sin matorral que acumulan en superficie hojarasca de los propios árboles que los forman. Su distribución es la que sigue:

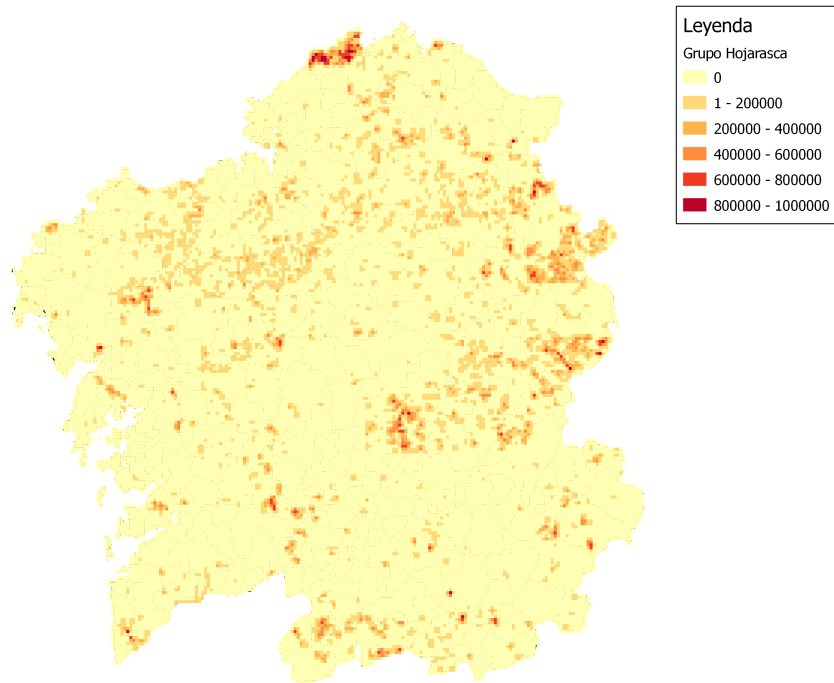


Figura A.11: Distribución de la Hojarasca bajo arbolado en Galicia

Un pequeño porcentaje de toda la superficie forestal está ocupada por este grupo de modelos de combustible.

A.2.5. Grupo de Restos y/o Operaciones Selvícolas

Este grupo está caracterizado por grandes acumulaciones de restos gruesos y pesados, cubriendo todo el suelo, provocado por claras tardías, talas rasas y cortas parciales intensas en masas maduras y sobremaduras. La distribución de este grupo de modelos de combustible es la que sigue a continuación:

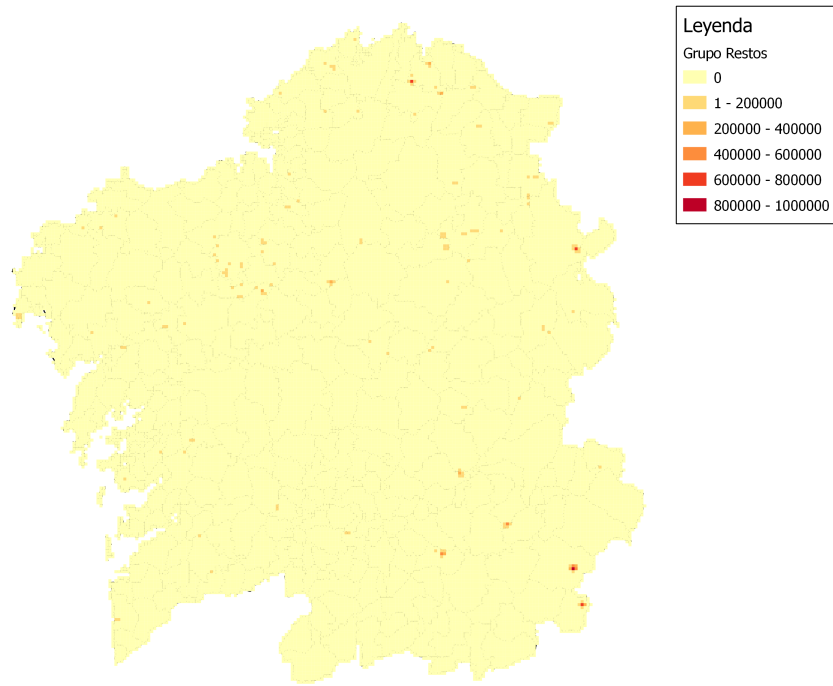


Figura A.12: Distribución de los Restos de corta y operaciones selvícolas en Galicia

Prácticamente no se aprecian existencias de este grupo de modelos de combustible en Galicia, aspecto a tener en cuenta a la hora de modelizar estas variables.

Apéndice B

Software

Para la realización de este trabajo se ha utilizado en software estadístico R. Se han utilizado las siguientes librerías:

- **BayesX:** Este paquete proporciona la funcionalidad para explorar y visualizar resultados de las estimaciones obtenidas con el paquete de software BayesX en la regresión aditiva estructurada. También proporciona funciones que permiten leer, escribir y manipular objetos de mapa que se requieren en el análisis espacial realizados con BayesX.
- **R2BayesX:** Paquete interfaz de BayesX (<http://www.BayesX.org>) de línea de comandos binarios de R.

Para programar los algoritmos que modelizan la ocurrencia de incendios forestales provocados y totales para variables meteorológicas, de tipos de modelos de combustible y ambas, se ha hecho uso de la función:

- **bayesx:** Función principal para el ajuste del modelo de la librería R2BayesX. La estimación inferencial usada en este trabajo que proporciona la función ha sido la estimación basada en el modelo de tecnología mixta y máxima verosimilitud restringida (REML).

B.1. Un ejemplo de código R para modelos STAR

B.1.1. Script de la modelización de la ocurrencia de incendios totales con variables climatológicas

```
# Cargamos las librerías

library(BayesX)
library(R2BayesX)

# Cargamos los datos: base de datos y archivo .gra

incendios1<-read.table("C:Users/nuevo_comb1.txt",header=T)
m=read.gra("C:Users/Galicia2.gra")

# Modelo climatológico (incendios totales)

f <- numero_incendios ~ sx(altitud) + sx(ta_media) + sx(hr) + sx(pp) +
sx(cdconc, bs = "mrf", map = m) + sx(cdconc, bs = "re")
b <- bayesx(f, family = "binomial", method = "REML", data = incendios1)
```

B.1.2. Script para la representación de los efectos de las covariables continuas

```
# Abrir ficheros externos que generó la modelización. Un fichero por covariable.

setwd("C:/Users/Laura/Desktop/Proyecto_master/Resultados1/Modelo_clima_todos")
res_altitud<-read.table("bayesx.estim_f_altitud_pspline.res", header=TRUE)
res_ta<-read.table("bayesx.estim_f_taIi11iImedia_pspline.res", header=TRUE)
res_hr<-read.table("bayesx.estim_f_hr_pspline.res", header=TRUE)
res_pp<-read.table("bayesx.estim_f_pp_pspline.res", header=TRUE)

# Representación de los efectos de las covariables continuas.

par(mfrow=c(1,2))
plotnonp(res_altitud, ylab= "f(altitud)", xlab="altitud", main="Efecto altitud")
plotnonp(res_ta, ylab= "f(ta_media)", xlab="ta_media", main= "Efecto T2 Media")
plotnonp(res_hr, ylab= "f(hr)", xlab="hr", main= "Efecto de la Hum. Relativa")
plotnonp(res_pp, ylab= "f(pp)", xlab="pp", main= "Efecto de la Precipitación")
```

B.1.3. Script para dibujar los efectos espaciales estructurados y no estructurados

```
# Abrir ficheros externos que generó la modelización.

res_estr <- read.table("bayesx.estim_f_cdconc_spatial.res", header=TRUE)
res_noestruc <- read.table("bayesx.estim_f_cdconc_random.res", header=TRUE)

# Cargamos el boundary

m1=read.bnd("C:/Users/Bases_de_datos/Galicia.bnd")

# Representación del efecto espacial estructurado (95\% credibilidad)

drawmap(res_estr, map=m1, regionvar="cdconc", plotvar="pmode",
swapcolors=FALSE, cols="hsv", main="Efecto espacial estructurado")
drawmap(res_estr, map=m1, regionvar="cdconc", plotvar="pcat95",
swapcolors=TRUE, cols="grey", main="95% de probabilidades a posteriori",
limits=c(-1,1))

# Representación del efecto espacial no estructurado (95\% credibilidad)

drawmap(res_noestruc, map=m1, regionvar="cdconc", plotvar="pmode",
main="Efecto espacial no estructurado")
drawmap(res_noestruc, map=m1, regionvar="cdconc", plotvar="pcat95",
swapcolors=TRUE, cols="grey",
main="95% de probabilidades a posteriori",
limits=c(-1,1))
```


Bibliografía

AKAIKE, H.,(1973). Information theory and an extension of the maximum likelihood principle. In *B. N. Petrov & F. Csaki (Eds.), 2nd International Symposium on Information Theory, pp. 267-281. Akademia Kiao, Budapest.*

ALONSO BETANZOS A., FONTELA ROMERO O., GUIJARRO BERDIÑAS B., HERNÁNDEZ PEREIRA E., PAZ ANDRADE M., JIMÉNEZ E., LEGIDO SOTO J.L. & CABALLAS T.,(2003). An intelligent system for forest fire risk prediction and fire fighting management in galicia. *Expert Systems with Applications, 25, 2003.*

ANDREWS, P.L., (1986). BEHAVE: Fire behavior prediction and fuel modeling system-BURN subsystem, Part 1, Gen. Tech. Rep. INT-194. *Ogden, UT: US Department of Agriculture, Forest Service, Intermountain Research Station, 130 p.*

BayesX Web-Seite: <http://www.BayesX.org>

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.*

BILLER, C.,(2000). Adaptive bayesian regression splines in semiparametric generalized linear models. *Journal of Computational and Graphical Statistics 9, 122-140.*

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.*

BRILLINGER, D.R., PREISLER, H.K. & BENOIT, J.W.,(2003). Risk assessment: a forest fire example. In "Science and Statistics: a Festschrift for Terry Speed". (Ed. DR Goldstein). *Lecture Notes, Monograph Series, vol. 40, pp. 177-196.*(Institute of Mathematical Statistics: Beechwood, OH.)

BRILLINGER, D.R., PREISLER, H.K. & BENOIT, J.W.,(2006). Probabilistic risk assessment for wildfires. *Environmetrics 17, 623-633. doi:10.1002/ENV.768.*

BURNHAM, K.P. & ANDERSON, D.R.,(1998).Model Selection and Multimodel Inference. *Springer.*

- CARDILLE, J.A., VENTURA, S.J. & TURNER, M.G.,(2001). Environmental and social factors influencing wildfires in the upper midwest, United States. *Ecological Applications* 11(1), 111–127. doi:10.1890/1051-0761(2001)011 [0111:EASFIW] 2.0.CO;2.
- CAVANAUGH, J.E.,(1997). Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria. *Statistics and Probability Letters* 33, 201-208.
- CHAS, M.,(2007). Forest fires in galicia (spain): Threats and challenges for the future. *Journal of Forest Economics* 13, 1 -5.
- CHEN, Z.,(1993). Fitting multivariate regression functions by interaction spline models. *Journal of the Royal Statistical Society B* 55 (2), 473-491.
- CHUVIECO, E., SALAS, F.J., CARVACHO, L. & RODRÍGUEZ SILVA, F.,(1999). Integrated fire risk mapping. In: *Chuvieco E (ed) Remote sensing of large wildfires in the European Mediterranean Basin*. Springer, Berlin, pp 61–84.
- CUNNINGHAM, A.A. & MARTELL, D.L.,(1973). A stochastic model for the occurrence of man-caused forest fires. *Canadian Journal of Forest Research* 3(2), 282–287. doi:10.1139/X73-038.
- DE BOOR, C.,(1978). A Practical Guide to Splines. *Springer, New York*.
- DÍAZ DELGADO, R., LLORET, F. & PONS, X.,(2004). Spatial patterns of fire occurrence in Catalonia, NE, Spain. *Landscape Ecology*, 19: 731- 745.
- DICKSON, B.G, JOHN W., PRATHER, J.W., YAGUANG X., HAYDEE, M., HAMPTON, H.M, AUMACK, E.N., & SISK, T.D. (2006). Mapping the probability of large fire occurrence in northern Arizona, USA. *Landscape Ecology* 21:747– 761.
- DIERCKX, P.,(1995). Curve and Surface Fitting with Splines. *Oxford: Clarendon Press*.
- EILERS, P.H. & MARX, B.D.,(1996). Flexible smoothing with B-splines and Penalties. *Statistical Science* 11, 89-121.
- FAN, J. & GIJBELS, I.,(1984). Local Polynomial Modelling and Its Applications. *Chapman & Hall, London*.
- FAHRMEIR, L. & TUTZ, G.,(2001). Multivariate Statistical Modelling Based on Generalized Linear Models. *Springer*.
- 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.
- FAHRMEIR, L., KNEIB, T. & LANG, S.,(2004). Penalized additive regression for spacetime data: a Bayesian perspective. *Statistica Sinica* 14, 731-761.

- FELICÍSIMO, A.M., (1994). Modelos digitales del terreno. Introducción y aplicaciones en las ciencias ambientales. *Pentalfa Ediciones. Disponible en <http://www.etsimo.uniovi.es/~feli/pdf/libromdt.pdf>*
- FOOD AND AGRICULTURAL ORGANIZATION OF THE UNITED NATIONS (FAO) (2007). *Fire management global assessment. A thematic study prepared in the framework of the Global Forest Resources Assessment 2005. FAO Forestry Paper 151. Rome, 320 pp. <http://www.fao.org/forestry/fra2005/en/>. Accessed 6 Dec 2010.*
- FOTHERINGHAM, A., BRUNSDON, C., & CHARLTON, M.,(2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester.*
- GALIANA MARTÍN, L., ARAGONESES DOMÍNGUEZ, C., MONTIEL MOLINA, C., DE LA CITA BENITO, F., FERNÁNDEZ RAMIRO, M., (2009). *Caracterización de los Escenarios del Fuego en España. Sociedad Española de Ciencias Forestales*
- GONZÁLEZ, J.R., PALAHI, M., TRASOBARES, A. & PUKKALA, T.,(2006). *A fire probability model for forest stands in Catalonia (north-east Spain). Annals of Forest Science 63, 169-176.*
- GONZÁLEZ, J.R. & PUKKALA, T.,(2007). *Characterization of forest fires in Catalonia (north- east Spain) European Journal of Forest Research 126, 421-429.*
- GREEN, P.J. & SILVERMAN, B.W.,(1994). *Nonparametric Regression and Generalized Linear Models. Chapman & Hall, London.*
- HARDY, C.,(2005). *Wildland re hazard and risk: Problems, denitions and context. Forest Ecology and Management 211,73 - 82.*
- HASTIE, T.J. & TIBSHIRANI, R.J.,(1990). *Generalized Additive Models. Chapman and Hall.*
- HASTIE, T.J. & TIBSHIRANI, R.J.,(1993). *Varying-coefficient Models. Journal of the Royal Statistical Society B 55, 757-796.*
- KALABOKIDIS, K.D., KOUTSIAS, N., KONSTANTINIDIS, P. & VASILAKOS, C., (2007). *Multivariate analysis of landscape wildfire dynamics in a Mediterranean ecosystem of Greece. Area 39(3), 392-402. doi:10.1111/J.1475-4762.2007.00756.X*
- KAMMANN, E. & WAND, M.,(2003). *Geoadditive models. Journal of the Royal Statistical Society C 52, 1-18.*
- KASS, R.E. & RAFTERY, A.E.,(1995). *Bayes factors. Journal of the American Statistical Association 90 (430), 773-795.*
- KNEIB, T.,(2006). *Mixed model based inference in structured additive regression. PhD thesis, Dr.Hut-Verlag.*

- KOUTSIAS, N., KALABOKIDIS, K.D. & ALLGÖWER, B.,(2004). *Fire occurrence patterns at landscape level: beyond positional accuracy of ignition points with kernel density estimation methods*. *Nat Resour Model* 17(4):359–375
- LANG, S. & BREZGER, A.,(2004). *Bayesian P-Splines*. *Journal of Computational and Graphical Statistics* 13, 183-212.
- LOADER, C.,(1999). *Local Regression and Likelihood*. Springer, New York.
- LOEHLE, C.,(2004). *Applying landscape principles to FIRE hazard reduction*. *Forest Ecology and Management*, 198: 261-267.
- MAGRAMA Web-Seite: <http://www.magrama.gob.es>
- MALAMUD, B.D., MOREIN, G. & TURCOTTE, D.L.,(1998). *Forest fires: an example of self-organized criticality*. *Science*, 281, 1840-1842.
- MAREY PÉREZ, M.F., RODRÍGUEZ VICENTE, V., & CRECENTE MASEDA, R., (2006). “Using GIS to measure changes in the temporal and spatial dynamics of forestland: experiences from north-west Spain”. *Forestry* Vol. 79: 409-423.
- MAREY PÉREZ, M.F., & RODRÍGUEZ VICENTE, V., (2008). “Forest transition in Northern Spain: Local responses on large-scale programmes of field-afforestation”. *Land Use Policy*, Article in Press.
- MARTELL, D.L. & BELIVACQUA, E.,(1989). *A Modelling seasonal variation in daily people-caused forest fire occurrence*. *Canadian Journal of Forest Research* 19(12), 1555–1563. doi:10.1139/X89-237.
- MARTELL, D.L., OTUKOL, S. & STOCKS, B.J.,(1987). *A logistic model for predicting daily people-caused forest fire occurrence in Ontario*. *Canadian Journal of Forest Research* 17(5), 394–401. doi:10.1139/X87-068.
- MARX, B.D. & EILERS, P.,(1998). *Direct generalized additive modeling with penalized likelihood*. *Computational Statistics and Data Analysis* 28 (2), 193-209.
- 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.
- MILLER, A.,(2002). *Subset Selection in Regression*. Chapman and Hall.
- MINNICH, R.A. & BAHRE, C.J.,(1995). *Wildland fire and chaparral succession along the California-baja california boundary*. *International Journal of Wildland Fire*, 5, 13-24.
- MORENO J.M.,(1989). *Los ecosistemas terrestres mediterráneos y el fuego*. *Política Científica* 18:46–50.

- MORITZ, M.A.,(2003). *Spatiotemporal analysis of controls on shrubland and fire regimes: age dependency and fire hazard. Ecology, 84, 2003.*
- 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.*
- NÚÑEZ E.A. & COLHOUN, A., (1986). *Note on air temperature lapse rates on Mount Wellington. Pap. Proc. R. Soc. Tasmania 120, pp. 11–15.*
- PEW, K.L. & LARSEN, C.P.S.,(2001). *GIS analysis of spatial and temporal patterns of human-caused wildfires in the temperate rain forest of Vancouver Island, Canada. Forest Ecology and Management 140, 1–18. doi:10.1016/S0378-1127(00)00271-1.*
- PREISLER, H.K., BRILLINGER, D.R., BURGAN, R.E. & BENOIT, J.W.,(2004). *Probability based models for estimation of wildfire risk. International Journal of Wildland Fire 13, 133–142. doi:10.1071/WF02061.*
- PREISLER, H.K. & WESTERLING, A.L.,(2007). *Statistical model for forecasting monthly large wildfire events in western United States. Journal of Applied Meteorology and Climatology 46, 1020–1030. doi:10.1175/JAM2513.1.*
- R DEVELOPMENT CORE TEAM (2007). “R: a Language and Environment for Statistical Computing”. (R Foundation for Statistical Computing:Vienna) Available at <http://www.R-project.org> [Verified 19 March 2009].
- REED, W.,(1994). *Estimating the Historic Probability of Stand-Replacement Fire Using the Age-Class Distribution of Undisturbed Forest. Forest Science. 40: 104–119.*
- REED, W.J. & MCKELVEY, K.S.,(2002). *Power-law behaviour and parametric models for the size-distribution of forest fires. Ecological Modelling, 150, 239 – 254.*
- RIERA, P. & MOGAS, J.,(2004). *Evaluation of a risk reduction in forest fires in Mediterranean region. Forest Policy and Economics 6: 521-528.*
- RIGBY, R.A. & STASINOPOULOS, D.M,(2005). *Generalized Additive Models for Location, Scale and Shape. Applied Statistics 54, 507-554.*
- ROBIN, J.G., CARREGA, P. & FOX, D.,(2006). *Modelling fire ignition in the Alpes-Maritimes Department, France. A comparison. In: Proceedings of V international conference on forest fire research, 27–30 Nov, Figueira da Foz.*
- ROTHERMEL, R.C., (1983). *How to predict the Spreads and intensity of forest and range forest fires. USDA Forest Service. USA.*
- RUPPERT, D., WAND, M.P., & CARROLL, R.J.,(2003). *Semiparametric Regression. Cambridge University Press, Cambridge.*

- SCHWARZ, G.,(1978). *Estimating the dimension of a model. The Annals of Statistics* 6, 461-464.
- TÁBARA, D., SAURÍ, D. & CERDAN, R.,(2003). *Forest Fire Risk Mangement and Public Participation in Changing Socioenvironmental Conditions: a Case Study in a Mediterranean Region. Risk Analysis* 23: 249-260.
- THOMPSON, W.A., VERTINSKY, I., SCHREIER, H. & BLACKWELL, B.A.,(2000). *Using forest fire hazard modelling in multiple use forest management planning, Forest Ecology and Management.* 134: 163-176.
- VAN WAGNER, C.E., (1987). *Development and structure of the Canadian Forest Fire Weather Index System. Canadian Forest Service, Ottawa, Ontario Forestry Technical Report n^o 35.*
- VASCONCELOS, M.P.P., SILVA, S., TOMÉ, M., ALVIM, M. & PEREIRA, J.M.C., (2001). *Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. Photogrammetric Engineering and Remote Sensing* 67(1), 73-81.
- VÉLEZ, R.,(1990). *Mediterranean forest fires: A regional perspective. Unasylva.* 162: 10-122002. *Causes of fires in the Mediterranean Basin. EFI proceedings.* 45: 35-42.
- VIEDMA, O. MORENO, J.M. & RIEIRO, I.,(2006). *Interactions between land use/land cover change forest fires and landscape structure in Sierra de Gredos (Central Spain) Environmental Conservation* 33: 212-225.
- WESTERLING, A.L., HIDALGO, H.G., CAYAN, D.R. & SWETNAM, T.W.,(2006). *Warming and earlier spring increase western US forest wildfire activity. Science* 313:940-943.
- WOOD, S.N.,(2006A). *Generalized Additive Models: An Introduction with R. Chapman and Hall.*
- YANG, J., HE H.S., SHIFLEY, S.R. & GUSTAFSON, E.J.,(2007). *Spatial patterns of modern period human-caused fire occurrence in the Missouri Ozark Highlands. For Sci* 53:1-15.