

Aproximación a la modelización de la superficie afectada por los incendios forestales en Catalunya en función de variables meteorológicas e índices meteorológicos de riesgo.

F Xavier Castro, Antoni Tudela, David Montserrat, Esteve Canyameres i Eva Gabriel de Francisco
Servicio de Prevención de incendios forestales. Departament de Medi Ambient i Habitatge. Generalitat de Catalunya.

Resumen

Una de las fuentes de información más importantes para comprender de incendio forestal son los incendios ocurridos en el pasado. Son muchos los objetivos que se pueden plantear al realizar un análisis de este tipo: desde análisis de oportunidad de medidas de gestión de las causas de ignición hasta la conveniencia o no de tratamientos forestales en determinados lugares. Así mismo, son muchos y diversos los aspectos a tener en cuenta en este tipo de análisis y una forma razonable de abordar una problemática tan importante es hacer una división del problema y abordar sólo uno de sus aspectos. En este artículo nos centramos exclusivamente en la realización de diversos análisis estadísticos en relación con la superficie afectada por los incendios y su relación con diversas variables meteorológicas y diversos índices meteorológicos de riesgo de incendio forestal en Catalunya. Se comentan los problemas derivados de un análisis de este tipo y la interpretación de los resultados desde un punto de vista global. Este trabajo se ha realizado dentro del proyecto GRINFOMED+MEDIFIRE del programa Interreg 3B medocc.

1. INTRODUCCION

Una información que es relevante en la predicción del riesgo de incendio forestal es la relación entre las distintas variables que intervienen en el inicio y propagación de un incendio forestal y la superficie afectada por el mismo. Si bien esta información no es definitiva en determinación del riesgo de incendio si puede suponer una buena herramienta en la realización de cálculos de riesgo de incendio forestal en un momento dado.

2. MATERIAL Y METODOS

2.1 Series utilizadas y fuentes de datos

Las variables dinámicas de las que se dispone son variables meteorológicas y índices de vegetación derivados de variables meteorológicas.

Para realizar el trabajo, se han utilizado tres fuentes de datos: datos de incendios forestales, datos meteorológicos y índices de vegetación derivados de variables meteorológicas

a) Datos de incendios forestales

Los datos sobre incendios forestales proceden de la base de datos del Departamento de Medio Ambiente de la Generalitat de Catalunya. Para este trabajo se han considerado los

campos de fecha, termino municipal, coordenadas del incendio y superficie forestal afectada por el fuego. Se ha utilizado la base de datos del período 1996 – 2005.

b) Datos meteorológicos y índices meteorológicos de riesgo

Se han utilizado 3 fuentes de datos meteorológicos:

b.1) Datos meteorológicos de estaciones meteorológicas automáticas

De las estaciones automáticas se han obtenido los datos medios del intervalo entre las 12-13 horas solares de las distintas variables meteorológicas. También se han tratado estas variables meteorológicas para obtener los componentes del Fire Weather Index (índice meteorológico de riesgo de incendio forestal)

b.2) Datos meteorológicos de altura

Los datos meteorológicos de altura proceden del NCEP. De él se han extraído los datos medios diarios correspondientes a Catalunya para diversas variables de altura registradas a 850 hPa.

b.3) Datos meteorológicos de estaciones meteorológicas manuales

De las estaciones manuales se han obtenido los datos de pluviometría mensuales del período 1996-2005.

2.2 Variables consideradas

En concreto, se dispone información de las siguientes variables meteorológicas con su respectiva resolución.

Tabla 1 Variables disponibles para la obtención del modelo

Variable	Unidades	Resolución temporal
Fecha		
Coordenadas		
Superficie afectada	Hectáreas	
Temperatura	°C	Horaria
Humedad relativa	pct	Horaria
Velocidad viento	M/s	Horaria
Fine fuel moisture code	u.p.	Diaria
Duff moisture code	u.p.	Diaria
Drought code	u.p.	Diaria
Build up index	u.p.	Diaria
Initial spread index	u.p.	Diaria
Fire weather index	u.p.	Diaria
Altura geopotencial a 850hPa	Metros	Diaria
Temperatura a 850hPa	°C	Diaria
Humedad relativa a 850hPa	pct	Diaria
Velocidad del viento a 850hPa	M/s	Diaria
Dirección del viento a 850hPa	Grados	Diaria
Pluviometría del mes anterior al incendio respecto a la media mensual	pct	Mensual
Pluviometría de los 2 meses anteriores al incendio respecto a la media mensual	pct	Mensual
Pluviometría de los 3 meses anteriores al incendio respecto a la media mensual	pct	Mensual
Pluviometría de los 4 meses anteriores al incendio	pct	Mensual

respecto a la media mensual		
Pluviometría de los 5 meses anteriores al incendio respecto a la media mensual	pct	Mensual
Pluviometría de los 6 meses anteriores al incendio respecto a la media mensual	pct	Mensual

2.3 Asignación de las variables a los incendios

Con estas variables se ha generado una base de datos que hace corresponder los distintos valores a cada incendio. Se ha utilizado la coordenada de inicio del incendio forestal. La asignación se ha hecho de diferentes formas según la variable considerada:

- Para las variables de meteorología de altura se ha asignado la media diaria sobre Catalunya.
- Para las variables de superficie se ha asignado el dato medio de las 12-13 horas solares extrayendo el valor de una interpolación (kriging) previa construida con todas las estaciones disponibles para el territorio de Catalunya.

2.4 Método de análisis escogido

Se ha escogido el método de regresión logística binaria por ser un método flexible y que se ajusta a las necesidades de este tipo de análisis. Este tipo de análisis es un análisis robusto que utiliza variables cuantitativas y cualitativas como variables independientes y una variable dependiente dicotómica. Con él se obtiene una función lineal de las variables independientes que permite clasificar los individuos en uno de los dos grupos establecidos por los dos valores de la variable dependiente. La fórmula de probabilidad es:

$$P(Y=1)=1/(1+e^{-(b_0+b_1*X_1+b_2*X_2+\dots+b_n*X_n)})$$

Donde P es la probabilidad que el suceso tome el valor 1, Xi son las variables independientes, b0 la constante del modelo y b1,, bn son los coeficientes obtenidos en el análisis.

3. RESULTADOS

Antes de realizar el análisis estadístico se ha procedido a eliminar los registros de los que no se ha podido conseguir toda la información meteorológica de superficie necesaria. También se ha procedido a segmentar la población. Se ha procedido a separar los incendios forestales en tres grupos:

- Incendios no conducidos por el viento. Los incendios seleccionados son aquellos que cumplen tener vientos medios en superficie de 12 a 13 h ≤ 4 m/s y ≤ 6 m/s a 850 hPa y aquellos no considerados municipios de alta montaña.
- Incendios conducidos por el viento. Los incendios seleccionados son los complementarios a los incendios de no viento y que no se encuentran en municipios considerados de alta montaña
- Incendios de alta montaña. Son aquellos localizados en municipios considerados de alta montaña.

Para obtener información de la base de datos se ha realizado utilizando el análisis de regresión logística binaria.

Para ello se ha procedido a clasificar los incendios forestales según sus dimensiones para hacer 3 regresiones en función del número de hectáreas afectadas por el incendio. Las superficies escogidas para dividir la población en dos partes cada vez han sido 250 ha, 50 ha y 5 ha.

La gran mayoría de los incendios en Catalunya tiene menos de 1 hectárea. Por otro lado los incendios grandes o muy grandes son eventos raros. Es por ello que una vez obtenido el modelo se ha procedido a testarlo escogiendo al azar un 25por ciento de la población de incendios del grupo inferior y, conservando todos los del grupo superior, se ha vuelto a realizar el análisis para ver la robustez del modelo. Se comprueba que las variables escogidas con toda la población sigan siendo significativas al seleccionar una cuarta parte de ésta y que los porcentajes de clasificación sean similares.

3.1 Selección de incendios no conducidos por el viento

Este tipo de incendios se caracteriza por ocurrir en situaciones de viento habitual para la zona.

Se ha buscado una combinación de variables que clasifique el mayor número de casos posibles con un número de variables reducido. Las variables elegidas han sido la humedad relativa en superficie i el Drought Code. El resultado de la regresión logística es el siguiente:

Tabla 2. División de la población 250 ha

	Observado		Pronosticado		
	rha250		rha250 1	2	Porcentaje correcto
Paso 1	rha250	1	2827	406	87.4
		2	1	12	92.3
	Porcentaje global				87.5

a El valor de corte es .006

VARIABLES EN LA ECUACIÓN

	B	E.T.	Wald	G1	Sig.	Exp(B)
Paso 1(a) hr	-.129	.029	20.445	1	.000	.879
dc	.004	.001	11.395	1	.001	1.004
Constante	-3.903	1.211	10.390	1	.001	.020

a Variable(s) introducida(s) en el paso 1: hr, dc.

En la tabla 3 se aprecia el resultado del análisis sobre una selección de la cuarta parte de los incendios de menos de 250 ha conservando todos los de más de 250.

Tabla 3. División de la población 250 ha sobre una selección aleatoria 25por ciento de la población conservando los incendios superiores a 250ha

Tabla de clasificación(a)

	Observado		Pronosticado		
	rha250		rha250 1	2	Porcentaje correcto
Paso 1	rha250	1	691	112	86.1
		2	1	12	92.3
	Porcentaje global				86.2

a El valor de corte es .024

VARIABLES EN LA ECUACIÓN

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1(a) hr	-.118	.030	15.953	1	.000	.888
dc	.004	.001	7.514	1	.006	1.004

Constante	-2.572	1.391	3.417	1	.065	.076
-----------	--------	-------	-------	---	------	------

a Variable(s) introducida(s) en el paso 1: hr, dc.

Tabla 4. División de la población 50 ha

Tabla de clasificación(a)

Observado			Pronosticado		
			rha50		Porcentaje correcto
			1	2	
Paso 1	rha50	1	2646	572	82.2
		2	4	24	85.7
Porcentaje global					82.3

a El valor de corte es .010

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1(a) hr	-.116	.019	38.820	1	.000	.890
dc	.004	.001	15.554	1	.000	1.004
Constante	-2.753	.826	11.107	1	.001	.064

a Variable(s) introducida(s) en el paso 1: hr, dc.

Tabla 5. División de la población 5 ha

Tabla de clasificación(a)

Observado			Pronosticado		
			rha5		Porcentaje correcto
			1	2	
Paso 1	rha5	1	2166	955	69.4
		2	37	88	70.4
Porcentaje global					69.4

a El valor de corte es .040

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1(a) hr	-.067	.007	81.713	1	.000	.935
dc	.002	.000	16.279	1	.000	1.002
Constante	-1.441	.383	14.183	1	.000	.237

a Variable(s) introducida(s) en el paso 1: hr, dc.

3.2 Selección de incendios conducidos por el viento

Este tipo de incendios se caracteriza por ocurrir en situaciones de viento superior al habitual para la zona.

Se ha buscado una combinación de variables que clasifique el mayor número de casos posibles con un número de variables reducido. Las variables elegidas han sido la humedad relativa en superficie, la velocidad del viento y el Drought Code. El resultado de la regresión logística ha sido el siguiente:

Tabla 6. División de la población 250 ha.

Tabla de clasificación(a)

Observado			Pronosticado		
			rha250		Porcentaje correcto
			1	2	
Paso 1	rha250	1	2623	772	77.3
		2	3	10	76.9
Porcentaje global					77.3

a El valor de corte es .004

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1(a) hr	-.058	.026	5.093	1	.024	.943
dc	.003	.001	14.540	1	.000	1.003
vv	.401	.143	7.863	1	.005	1.493
Constante	-6.647	1.320	25.341	1	.000	.001

a Variable(s) introducida(s) en el paso 1: hr, dc, vv.

Tabla 7. División de la población 50 ha.

Tabla de clasificación(a)

Observado			Pronosticado		
			rha50		Porcentaje correcto
			1	2	
Paso 1	rha50	1	2317	1056	68.7
		2	11	24	68.6
Porcentaje global					68.7

a El valor de corte es .010

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1(a) hr	-.027	.014	3.846	1	.050	.973
dc	.002	.001	11.571	1	.001	1.002
vv	.384	.083	21.189	1	.000	1.467
Constante	-5.932	.808	53.857	1	.000	.003

a Variable(s) introducida(s) en el paso 1: hr, dc, vv.

Tabla 8. División de la población 5 ha.

Tabla de clasificación(a)

Observado			Pronosticado		
			rha5		Porcentaje correcto
			1	2	
Paso 1	rha5	1	2027	1222	62.4
		2	58	101	63.5
Porcentaje global					62.4

a El valor de corte es .046

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
--	---	------	------	----	------	--------

Paso 1(a)	hr	-.037	.007	31.558	1	.000	.963
	dc	.001	.000	18.647	1	.000	1.001
	vv	.221	.046	23.046	1	.000	1.247
	Constant						
	e	-2.892	.377	58.981	1	.000	.055

a Variable(s) introducida(s) en el paso 1: hr, dc, vv.

3.3 Selección de incendios de alta montaña

Este tipo de incendios se caracteriza por ocurrir, generalmente, durante el período invernal en localidades de alta montaña. En esta categoría el número de incendios es claramente inferior a las otras dos. Sólo se han podido analizar las categorías de 50 ha y 5 ha. No hay incendios superiores a 500 ha y el número de incendios mayores de 250 ha es reducido.

En este caso sólo se ha seleccionado la variable humedad a 850 hPa. El resultado de la regresión logística ha sido el siguiente:

Tabla 9. División de la población 50 ha

Tabla de clasificación(a)

	Observado	Pronosticado			
		rha50		Porcentaje correcto	
		1	2		
Paso 1	rha50	1	407	64	86.4
		2	1	7	87.5
	Porcentaje global				86.4

a El valor de corte es .030

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)	
Paso 1(a)	u_850	-.092	.035	6.919	1	.009	.912
	dc	-.008	.004	3.103	1	.078	.992
	Constant						
	e	-.075	.867	.007	1	.931	.928

a Variable(s) introducida(s) en el paso 1: u_850, dc.

Tabla 10. División de la población 5 ha

Tabla de clasificación(a)

	Observado	Pronosticado			
		rha5		Porcentaje correcto	
		1	2		
Paso 1	rha5	1	259	167	60.8
		2	19	34	64.2
	Porcentaje global				61.2

a El valor de corte es .110

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)	
Paso 1(a)	u_850	-.031	.010	10.712	1	.001	.969

dc	-.002	.001	3.852	1	.050	.998
Constant	-.453	.373	1.474	1	.225	.636
e						

a Variable(s) introducida(s) en el paso 1: u_850, dc.

4. DISCUSION

4.1 Incendios no conducidos por el viento

Del análisis anterior se pueden extraer diversas conclusiones:

- Con las variables escogidas (humedad relativa y Drought code) los resultados son bastante buenos pues se obtienen porcentajes relativamente altos bien clasificados.
- El número de variables escogidas es de 2, lo que hace el sistema sencillo de manejar
- Las dos variables escogidas se mantienen significativas a lo largo de todos los análisis.
- Los análisis utilizando el 25por ciento de la población (conservando los incendios del grupo superior) replican razonablemente los resultados obtenidos con toda la población

4.2 Incendios conducidos por el viento

De los resultados se pueden extraer diversas conclusiones:

- En general, los resultados no son tan buenos como en el caso de los incendios de no viento.
- Las tres variables elegidas (velocidad del viento en superficie, humedad relativa en superficie y Drought code) se mantienen significativas a lo largo de los distintos análisis al variar las hectáreas.
- Los análisis utilizando el 25por ciento de la población (conservando los incendios del grupo superior) replican razonablemente los resultados obtenidos con toda la población

4.3 Incendios de alta montaña

Los resultados en este apartado no son tan buenos como en los anteriores:

- En los dos análisis (sobre 50 y 5 ha) la variable común es la humedad relativa a 850ha.
- Las variables Drought code y velocidad del viento a 850 hPa aparecen como significativas en algunos análisis pero no se mantienen en todos. Por otro lado, al segregar el 25 por ciento de la base de datos y repetir el análisis los resultados, en general, la variable que se mantiene como significativa es la humedad a 850 hPa mientras que las otras no se mantienen en todos los modelos.

Esto es debido a:

- El bajo número de incendios en esta categoría de alta montaña limita mucho el análisis, tanto en su robustez como en la generalización de los resultados.
- La falta de estaciones automáticas en alta y media montaña. Este hecho queda reflejado en el modelo en la elección de la humedad relativa a 850 hPa posiblemente como sustituto de la humedad relativa de superficie.

Así pues, la variable que se puede aceptar como correlacionada con los incendios de alta montaña es la humedad relativa a 850 hPa

4.4 Representatividad limitada de los datos

Hay que realizar una serie de consideraciones respecto a las variables que limitan su representatividad:

- La asignación de las variables a cada uno de los incendios de ha hecho, como se ha explicado anteriormente, extrayendo el valor de una interpolación previa construida con todas las estaciones disponibles para el territorio de Catalunya con criterios de distancia. Esta interpolación no tiene en cuenta el relieve. Es por ello que el dato puede no ser representativo debido a, por ejemplo, que la estación meteorológica más cercana no sea representativa o bien que no haya una estación meteorológica razonablemente cerca. Este es el caso de, en general, de los incendios que hemos denominado de “alta montaña”.
- Las variables meteorológicas de superficie se refieren al dato medio de las 12-13 horas solares. Si bien en general es un buen dato, frecuentemente no es un dato tan definitivo como se desearía de ciertas situaciones (por ejemplo en el caso del viento) Por otro lado, la ubicación de las estaciones no es siempre la más favorable. Las estaciones automáticas se han instalado pensando en distintos criterios como la agricultura o la densidad de población. Sólo últimamente se están instalando estaciones en zonas forestales. Este hecho afecta al análisis de los datos pues los incendios forestales pueden suceder lejos de la ubicación de la estación meteorológica más cercana (o más representativa).
- Las variables meteorológicas de altura ofrecen un único dato para todo el territorio catalán limitando su utilización en la confección de cartografía de superficie. En general, frente a ellas, los modelos eligen variables de superficie. Su uso es importante fijando las condiciones de contorno y como variables suplentes.
- Algo similar pasa con las variables de pluviometría mensual. Frente a ellas, los modelos eligen un índice de sequía.

4.5 Sobre la simplicidad del modelo

Los modelos son sencillos. Esto es adecuado por varias razones:

- Casi todas las variables citadas en la tabla 1 presentan un elevado grado de correlación. La introducción de variables altamente correlacionadas implica incrementos muy pobres en la explicación de la variable dependiente a costa de incrementar el número de variables del modelo. Las variables seleccionadas presentan un grado de correlación bajo.
- Pensando en la utilización práctica de los modelos no siempre se dispone de todas las variables. Los problemas en el suministro de variables son más importantes cuanto mayor es el número de variables del modelo.
- La sencillez es una cualidad apreciable, que da robustez al modelo y facilita la interpretación.

5. CONCLUSIONES

El modelo de regresión logística binaria de incendios del tipo no viento clasifica bien el 87por ciento de los incendios mayores de 250ha, el 82por ciento de los mayores de 50ha y el 69por ciento de los incendios mayores de 5 ha.

El modelo de regresión logística binaria de incendios del tipo viento clasifica bien el 77por ciento de los incendios mayores de 250ha, el 68por ciento de los mayores de 50ha y el 62por ciento de los incendios mayores de 5 ha.

El modelo de regresión logística binaria de incendios del tipo alta montaña clasifica bien el 86por ciento de los mayores de 50ha y el 61por ciento de los incendios mayores de 5 ha.

Está previsto mejorar el modelo incorporando variables estáticas de territorio con datos de orografía y vegetación

Este trabajo se ha realizado dentro del proyecto GRINFOMED+MEDIFIRE del programa Interreg 3B medocc.

6. REFERENCIAS BIBLIOGRAFICAS

Hair y otros, 'Análisis multivariante'. Ed Prentice Hall.

Montserrat, David 'Situaciones sinópticas relacionadas con el inicio de grandes incendios forestales en Cataluña' Nimbus 1-2. 1998

Peña, Daniel 'Estadística. Modelos y métodos'. Ed. Alianza Universidad 1993

SPSS Ibérica, 'Técnicas de segmentación con SPSS'