



6º CONGRESO FORESTAL ESPAÑOL

6CFE01-340

Montes: Servicios y desarrollo rural
10-14 junio 2013
Vitoria-Gasteiz



Edita: Sociedad Española de Ciencias Forestales
Vitoria-Gasteiz, 10-14 junio de 2013
ISBN: 978-84-937964-9-5
© Sociedad Española de Ciencias Forestales

Utilización de modelos semiparamétricos para la predicción de superficie quemada: Validación en Galicia

BOUBETA-MARTÍNEZ, M¹., FUENTES-SANTOS, I.^{1,2}, MAREY-PÉREZ, M.F.^{3,4},
GONZÁLEZ-MANTEIGA, W.^{1,3}

¹ GI-1914 Modelos de Optimización, decisión, estadística y aplicaciones. Departamento de Estadística e Investigación Operativa Universidade de Santiago de Compostela. Facultad de Matemáticas. Campus Vida de Santiago de Compostela. 15782.

² Instituto de Investigaciones Marinas de Vigo (IIM) – Consejo Superior de Investigaciones Científicas (CSIC).

³ Instituto de Estudos e Desenvolvimento de Galicia (IDEGA).

⁴ GI-1716 Proxectos e Planificación. Departamento de Enxeñaría Agroforestal. Universidade de Santiago de Compostela. E.P.S. Campus Universitario de Lugo. 27002. manuel.marey@usc.es

Resumen

La clase de modelos ARMA sigue siendo hoy en día, la familia paramétrica más utilizada en la modelización de las series temporales. Las razones de su éxito se deben a la generalidad, la relativa facilidad de implementación y la habilidad de proporcionar predicciones lineales óptimas. Sin embargo, existen situaciones donde las familias paramétricas no pueden ser adoptadas con seguridad. En tales situaciones los modelos no paramétricos ofrecen una alternativa idónea. Los modelos semiparamétricos que se presentan en este trabajo ofrecen una modelización alternativa de las series temporales. En ellos, se descompone la predicción en dos componentes: una no paramétrica y otra paramétrica, modelizable a través de la metodología Box-Jenkins. Su comportamiento ha resultado ser tremendamente competitivo respecto a otros modelos de predicción, tal y como se observa en la aplicación a datos reales de la superficie semanal quemada en los incendios forestales gallegos desde el año 1999 hasta el 2008.

Palabras clave

Incendios forestales, modelos semiparamétricos, superficie quemada.

1. Introducción

El fuego es un componente importante en muchos 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.

La mayoría de los incendios forestales contemporáneos son intencionados (ROMERO-CALCERRADA et al., 2008; CATRY et al., 2009 ; CHUVIECO et al., 2010 ; GONZÁLEZ-OLABARRIA et al., 2011; JUANet al., 2012). Diferentes causas se han considerado confirmando la importancia del aumento de la superficie forestal (BADIA et al., 2003 ; CATRY et al., 2009). También existen trabajos que lo relacionan con el patrón del paisaje agrario (AMATULLI et al., 2007 ; CORCORAN et al. 2007 ; KALABOKIDIS et al. 2007

;CATRY et al., 2009; MARTÍNEZ et al., 2009; FERNANDES, 2009; VERDE & ZÉZERE, 2010; MOREIRA et al. 2009 y 2011) o tradicional (VÁZQUEZ & MORENO, 1998 y 2001). El abandono agrícola (ACOSTA et al., 2005; FALCUCCI et al., 2007 ; ARANZABAL et al., 2008), asociado con el aumento de la superficie de los matorrales (DÍAZ-DELGADO et al., 2004; NUNES et al., 2005; MOUILLOT et al., 2005 ; GONZÁLEZ et al., 2006; GONZÁLEZ & PUKKALA, 2007; SEBASTIÁN-LÓPEZ et al., 2008 ; MOREIRA et al., 2009; BAJOCOCCO & RICOTTA, 2008 ; KOUTSIAS et al., 2009). Por último el desarrollo de procesos de cambios sociales y económicos (ARIANOUTSOU, 2001; CARDILLE et al., 2001; VASCONCELOS et al., 2001; BADIA-PERPINY & PALLARES-BARBERA, 2006; LOBODA & CZIZAR, 2007; ZEA-BERMÚDEZ et al. 2009 ; MARTÍNEZ et al., 2009; WITTENBERG & MALKINSON, 2009; ROMERO-CALCERRADA et al., 2010; VILAR et al., 2010; CURT & DELCROS, 2010) y también relacionado con niveles más altos de desempleo (GONZÁLEZ-OLABARRIA et al., 2011).

Diferentes enfoques han sido considerados para resolver las dificultades de modelar igniciones de origen humano en los incendios en Europa, con una metodología espacialmente explícita. Algunos estudios han intentado predecir la probabilidad de ocurrencia de incendios mediante modelos logísticos (CARDILLE et al., 2001; VASCONCELOS et al., 2001; MARTÍNEZ et al., 2004; MARTÍNEZ et al., 2009; CATRY et al., 2009; WITTENBERG & MALKINSON, 2009; CHUVIECO et al., 2010), redes neuronales artificiales (VASCONCELLO et al., 2001; VILAR DEL HOYO et al., 2007), pesos de la evidencia (AMATULLI et al., 2007 ; ROMERO-CALCERRADA et al., 2008 y 2010), estadísticas multivariantes (KALABOKIDIS et al., 2007) , regresión lineal generalizada (WOOD & AGUSTÍN, 2002), distribución de Pareto generalizada (ZEA-BERMÚDEZ et al., 2009), lógica difusa (LOBODA & CZIZAR, 2007), modelos bayesianos (VERDE & ZÉZERE, 2010), modelos de Poisson (MANDALLAZ & YE, 1997 ; WOTTON et al., 2003; BENAVENT-CORA et al., 2007), modelos aditivos generalizados (VILAR et al., 2010) , densidad kernel (DE LA RIVA et al., 2004; AMATULLI et al., 2007), simulaciones de Monte Carlo (CONEDERA et al., 2011) y Comaps (CORCORAN et al., 2007).

El análisis de las componentes espaciales y temporales de los incendios forestales han sido estudiadas en trabajos interesantes como los propuestos por LEE et al. (2006), que relacionan el área quemada y el número de incendios ocurridos en Korea entre los años 1970 y 2003, o LI et al. (2003) donde se modeliza una variable importante para la presencia de incendios, como es el índice de sequedad del suelo (SDI) a través de modelos ARMA. Autores como BECKAGE & PLATT (2003), para el caso de los incendios en Florida (EEUU), y RIAÑO et al. (2007) para los incendios en África Tropical, modelizan el área quemada por medio de modelos ARIMA. En todos estos trabajos la metodología desarrollada permite anticiparse en el tiempo y activar los sistemas de prevención y lucha contra el fuego.

En los países Mediterráneos, el fuego se ha convertido en un problema muy grave durante las tres últimas décadas y es actualmente la principal causa de destrucción de los bosques, con una media de área quemada de 500.000 hectáreas por año (VÉLEZ, 2002). Desde 1961, año en el que comienzan las estadísticas sobre incendios forestales, hasta octubre de 2011, se han producido en Galicia 245.593 incendios forestales. Estos incendios arrasaron una superficie de 1.794.578 ha, lo que equivale a un 62% del área geográfica de la Comunidad Autónoma y prácticamente el total de la superficie forestal (87%). Por ello los incendios forestales son percibidos como el principal problema medioambiental de la región (ALONSO-BETANZOS et al., 2003; CHAS, 2007).

2. Objetivos

El objetivo principal de este trabajo consiste en la implementación y validación de un modelo de análisis temporal con capacidad predictiva a un año para variables relacionadas con los incendios forestales. En este caso se ha tomado la superficie semanal quemada en los incendios forestales ocasionados en Galicia desde el año 1999 hasta el 2008, examinándose para ello tres metodologías diferentes: la clase de modelos ARIMA, por ser ésta la familia paramétrica más utilizada en la actualidad, dada su generalidad y su habilidad de proporcionar predicciones lineales óptimas, y dos modelos de predicción semiparamétricos. Finalmente se ha realizado un análisis comparativo a través de diversos criterios de error.

3. Metodología

3.1 Modelos semiparamétricos

Los modelos semiparamétricos permiten modelizar series temporales y calcular predicciones para las mismas en instantes futuros del tiempo. Estos modelos suponen una alternativa a la metodología clásica Box-Jenkins, pero a diferencia de estos últimos en los que la modelización es únicamente paramétrica, en los modelos semiparamétricos se descompone la predicción en una componente no paramétrica y otra paramétrica modelizable a través de un proceso ARMA $(p, q) \times (P, Q)_s$, expresado mediante la ecuación

$$\phi(B)\Phi(B^s)X_t = c + \theta(B)\Theta(B^s)a_t,$$

donde

$$\begin{aligned}\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \\ \Phi(B) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}, \\ \theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \\ \Theta(B) &= 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs},\end{aligned}$$

con B y B^s denotando los operadores retardo y retardo estacional, respectivamente.

3.1.1. Modelización ARMA posterior a la suavización

Sea (Z_t, Y_t) , con $t = 0, \pm 1, \pm 2, \dots$ una serie estrictamente estacionaria, donde Z_t es una serie r -dimensional e Y_t una serie respuesta unidimensional. El objetivo consiste en estimar

$$\varphi \left(\left. \cdot \right| Z_t = z_t^0 \right)$$

donde $F \left(\left. \cdot \right| Z_t = z_t^0 \right)$ denota la distribución condicional de Y_t dado $Z_t = z_t^0$, usando una serie temporal $\left(\left. \cdot \right| Y_t \right)$ de longitud n . Frecuentemente, φ es la media o mediana funcional. En

particular, cuando φ es la media funcional, $Y_t = X_{t+k}$, con $k \geq 1$ y $Z_t = (X_t, \dots, X_{t-r+1})$, siendo X_t una serie estacionaria, estamos estimando la función de autorregresión de orden k ,

$$\varphi(x_1^0, \dots, x_r^0) = E[X_{t+k} \mid X_t, \dots, X_{t-r+1} = x_1^0, \dots, x_r^0],$$

usando una muestra (X_{t-m+1}, \dots, X_t) de tamaño m .

Dada una muestra $(Z_i, Y_i)_{i=1}^m$, este enfoque permite estimar la función $\varphi(z_t^0) = E[Y_t \mid Z_t = z_t^0]$ sin realizar ninguna hipótesis paramétrica sobre ella, a través del estimador

$$\hat{\varphi}_n(z_t^0) = \sum_{i=1}^n W_{ni}(z_t^0, Z_1, Y_1, \dots, Z_n, Y_n) Y_i, \quad (1)$$

siendo (W_{ni}) una sucesión de pesos.

En YAKOWITZ (1985) se predice $Y_t = X_{t+1}$ a partir de $Z_t = X_t$, usando una muestra (X_1, \dots, X_m) de un modelo Markoviano estacionario. La función $\varphi(x) = E[X_{t+1} \mid X_t = x]$ es estimada a partir de (1) usando pesos tipo kernel,

$$W_{ni}(x, (X_1, X_2), \dots, (X_n, X_{n+1})) = \frac{K\left(\frac{x - X_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{x - X_j}{h_n}\right)}, \quad i = 1, \dots, n - m - 1, \quad (2)$$

donde K es la función kernel y h_n el parámetro ventana. Este autor propone la predicción $\hat{\varphi}_n(X_t)$ como una buena alternativa a los modelos ARMA.

Este modelo semiparamétrico (modelización ARMA posterior a la suavización), propuesto por GARCÍA JURADO et al. (1995), es presentado como una alternativa a la metodología Box-Jenkins, generalizando el procedimiento no paramétrico al semiparamétrico. Consideremos para ello el modelo

$$Y_t = \varphi(Z_t) + e_t,$$

donde e_t tiene una estructura ARMA(p, q) independiente de Z_t . El objetivo consiste en obtener la predicción de Y_t una vez observada la serie Y_t hasta el tiempo $t-k$ y Z_t hasta el tiempo t . En particular, usando la muestra $(X_{t-n+1-k}, Y_{t-n+1-k}), \dots, (X_{t-k}, Y_{t-k})$ de tamaño n , la predicción \hat{Y}_t de Y_t es definida mediante

$$\hat{\varphi}_n(Z_t) + \hat{e}_t, \quad (3)$$

donde $\hat{\varphi}_n$ es la estimación no paramétrica dada por (1), considerando por ejemplo pesos del tipo (2), y \hat{e}_t es la predicción Box-Jenkins a k retardos construida a partir de la componente ARMA estimada de la serie,

$$\hat{e}_t = Y_t - \hat{\varphi}_n(Z_t).$$

Las predicciones obtenidas con este modelo semiparamétrico suelen ser mejores que las obtenidas utilizando los modelos no paramétricos, pues en estos últimos, los residuos pueden no ser ruido blanco y en consecuencia podríamos estar dejando información sin tratar.

3.1.2. Modelización ARMA posterior a la suavización

Después de décadas de modelos de series temporales no lineales, la clase ARMA sigue siendo hoy en día, la familia paramétrica más utilizada. Las razones de su éxito se deben a la generalidad, la relativa facilidad de implementación y la habilidad de proporcionar predicciones lineales óptimas.

Si un proceso estacionario $\{X_t\}_{t \in \mathbb{Z}}$ es un modelo ARMA, su predictor lineal óptimo,

$$EL[X_t | X_u, u < t] = \sum_{i=1}^{\infty} a_i X_{t-i},$$

es obtenido a partir del modelo ARMA. Sin embargo, estos modelos también tienen importantes inconvenientes, como por ejemplo el desvanecimiento de la generalidad cuando se hacen fuertes hipótesis sobre el ruido. En tal caso, el predictor lineal óptimo puede no coincidir con el predictor óptimo,

$$E[X_t | X_u, u < t] = \phi(X_{t-1}, X_{t-2}, \dots) \quad (4)$$

En situaciones donde las familias paramétricas no pueden ser adoptadas con seguridad, los modelos no paramétricos ofrecen una alternativa idónea. La regresión tipo núcleo puede ser una elección atractiva, pues tiene por objetivo la estimación de la función de regresión del proceso X_t en sus valores pasados X_{t-1}, \dots, X_{t-d} , es decir

$$r(X_{t-1}, \dots, X_{t-d}) = E[X_t | X_{t-1}, \dots, X_{t-d}],$$

sin el requerimiento de fuertes hipótesis sobre el proceso generador de los datos.

La idea de este segundo modelo semiparamétrico (suavización posterior a la modelización ARMA), propuesto por DABO NIANG et al. (2010), consiste en utilizar los residuos ARMA como regresores en el enfoque no paramétrico. Más precisamente, se consideran dos enfoques. En el primero se usa

$$\tilde{r}(X_{t-1}, \dots, X_{t-l}, \varepsilon_{t-1}, \dots, \varepsilon_{t-m})$$

como aproximación del predictor óptimo dado en (4), donde

$$\varepsilon_t = X_t - EL[X_t | X_u, u < t]$$

denota la innovación lineal del proceso estacionario X_t . El uso de la regresión no paramétrica tiene como objetivo captar la estructura no lineal de X_t , mientras que el uso de las innovaciones lineales alivia los efectos del problema de la dimensionalidad. Dado que los ε_t 's no son observables, éstos serán reemplazados por los residuos $\hat{\varepsilon}_t$ de un modelo ARMA.

El segundo enfoque utiliza la descomposición del predictor óptimo dado en (4) como suma del predictor óptimo lineal y el predictor óptimo (no lineal) de las innovaciones del proceso

$$E[X_t | X_u, u < t] = EL[X_t | X_u, u < t] + E[\varepsilon_t | \varepsilon_u, u < t]. \quad (5)$$

La idea consiste en estimar el primer término de forma paramétrica y el segundo de forma no paramétrica. Al igual que sucedía en el caso anterior, las innovaciones no son conocidas por lo que serán reemplazadas por los residuos del modelo ARMA.

4. Resultados

Predecir la superficie semanal quemada en los incendios forestales es de vital importancia, pues permite una mejora en la gestión de los servicios forestales. Por este motivo se tratará de estimar esta cantidad a través de los modelos semiparamétricos vistos anteriormente, utilizando para ello los datos del Área de Defensa contra Incendios Forestales (ADCIF), del actual Ministerio de Agricultura Alimentación y Medio Ambiente (MAGRAMA). Este organismo es el encargado de la elaboración y publicación de la información registrada referente a estos siniestros. A partir de ella, se ha obtenido la superficie semanal quemada en Galicia para el periodo que comprende desde el año 1999 hasta el 2008. Es decir, se tiene

$$\{S_1, \dots, S_n\} \quad (6)$$

donde $n = 520$ y S_i con $i = 1, \dots, n$ refleja la superficie total quemada en la semana i .

En la Figura 1 se muestra el comportamiento de la serie temporal, donde se puede decir que es aproximadamente constante, incorporando bruscos incrementos ocasionados fundamentalmente en el periodo estival.

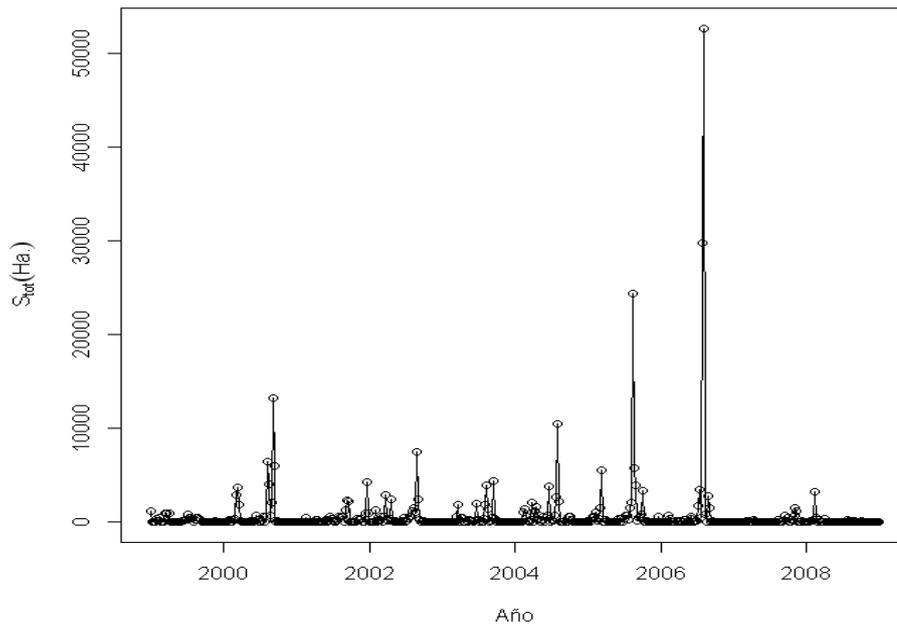


Figura 1. Gráfico secuencial de la superficie semanal quemada.

Para evaluar la calidad de las predicciones obtenidas, se ha considerado la muestra de entrenamiento

$$S_1, \dots, S_T, \text{ con } T=468, \quad (7)$$

donde se han excluido de la serie de tiempo original (6), las observaciones correspondientes al año 2008. Con esta muestra de entrenamiento será calculado el modelo de predicción y se deja la parte restante, $\{S_{T+1}, \dots, S_n\}$, para la posterior validación.

La estimación no paramétrica óptima elegida será aquella que proporcione mejores resultados entre: regresión tipo kernel, lineal local, B-Splines y P-Splines. Para hallarla se ha utilizado el siguiente algoritmo:

- i) Dada la muestra de entrenamiento de tamaño T , s_1, \dots, s_T , se considera la matriz histórica generada por esta muestra, definida por

$$MH = (s_t, s_{t+1}), \text{ con } t = 1, \dots, T-1.$$

- ii) De la matriz histórica anterior (MH), se selecciona un 75% de las observaciones al azar (matriz histórica de entrenamiento), dejando el 25% restante para la posterior validación (matriz histórica de validación).
- iii) Con la matriz histórica de entrenamiento se construyen las correspondientes estimaciones no paramétricas de la función de regresión.
- iv) Se calcula el ECM cometido sobre la matriz histórica de validación,

$$ECM = \frac{1}{K} \sum_{i=1}^K (s_i - \hat{s}_i)^2,$$

donde K denota el tamaño de la matriz histórica de validación.

v) Repetimos los pasos ii)- iv) $M = 1000$ veces.

4.1 Modelización ARMA posterior a la suavización

Para el ajuste de este primer modelo de predicción se tomará una expresión particular de (3), donde $Y_t = S_{t+1}$ y $Z_t = S_t$. Esta convención permite reformular el modelo mediante

$$\hat{S}_{t+1} = \hat{\varphi}(S_t) + \hat{e}_t. \quad (8)$$

Así, se estimará la componente no paramétrica anterior utilizando la matriz histórica asociada a la muestra de entrenamiento (7). Dado que ésta sugiere presencia de heterocedasticidad, se propone una transformación logarítmica para estabilizar la variabilidad (véase Figura 2).

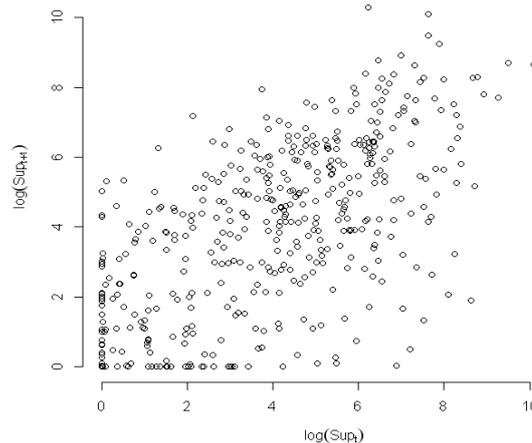


Figura 2. Matriz histórica transformada (mediante log) asociada a la distribución temporal de los incendios en Galicia (1999-2007).

A continuación se estimará la función de regresión de $\log(S_{t+1})$ sobre $\log(S_t)$,

$$m(x) = E \log(S_{t+1}) | \log(S_t) = x ,$$

mediante regresión tipo kernel, lineal local, B-Splines y P-Splines (sobre una matriz histórica de entrenamiento elegida al azar). En los dos primeros casos se utiliza como función kernel la densidad gaussiana y se selecciona el parámetro ventana usando el método de validación cruzada, mientras que en los dos últimos se eligen los grados de libertad mediante el criterio GCV. Una vez realizadas las cuatro estimaciones de la función de regresión, se toma como mejor modelo no paramétrico aquél que proporcione un menor error cuadrático medio (sobre la matriz histórica de validación), tal y como se ha explicado en el algoritmo anterior.

La Tabla 1 sugiere que el mejor estimador no paramétrico de la función de regresión es P-Splines, pues es el que mejores resultados proporciona en 486 de las $M=1000$ iteraciones. Por este motivo, se considera esta regresión como la mejor estimación no paramétrica de las utilizadas y se aplican los modelos ARMA sobre los residuos proporcionados por este modelo.

Tabla 1. Contadores del mejor estimador no paramétrico de la función de regresión obtenidos a partir del algoritmo anterior.

Análisis comparativo			
Tipo kernel	Lineal local	B-Splines	P-Splines
209	207	98	486

Además, la representación mediante boxplots (véase Figura 3), de los ECM para los cuatro estimadores permite observar como su distribución en el caso de la regresión lineal local y P-Splines se comporta de modo similar. En los casos restantes (regresión tipo kernel y B-Splines), la distribución de los ECM toma valores más altos, por lo que su uso no es aconsejable en este estudio.

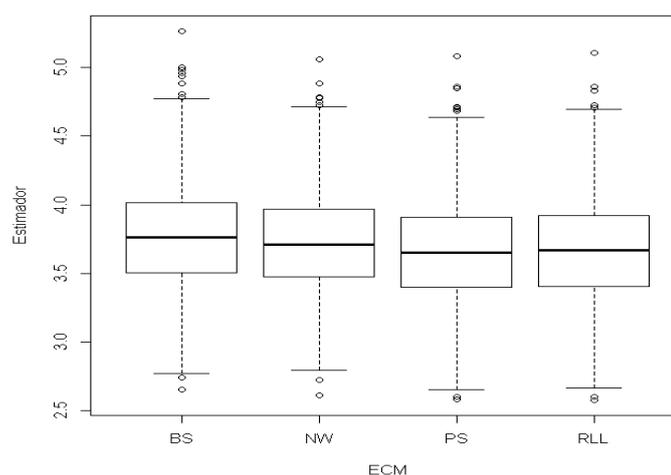


Figura 3. Boxplots de los ECM.

Seleccionado el modelo de regresión, se aplica P-Splines sobre la matriz histórica transformada mediante el logaritmo y se obtiene la serie de los residuos que será tratada a través de los modelos ARMA.

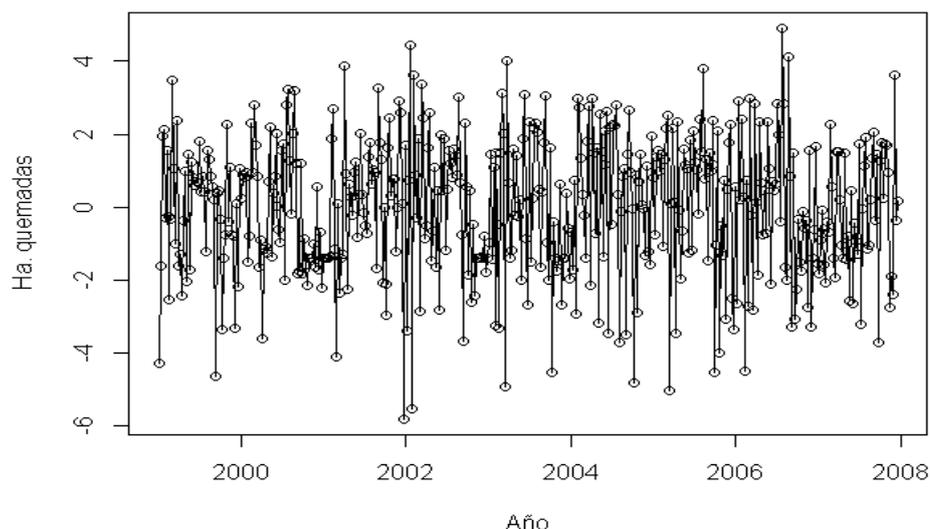


Figura 4. Gráfico secuencial de la serie temporal de los residuos (P-Splines).

A partir de la serie de los residuos, Figura 4, se tratará de descubrir algún proceso estocástico sencillo susceptible de haberla generado. Para ello se considera la clase ARMA y se selecciona aquél modelo que proporciona el criterio BIC (criterio de información bayesiana), con órdenes máximos de 10 para (p, q) y 3 para (P, Q) . El modelo óptimo propuesto según este criterio y que ha superado con éxito las pruebas de diagnóstico es

$$\text{ARMA}(9, 9) \times (2, 0)_{50},$$

con los siguientes parámetros significativos estimados.

Tabla 2. Parámetros significativos del modelo $\text{ARMA}(9, 9) \times (2, 0)_{50}$ estimados por mínimos cuadrados.

Parámetros estimados										
	ϕ_1	ϕ_3	ϕ_9	θ_3	θ_5	θ_7	θ_8	θ_9	Φ_1	Φ_2
Coef.	-0.094	-0.350	0.494	0.414	0.111	0.193	0.156	-0.512	0.102	0.096
s.e.	0.040	0.069	0.074	0.065	0.039	0.034	0.039	0.063	0.046	0.046

Calculada la componente no paramétrica y paramétrica de este modelo semiparamétrico se pueden obtener sus predicciones. Para ello, en este estudio se procede de acuerdo al siguiente algoritmo de predicción:

1. Observada la serie temporal S_1, \dots, S_T , se calcula:
2. Los residuos del modelo de regresión P-Splines, e_1, \dots, e_{T-1} .

3. La predicción Box-Jenkins a un retardo, \hat{e}_T .

4. La predicción del modelo semiparamétrico,

$$\hat{S}_{T+1} = \hat{m}(S_T) + \hat{e}_T.$$

5. El residuo del modelo,

$$res = S_{T+1} - \hat{S}_{T+1}.$$

6. Finalmente se actualizan los valores de la serie, S_1, \dots, S_T, S_{T+1} ,

y se repite tantas veces como instantes se quieran predecir (en este caso, se desean obtener las predicciones del año 2008, por lo que se realizará durante $n - T = 52$ semanas. Una vez hecho esto, se pueden comparar los valores obtenidos por el modelo (8) y los futuros valores observados de la serie (muestra de validación), sin más que deshacer los cambios hechos anteriormente aplicando una transformación exponencial.

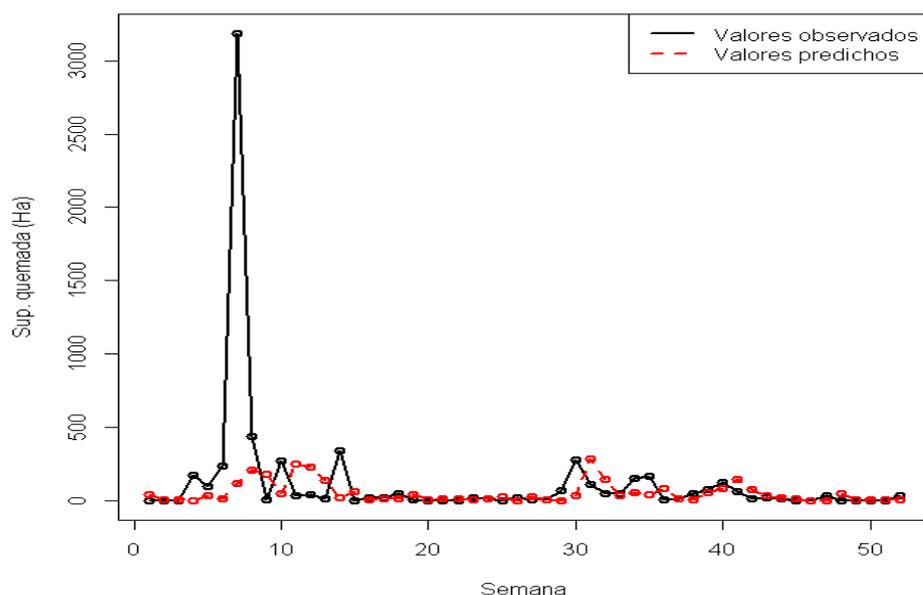


Figura 5. Predicciones proporcionadas por el modelo semiparamétrico (8) y futuros valores de la serie observados.

En la Figura 5 se muestran las predicciones de la superficie semanal quemada en los incendios forestales gallegos proporcionadas por (8), y como puede observarse éstas son razonablemente buenas. Finalmente se calcula el error cometido por este modelo a través de diversos criterios.

Tabla 3. Tabla de errores: Error Cuadrático Medio, Error Absoluto, Error Relativo Cuadrático y Error Relativo Absoluto.

Criterios de error			
ECM	EA	ERC	ERA
179283.7	113.688	4882.217	13.454

4.2. Suavización posterior a la modelización ARMA

Para el ajuste de este segundo modelo (5), se utilizará la expresión

$$\hat{X}_{T+1}^M = \hat{X}_{T+1}^L + \tilde{r}(\hat{\varepsilon}_T), \quad (9)$$

donde la primera componente se estima mediante técnicas paramétricas (metodología Box-Jenkins), mientras que la segunda se lleva a cabo de forma no paramétrica.

Para estimar la componente paramétrica del modelo se considera la serie temporal (7), pero dado que ésta presenta variabilidad no constante se ha decidido aplicar una transformación logarítmica, Figura 6.

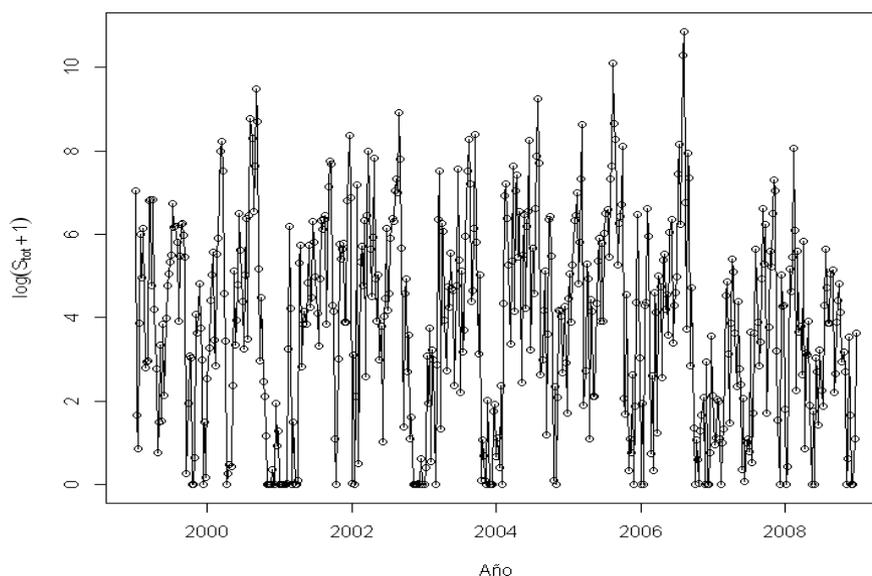


Figura 6. Gráfico secuencial de la serie temporal transformada mediante el logaritmo.

Es decir, se utiliza la metodología Box-Jenkins sobre

$$\log(S_1 + 1), \dots, \log(S_T + 1), \quad \text{con } T = 468.$$

Para la selección del modelo ARMA óptimo se aplica el criterio BIC, con órdenes máximos de 3 para p y q , obteniendo un $AR(1)$ con parámetros estimados dados en la Tabla 4.

Tabla 4: Parámetros del modelo AR(1) estimados por mínimos cuadrados.

Parámetros estimados		
	ϕ_1	c
coef.	0.644	3.917
s.e.	0.035	0.247

Este ajuste proporciona los residuos $(\varepsilon_1, \dots, \varepsilon_T)$, que serán usados para calcular la componente no paramétrica de (5). Razonando de modo análogo al primer modelo, se calcula la serie de los residuos asociada a la parte residual y se estima la función de regresión de ε_{t+1} sobre ε_t ,

$$m(x) = E \varepsilon_{t+1} | \varepsilon_t = x ,$$

mediante los cuatro estimadores no paramétricos de la función de regresión considerados: regresión tipo kernel, lineal local, B-Splines y P-Splines. La elección de la función kernel, el parámetro ventana y los grados de libertad se realiza de igual forma que en el caso anterior, tomándose como mejor modelo no paramétrico aquél que proporcione un menor error cuadrático medio (sobre la matriz histórica de validación).

En la Tabla 5 se sugiere que el mejor estimador no paramétrico de la función de regresión es P-Splines, pues es el que mejores resultados proporciona en 526 de las $M=1000$ iteraciones, por lo que se toma como la mejor estimación no paramétrica de las consideradas.

Tabla 5: Contadores del mejor estimador de la función de regresión en cada iteración.

Análisis Comparativo			
Tipo kernel	Lineal local	B-Splines	P-Splines
231	179	64	526

La Figura 7 muestra que los ECM obtenidos en la regresión lineal local incluyen más valores atípicos superiores que los demás estimadores, en consecuencia su uso no parece adecuado en este estudio. Obsérvese también como la distribución de los ECM en la regresión tipo kernel y P-Splines se comportan de modo similar.

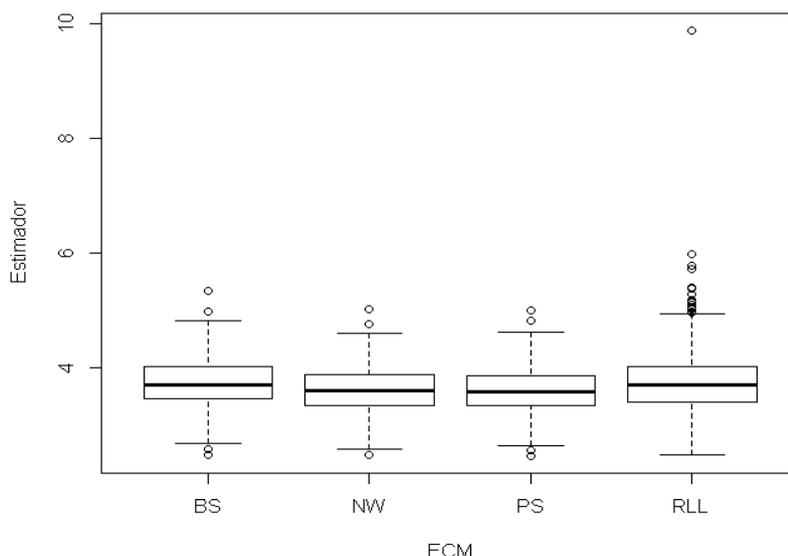


Figura 7. Boxplots de los ECM.

Una vez calculada la componente paramétrica y no paramétrica del modelo, se pueden obtener las predicciones proporcionadas por el mismo. Para ello, en este estudio se procede de acuerdo al siguiente algoritmo de predicción:

1. Se considera la serie temporal S_1, \dots, S_T .
2. Se ajusta el modelo $AR(1)$ y se obtiene la predicción, \hat{S}_{T+1} .
3. Dados los residuos del modelo AR, $\varepsilon_1, \dots, \varepsilon_T$, se construye la matriz histórica asociada, $(\varepsilon_t, \varepsilon_{t+1})_t$.
4. Se calcula la predicción de los residuos por P-Splines, $\hat{\varepsilon}_{T+1}$.
5. Se obtiene la predicción del modelo (5),

$$\hat{S}_{T+1}^M = \hat{S}_{T+1}^L + \hat{\varepsilon}_{T+1},$$

y el error, $err = S_{T+1} - \hat{S}_{T+1}^M$.

6. Finalmente se actualizan los valores de la serie, S_1, \dots, S_T, S_{T+1} ,

y se repite $n-T=52$ iteraciones, obteniendo así las predicciones que este modelo otorga al año 2008. Hechos los cálculos anteriores, se pueden comparar las estimaciones proporcionadas por (9) y los futuros valores observados de la serie (muestra de validación), deshaciendo previamente la transformación realizada.

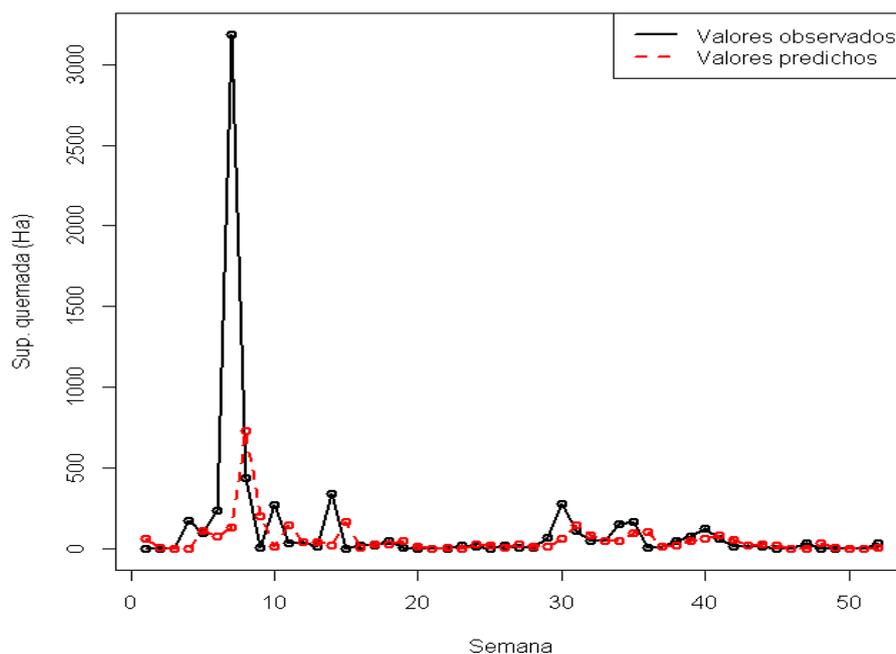


Figura 8. Predicciones del modelo (9) y futuros valores de la serie observados.

La Figura 8 muestra que las predicciones proporcionadas por este modelo y los futuros valores de la serie son bastante similares. Finalmente, se pueden calcular las discrepancias existentes entre los valores reales y los estimados, a través de diferentes criterios.

Tabla 6: Contadores del mejor estimador de la función de regresión en cada iteración.

Criterios de error			
ECM	EA	ERC	ERA
188610.400	115.620	36040.950	32.468

4.3 Análisis comparativo

Ajustados los dos modelos semiparamétricos, se puede realizar un análisis comparativo entre ambos con el fin de estudiar cuál de ellos proporciona mejores predicciones. Para ello se utilizarán los cuatro tipos de errores considerados, y además, se toma como modelo referencia el puramente paramétrico (Box-Jenkins), que ha resultado ser un $ARIMA(3,0,0) \times (2,1,3)_{52}$.

Una apreciación visual de las tres estimaciones puede contemplarse en la Figura 9.

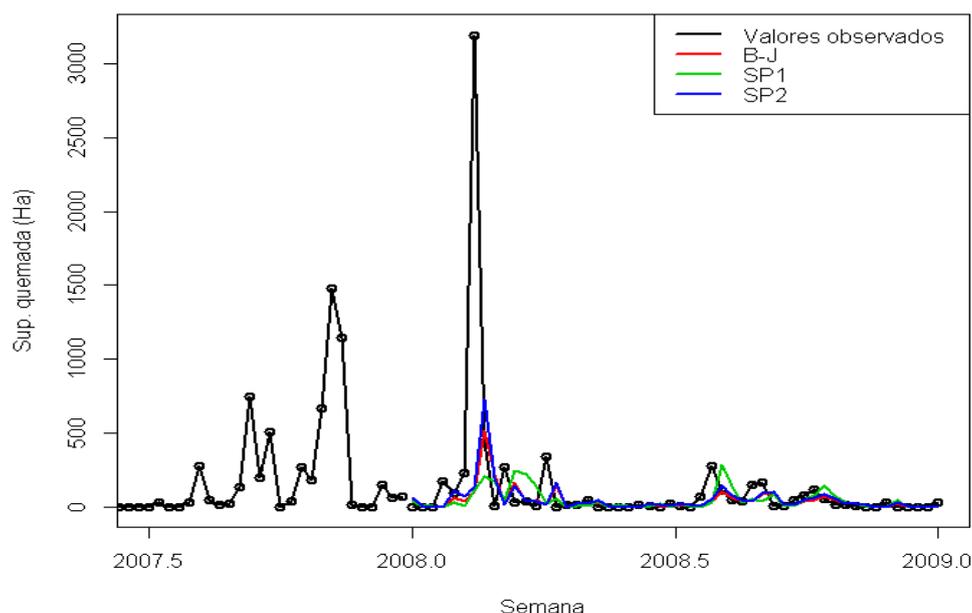


Figura 9. Predicciones obtenidas por Box-Jenkins y los dos modelos semiparamétricos.

Además, se muestra una tabla comparativa de los tres modelos de predicción, en donde se reflejan los diversos errores considerados.

Tabla 7. Tabla de errores: Error Cuadrático Medio, Error Absoluto, Error Relativo Cuadrático y Error Relativo Absoluto (para las tres predicciones).

Criterios de error				
Modelo	ECM	EA	ERC	ERA
BJ	190913.60	111.94	32120.26	30.40
SP1	179283.70	113.68	4882.21	13.45
SP2	188610.40	115.62	36040.95	32.47

La Tabla 7 sugiere que el modelo (8) es el que proporciona mejores resultados, pues es el que aporta menores errores (excepto en el Error Absoluto donde el modelo puramente paramétrico proporciona mejores resultados).

5. Conclusiones

Los incendios forestales son una de las mayores amenazas ambientales, económicas y sociales de muchas zonas del planeta. En el Sur de Europa se han convertido en el principal problema para las autoridades ambientales. En el caso de Galicia presentan una especial incidencia en determinadas zonas de la comunidad autónoma y que debido a su origen provocado suponen un desafío para el futuro de la región.

El desarrollo de metodologías de cualquier tipo, pero en especial de aquellas contrastadas por la evidencia de los datos, permitirá una más eficaz organización y planificación de la lucha contra el fuego, lo que redundará en menor número de superficie quemada y menor riesgo para vidas y bienes.

En este trabajo se ha realizado una revisión de dos modelos de predicción semiparamétricos para series temporales, los cuales descomponen la predicción en dos componentes. Una de ellas es estimada mediante técnicas de regresión no paramétrica, mientras que la otra se lleva a cabo con modelos Box-Jenkins.

El primer modelo descompone la predicción en una componente no paramétrica que estima la tendencia seguida de una estimación Box-Jenkins de los residuos, entretanto el segundo está basado en la estimación de la representación ARMA, seguida por una regresión no paramétrica para capturar la parte no lineal.

El comportamiento de estos sistemas de predicción ha resultado ser competitivo en comparación con otros modelos de predicción de series temporales, tales como los modelos no paramétricos o la metodología Box-Jenkins.

6. Bibliografía

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

ACOSTA, A.; CARRANZA, M.L.; GIANCOLA, M. 2005. Landscape change and ecosystem classification in a municipal district of a small city (Isernia, central Italy). *Environmental Monitoring and Assessment* 108, 323-335.

AMATULLI, G.; PÉREZ-CABELLO, F.; DE LA RIVA, J. 2007. Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. *Ecological Modelling* 200, 321-333.

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.

ARANZABAL, I.; SCHMITZ, M.F.; AGUILERA, P.; PINEDA, F.D. 2008. Modelling of landscape changes derived from the dynamics of socio-ecological systems: a case of study in a semiarid Mediterranean landscape. *Ecological Indicators* 8, 672-685.

ARIANOUTSOU, M. 2001. Landscape changes in Mediterranean ecosystems of Greece: implications for fire and biodiversity issues. *Journal of Mediterranean Ecology* 2, 165-178.

BADIA-PERPINYA, A.; PALLARES-BARBERA, M. 2006. Spatial distribution of ignitions in Mediterranean periurban and rural areas: the case of Catalonia. *International Journal of Wildland Fire* 15, 187-196.

BADIA, A.; PALLARÈS, M.; PIÑOL, J. 2003. Incendis periurbans front a incendis rurals. Els casos de l'Àrea Metropolitana de Barcelona i la comarca del Bages. *Documents d'Anàlisi Geogràfica* 41, 17-42.

- BAJOCCO, S.; RICOTTA, C. 2008. Evidence of selective burning in Sardinia (Italy): which land-cover classes do wildfires prefer?. *Landscape Ecology* 23, 241-248.
- BECKAGE, B., PLATT, W.J. 2003. Predicting severe wildfire years in the Florida Everglades. *Frontiers in Ecology and Environment*, 1, 235 – 239.
- CARDILLE, J.A.; VENTURA, S.J.; TURNER, M.G. 2001. Environmental and social factors influencing wildfires in the Upper Midwest, United States. *Ecol. Appl.* 11, 111–127.
- CATRY, F.X.; REGO, F.C.; BAÇAO, F.; MOREIRA, F. 2009. Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire* 18, 921–931.
- CHAS, M. 2007. Forest fires in Galicia (Spain): Threats and challenges for the future. *Journal of Forest Economics* 13, 1–5.
- CONEDERA, M.; TORRIANI, D.; NEFF, C.; RICOTTA, C.; BAJOCCHO, S.; PEZZATTI, G.B. 2011. Using Monte Carlo simulations to estimate relative fire ignition danger in a low-to-medium fire-prone region. *Forest Ecology and Management* 261, 2179-2187.
- CORCORAN, J.; HIGGS, G.; BRUNSDON, C.; ANDRE, W. 2007. The Use of Comaps to explore the spatial and temporal dynamics of fire incidents: A case study in South Wales, United Kingdom. *The Professional Geographer* 59, 521-536.
- CURT, T.; DELCROS, P. 2010. Managing road corridors to limit fire hazard. A simulation approach in southern France. *Ecological Engineering* 4, 1-12.
- DABO NIANG, S., FRANCO, C., ZAKOIAN, J.M., 2010. Combining nonparametric and optimal linear time series predictions. *JASA* 105, 1554 – 1565.
- DE LA RIVA, J.; PÉREZ-CABELLO, F.; LANA-RENAULT, N.; KOUTSIAS, N. 2004. Mapping wildfire occurrence at regional scale. *Remote Sensing of Environment* 92, 288 – 294.
- CHUVIECO, E.; ALLGÖWER, B.; SALAS, J. 2003. Integration of physical and human factors in fire danger assessment. In 'Wildland Fire Danger Estimation and Mapping. The Role of Remote Sensing Data'. (Ed. E Chuvieco) Vol. 4, pp. 197–218.
- CHUVIECO, E.; AGUADO, I., YEBRA, M.; NIETO, H.; SALAS, J.; MARTÍN, M.P.; VILAR, L.; MARTÍNEZ, J.; MARTÍN, S.; IBARRA, P.; DE LA RIVA, J.; BAEZA, J.; RODRÍGUEZ, F.; MOLINA, J.R.; HERRERA M.A.; ZAMORA, R. 2009. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modeling* 221, 46–58.
- DGCN. 2006. Los incendios forestales en España. Decenio 1996-2005, Dirección General de Conservación de la Naturaleza. MMA, Madrid.
- FALCUCCI, A.; MAIORANO, L.; BOITANI, L. 2007. Changes in land-use/land-cover patterns in Italy and their implications for biodiversity conservation. *Landscape Ecology* 22, 617-631.

FERNANDES, P. 2009. Combining forest structure data and fuel modelling to classify fire hazard in Portugal. *Annals of Forest Science* 66, 121 – 130.

GARCÍA-JURADO, I., GONZÁLEZ-MANTEIGA, W., PRADA-SÁNCHEZ, J.M., FEBRERO-BANDE, M., CAO, R. 1995. Predicting using box-jenkins, nonparametric, and bootstrap techniques. *Technometrics* 37, 303 – 310.

GONZALEZ-OLABARRIA, J.R.; MOLA-YUDEGO, B.; PUKKALA, T.; PALAHI, M. 2011. Using multiscale spatial analysis to assess fire ignition density in Catalonia, Spain. *Annals of Forest Science* 68, 861–871.

GONZALEZ, 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.

GONZALEZ, J.R.; PUKKALA, T. 2007. Characterization of forest fires in Catalonia (north-east Spain). *European Journal of Forest Research* 126, 421-429.

HARDY, C. 2005. Wildland re hazard and risk: Problems, denitions and context. *Forest Ecology and Management* 211, 73 - 82.

JUAN, P.; MATEU, J.; SAENZ, M. 2012. Pinpointing spatio-temporal interactions in wildfire patterns. *Stoch Environ Res Risk Assess* (in press).

KALABOKIDIS, K.D.; KOUTSIAS, N.; KONSTANTINIDIS, P.; VASILAKOS, C. 2007. Multivariate analysis of landscape wildfire dynamics in a Maditerranean ecosystems of Greece. *Area* 39, 392 - 402.

KOUTSIAS, N.; ARIANOUTSOU, M.; KALLIMANIS, A.S.; DIMOPOULOS, P. 2009. Is there any special pattern of the extreme wildland fires occurred in Greece in the summer of 2007? In: 52 nd International Symposium of the International Association for Vegetation Science, Vegetation Processes and Human Impact in a Changing World, Chania, Greece.

LEE, B., PARK, P. S., CHUNG, J. 2006. Temporal and spatial characteristics of foresters in South Korea between 1970 and 2003. *International Journal of Wildland Fire* 15, 389 – 396.

LI, Y., CAMPBELL, E., HASWELL, D., SNEEUWJAGT, R., AND VENABLES, W. 2003. Statistical forecasting of soil dryness index in the southwest of Western Australia. *Forest Ecology and Management* 183, 147–157.

LOBODA T.V.; CSIZAR I.A. 2007. Assessing the risk of ignition in the Russian Far East within a modeling framework of fire threat. *Ecol Appl* 17, 791–805.

MALAMUD, B.D. MOREIN, G. TURCOTTE D.L. 1998. Forest fires: an example of self-organized criticality. *Science*, 281, 1840-1842.

MARTINEZ, J.; VEGA-GARCIA, C.; CHUVIECO, E. 2009. Human-caused wildfire risk rating for prevention planning in Spain. *Journal of Environmental Management* 90, 1241-1252.

- MERMOZ, M.; KITZBERGER, T.; VEBLÉN, T.T. 2005. Landscape influences on occurrence and spread of wildfires in Patagonian forests and shrublands. *Ecology* 86, 2705–2715.
- 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.
- MOREIRA, F.; VIEDMA, O.; ARIANOUTSOU, M.; CURT, T.; KOUTSIAS, N.; RIGOLOT, E.; BARBATI, A.; CORONA, P.; VAZ, P.; XANTHOPOULOS, G.; MOUILLOT, F. & BILGILI, E. 2011. Landscape wildfire interactions in southern Europe: implications for landscape management. *Journal of Environmental Management* 92, 2389–2402.
- MOREIRA, F.; VAZ, P.; CATRY, F.; SILVA, J.S. 2009. Regional variations in wildfire susceptibility of land-cover types in Portugal: implications for landscape management to minimize fire hazard. *International Journal of Wildland Fire* 18, 563-574.
- MORITZ, M.A. 2003. Spatiotemporal analysis of controls on shrubland and fire regimes: age dependency and fire hazard. *Ecology*, 84, 2003.
- MOUILLOT, F.; RATTE, J.; JOFFRE, R.; MOUILLOT, D.; RAMBAL, S. 2005. Long-term forest dynamic after land abandonment in a fire prone Mediterranean landscape (central Corsica, France). *Landscape Ecology* 20, 101-112.
- NATIONAL INTERAGENCY FIRE CENTER. 2011. The science of wildland fire (available in: <http://www.nifc.gov/>).
- 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.
- PYNE, S.J.; ANDREWS P.L.; LAVEN R.D. 1996. Introduction to wildland fire, 2nd edn. Wiley.
- R DEVELOPMENT CORE TEAM. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Http://www.R-project.org](http://www.R-project.org).
- 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.
- RIAÑO, D., MORENO RUIZ, J.A., BARÓN MARTÍNEZ, J, USTIN, S.L. 2007. Burned area forecasting using past burned area records and southern oscillation index for tropical africa (1981-1999). *Remote Sensing of Environment* 107, 571 – 581.
- ROMERO-CALCERRADA, R.; NOVILLO, C.J.; MILLINGTON, J.D.A.; GOMEZ-JIMENEZ, I. 2008. GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (Central Spain). *Landscape Ecology* 23, 341–354.

SAN-MIGUEL-AYANZ, J.; SCHULTE, E.; SCHMUCK, G.; CAMIA, A.; STROBL, P.; LIBERTA, G.; GIOVANDO, CV BOCA, RV SEDANO, F.; KEMPENEERS, P.; MCINERNEY, DV.; WITHMORE, C.; SANTOS DE OLIVEIRA, S.; RODRIGUES, M.; DURRANT, T.; CORTI, P.; OEHLER, F.; VILAR L.; AMATULLI, G. 2012. Comprehensive monitoring of wildfires in europe: the European Forest Fire Information System (EFFIS), in John Tiefenbacher (Ed.), *Approaches to Managing Disaster - Assessing Hazards, Emergencies and Disaster Impacts*, , pp. 87-105, InTech.

SEBASTIAN-LOPEZ, A.; SALVADOR-CIVIL, R.; GONZALO-JIMENEZ, J.; SANMIGUEL-AYANZ, J. 2008. Integration of socio-economic and environmental variables for modelling long-term fire danger in southern Europe. *European Journal of Forest Research* 127, 149-163.

VÉLEZ, R. 2002. Causes of fires in the mediterranean basin. *EFI proceedings*, 45.Finlad.

VILAR DEL HOYO, L.; GÓMEZ NIETO, I.; MARTÍN ISABEL, M.P.; MARTÍNEZ VEGA, F.J. 2007. Análisis comparativo de diferentes métodos para la obtención de modelos de riesgo humano de incendios forestales. 4th International Wildland fire Conference-Wildfire 2007. Sevilla.

VILAR, L.; WOOLFORD, D.G.; MARTELL, D.L.; PILAR-MARTÍN, M. 2010. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. *International Journal of Wildland Fire* 19, 325–337.

WITTENBERG, L.; MALKINSON, D. 2009. Spatio temporal perspectives of forest fires regimes in a maturing Mediterranean mixed pine landscape. *European Journal of Forest Research* 128, 297-304.

WOOD, S.N.; AUGUSTIN, N.H. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling* 157, 157-177.

YAKOWITZ, S. J.1985. Nonparametric density estimation, prediction and regression for markov sequences. *Journal of the American Statistical Association* 339, 215-221.

ZEA-BERMUDEZ, P.; MENDES, J.; PEREIRA, J.M.C.; TURKMAN, K.F.; VASCONCELOS, M.J.P. 2009. Spatial and temporal extremes of wildfire sizes in Portugal (1984–2004). *International Journal of Wildland Fire* 18, 983–991.