

# Propuesta metodológica para la predicción diaria de incendios forestales

Cristina Vega-García<sup>1</sup>

## Resumen

Este estudio se basó en el análisis espacial y temporal de diversos factores geográficos y meteorológicos potencialmente relacionados con la ocurrencia de incendios forestales por agentes de ignición humanos en Cataluña, España, mediante su integración en un SIG.

Se desarrollaron varios modelos de red neuronal para la predicción diaria de incendios causados por personas, con fiabilidad suficiente para ser utilizados operativamente en la prevención y extinción de incendios en las cuadrículas de 10x10 km<sup>2</sup> que son habitualmente utilizadas por el Area de Defensa contra Incendios Forestales del Ministerio de Medio Ambiente español.

Se propone la utilización de este tipo de modelos de red neuronal artificial, cuyas características y proceso de desarrollo se discuten, a nivel nacional.

## Introducción

La predicción de incendios forestales cuenta con numerosos antecedentes basados en la modelización estadística a partir de registros históricos de incendios (Martell y otros 1987, Loftsgaarden y Andrews 1992, Chou y otros 1993, Vega-García y otros 1995, Vega-García y otros 2004). Sin embargo, las precisiones alcanzadas por estos modelos rara vez han sido suficientes para su aplicación operativa; o bien han resultado limitados en cuanto a su resolución temporal o espacial. En 2003 se ensayó operativamente un modelo logit creado para Cataluña, España, que ha sido descrito en Vega-García y otros (2004). La precisión global de este modelo era del 73 por ciento, tanto en datos de construcción como de validación, para la predicción diaria de incendios (sí/no) en las cuadrículas de 10x10 km<sup>2</sup> que son habitualmente utilizadas por el Area de Defensa contra Incendios Forestales del Ministerio de Medio Ambiente español. Sobre la misma base de datos históricos de análisis, el presente estudio exploró el potencial de varios modelos de redes neuronales artificiales para la mejora de las predicciones obtenidas.

## Métodos

El área de estudio abarcaba la superficie de la Comunidad de Cataluña que disponía de cartografía de modelos de combustible (categorizados según Vélez Muñoz, 2000) al inicio del proyecto, las provincias de Barcelona y Gerona, y partes, salvo algunas comarcas, de las provincias de Tarragona y Lérida (Figura 1).

---

<sup>1</sup> Departamento Ingeniería Agroforestal, ETSE Agraria, Universitat de Lleida, Lleida, España



**Figura 1**—Localizaciones con combustibles clasificados al inicio del proyecto: Área de predicción.

Los diversos factores geográficos y meteorológicos analizados en relación a la ocurrencia de incendios forestales por agentes de ignición humanos en Cataluña, España, fueron los siguientes (tablas 1 y 2):

**Tabla 1**—Variables geográficas de análisis para cada cuadrícula de 10x10 km<sup>2</sup>.

Denominación	Descripción	Unidades
Elev	Media de elevación de la cuadrícula a partir de MDT de 50 metros	Metros
Clase-O1,O2,O3,O4,O5	Orientaciones presentes en la cuadrícula Llano Norte Este Sur Oeste	pct de superficie ocupada por cada clase. En la fase de análisis se ha considerado la clase 1 como sur (clase 4)
Clase-P1,P2,P3,P4	Superficie ocupada según las siguientes clases de pendientes (pct) 0-5 5-15 15-30 >30	pct de superficie ocupada por cada clase
Vias1	Distancia media por cuadrícula a autopistas y carreteras nacionales	Metros
Vias2	Distancia media por cuadrícula a carreteras asfaltadas (sin considerar autopistas y carreteras nacionales)	Metros
Vias3	Distancia media por cuadrícula a carreteras no asfaltadas, considerando como tales las pistas y los caminos	Metros
Ferroc	Distancia media por cuadrícula a líneas férreas	Metros
Línelec	Distancia media por cuadrícula a líneas eléctricas	Metros

Denominación	Descripción	Unidades
Urban	Distancia media por cuadrícula a urbanizaciones	Metros
Núcleos	Distancia media por cuadrícula a núcleos urbanos	Metros
PEIN	Superficie ocupada por área PEIN-protégida	pct de superficie ocupada
Mod2,mod3,mod4,mod5,mod6,mod7,mod8,mod9,mod10	Superficie ocupada por cada modelo de combustible	pct de superficie ocupada por cada modelo

**Tabla 2**—Variables meteorológicas de análisis para cada cuadrícula de 10x10 km<sup>2</sup>.

Denominación	Descripción	Unidades
Intensidad	Intensidad de la radiación solar potencial - Volk	mm
P24	Precipitación en 24 h, diaria	Décimas de mm
Tmedia	Temperatura media diaria	Grados centígrados
HRmedia	Humedad relativa media diaria	Porcentaje, 0-100
VVmedia	Velocidad del viento media del día	m/s
DVmedia	Dirección media del viento	Grados
Tmax	Temperatura máxima diaria	Grados centígrados
HRmin	Humedad relativa mínima del día	Porcentaje, 0-100
VVmax	Velocidad máxima del viento diaria	m/s
MV	Magnitud del viento	m/s
DS	Desviación estándar del viento	Grados
RadG	Radiación solar efectiva en la estación	mj/m <sup>2</sup>
FWI	Fire Weather Index	Escala abierta, 0 a +100
KBDI	Keetch-Byram Drought Index	Escala 0-800
HCFM	Humedad del Combustible Fino muerto, Probabilidad de Ignición de ADIF-MMAM	Porcentaje, 0-100

La construcción de redes neuronales se basó en la misma base de datos de 598 observaciones utilizada en Vega-García y otros (2004), equilibrada en cuanto a observaciones diarias de incendio-no incendio en las diversas cuadrículas de 10x10 km<sup>2</sup> en Cataluña durante las campañas de verano de un periodo de 5 años (1996-2000). De las 598 observaciones se segregó un 20 por ciento (117 observaciones) para disponer de un conjunto de observaciones de validación, datos no utilizados en la construcción de los modelos.

Los modelos de red neuronal fueron simulados con NeuralSIM™ (NeuralWare 2000), mediante el enfoque constructivo desarrollado por Scott Fahlman en Carnegie Mellon University. En sus comienzos esta técnica se denominó de correlación en cascada (*Cascade Correlation*). El término cascada se refiere a la arquitectura de la red y su modo de construcción que permite añadir nodos ocultos (entre los de entrada y salida) de uno en uno y siempre conectando todos los anteriores al nuevo. El término correlación en el nombre de esta técnica hace referencia al modo en que los nodos ocultos se entrenan para maximizar la correlación entre la salida del nodo oculto y la salida deseada de la red (output) a través de los datos de entrenamiento.

La regla de aprendizaje utilizada fue el gradiente adaptativo (*Adaptative gradient learning rule*) de Falhman y Lebiere (1988).

Anteriormente al proceso de construcción de las redes se separaron parte de los datos de entrenamiento como datos de test, 141 observaciones (30 por ciento). Los datos de test fueron utilizados periódicamente durante el proceso de entrenamiento de las redes para evaluar el correcto desarrollo del proceso iterativo de construcción y asegurar la capacidad de generalización de los modelos creados. Con el objeto de conseguir redes adecuadas a la predicción operativa de incendios, también se trató de disminuir la complejidad de los modelos mediante un factor de disminución del peso de las conexiones durante el aprendizaje, y mediante la manipulación del número de variables de entrada, manual y automáticamente. Se diferenciaron varios subconjuntos de distintos tamaños a partir de las variables disponibles, como puntos de partida para la construcción de redes por correlación en cascada. El objetivo no era sólo reducir el número de variables de entrada para facilitar el entrenamiento de las redes, sino también explorar subconjuntos de variables de fácil adquisición o cómputo diario para el Área de Defensa contra Incendios Forestales, modelos simples en número de variables y en facilidad de cálculo de las variables, que fueran eficientes en caso de querer probarse operativamente en Cataluña u otra zona de características similares sin grandes disponibilidades de datos georreferenciados o de estaciones meteorológicas automáticas. El programa permitía además introducir un mecanismo de selección de variables significativas, previo al entrenamiento, y basado en un algoritmo genético, aplicable tanto a la base de datos completa como a los subconjuntos.

Entre los modelos construidos y evaluados se incluyeron modelos de una única variable basados en diferentes índices de peligro, como herramienta de comparación entre éstos.

## Resultados

Los resultados de estos análisis llevaron al desarrollo de numerosas redes que fueron evaluadas por su capacidad de predecir correctamente los datos de entrenamiento y test, pero sobre todo por su acierto en los datos de validación, no utilizados en la construcción de las redes. Los resultados obtenidos en las redes mas destacables se presentan en la Tabla 1:

**Tabla 1**—Resultados en porcentaje de aciertos de las redes construidas.

		Datos-Entren.	Datos-Test	Datos-Valid.
VARIABLES DE ENTRADA AL MODELO	VARIABLES SELECCIONADAS POR LA RED (O FORZADAS)	Precisión en pct Aciertos		
HCFM	HCFM	62.0	61.3	59.8
KBDI	KBDI	61.1	54.9	55.6
FWI	FWI	56.5	59.1	55.6
TMAX	TMAX	60.2	64.1	58.1
Todas	ELEV,PEIN,O1,P3,MOD3,MOD5,TMAX,2FWI	74.8	80.3	71.8
Todas Meteo	INTENS,2TMAX,2HRMIN,MV,2DS,KBDI	70.8	66.2	59.8
Todas Meteo+Combus+PEIN+MDT	2ELEV,PEIN,P4,MOD5,MOD6,TMAX,FWI	75.7	82.4	73.5

HCFM+PEIN+Combus	MOD4,MOD5,MOD6,MOD7,2MOD8,MOD9,HCFM	70.2	73.9	75.2
KBDI+PEIN+Combus	PEIN,MOD4,2MOD5,MOD6,2MOD7,3MOD8,MOD9,KBDI	75.1	76.8	74.4
TMAX+PEIN+ELEV+Combus	2ELEV,MOD4,2MOD5,MOD6,MOD8,MOD9,2TMAX	75.7	79.6	76.1
Todas GIS+ TMAX	ELEV,LINELECFI,PEIN,P3,MOD5,MOD6,MOD8,TMAX	74.8	79.6	74.4
PEIN+ELEV+P3+P4+Combus+TMAX +HRMIN+KBDI+HCFM	PEIN,ELEV,P3,P4,MOD4 A MOD10,TMAX,HRMIN, KBDI,HCFM	78.4	73.9	73.5
PEIN+ELEV+P3+P4+Combus+TMAX +HRMIN+KBDI+HCFM	2PEIN,2ELEV,MOD4,MOD6,2MOD7,MOD8,2TMAX	77.1	73.4	76.0

Durante el entrenamiento de las redes se comprobó que algunas variables siempre resultaban seleccionadas con preferencia a otras, que raramente entraban a formar parte del conjunto significativo en la construcción de la red. Así, ELEV, PEIN, los modelos de combustibles, TMAX, KBDI, eran favorecidas con mucha frecuencia, incluso duplicadas como entradas a la red a través de diferentes transformaciones (indicado con un 2 previo a la variable en la Tabla 1). Con menor frecuencia se seleccionaron HCFM, HRMIN, FWI, Intensitat, RadG, P3 y P4, LINELECFI y URBAFI. Prácticamente nunca formaron parte de la solución variables como FERROFI, NUCLISFI, y el resto de variables meteorológicas. Así, variables relacionadas con la topografía, existencia de zonas protegidas, combustibles, líneas eléctricas, urbanizaciones, temperatura máxima diaria, radiación solar, e índices de peligro aparecieron asociadas en distinto grado a la ocurrencia de incendios en el área de estudio.

Respecto de las redes construidas a partir de índices de peligro, HCFM resultó ser mejor variable de predicción que el resto de índices. La capacidad predictiva del Índice de Probabilidad de Ignición (HCFM) utilizado habitualmente por el Área de Defensa contra Incendios Forestales resultó aparentemente superior a la de KBDI y FWI. El modelo basado en la temperatura máxima diaria (Tmax), sin embargo, obtuvo resultados similares.

Del análisis global de todos los modelos se seleccionó como mejor red la basada en ELEV, TMAX y varios modelos de combustible (MOD4, MOD5, MOD6, MOD8, MOD9). Esta red contenía 10 nodos de entrada, 6 ocultos y dos de salida (resultado de la codificación 1 de N propia de las redes de clasificación). Era equilibrada en relación a sus aciertos en las tres bases de datos (alrededor del 78 por ciento), y dependía de variables de fácil incorporación operativa.

Esta determinación se hizo independientemente del análisis logit detallado en Vega-García y otros (2004). El mejor modelo logit incluía las variables ELEV, TMAX y los modelos de combustible de matorral, MOD4, MOD y MOD6. La semejanza en los resultados puso de manifiesto la capacidad de ambas técnicas para identificar la mejor solución al problema de la predicción de incendios en Cataluña, y el paralelismo entre ellas.

Para valorar mejor las posibilidades operativas de las redes, se comprobó la capacidad de predicción de las redes con datos reales anuales, correspondientes a todos los años del análisis. La mejor red, basada en variables de elevación, combustibles y temperatura máxima diaria, fue capaz de “acertar” un 64.7 por ciento de las observaciones en 1996 (sobre 10.174 observaciones), 69.8 por ciento en 1997 (8.611 observaciones), 59.2 por ciento en 1998 (7.603 observaciones), 66.7 por ciento en 1999 (8.639 observaciones) y un 72.4 por ciento en 2000 (1164 observaciones). La red predijo correctamente en esos años 92.3, 86.1, 77.2, 88.3, y

33.3 por ciento de los incendios. En 2000 la red predijo acertadamente sólo un incendio de tres ocurridos, igual que el mejor modelo logit, pero la baja precisión de ambos se consideró relacionada con los numerosos problemas sufridos por las estaciones meteorológicas automáticas durante ese año.

## Discusión e Implicaciones para la gestión

Los resultados indicaron que, sobre un mismo conjunto de datos históricos de incendios, el mejor modelo de red neuronal mostraba capacidades ligeramente superiores para la predicción de los incendios reales ocurridos en el territorio que el mejor modelo logit, 78 por ciento *vs* 73 por ciento de aciertos globales, es decir, tanto relativos a incendios como a no-incendios ocurridos en las unidades de predicción de 10x10 km<sup>2</sup>. Se consideró que la mejora que supuso utilizar una red para predecir incendios en este caso era aparentemente mas sustancial que cuando estos métodos se aplicaron en Canadá en 1996 (Vega-García y otros 1996). Además, dado que la mejor red dependía de un escaso número de variables, algunas de escasa variabilidad, la prueba en condiciones operativas de este modelo no debería suponer mas dificultades que las derivadas de la estimación de TMAX. Para cada día, sería necesario generar un mapa de TMAX prevista en el territorio, y la fiabilidad de la predicción para el día siguiente siempre dependería de la precisión con que se pudiera predecir la TMAX.

La aplicación de la red neuronal propuesta a los mapas de las variables seleccionadas (ELEV, MOD4, MOD5,MOD6,MOD8,MOD9 y TMAX) permitiría generar diariamente un mapa lógico de valores 0 y 1, traducible a una cartografía de zonas donde “probablemente no haya incendio hoy”, y zonas donde “probablemente haya al menos un incendio”, en las comarcas catalanas del estudio.

Respecto a las condiciones de aplicación, es necesario considerar que los análisis se centraron en la campaña de verano, que sólo incluía datos desde el 15 de mayo al 15 de septiembre de cada año en el periodo de estudio, 1996-2000. El comportamiento de los modelos fuera de estas fechas o condiciones (por ejemplo, TMAX oscilaba entre 12 y 38°C ) es desconocido.

Sería aconsejable, por tanto, actualizar en el tiempo y ampliar la utilización de este tipo de modelos de red neuronal artificial, cuyas características y proceso de desarrollo se discuten, a nivel nacional.

## Referencias bibliográficas

- Chow, Y.H., R.A. Minnich y R.A. Chase. 1993. Mapping Probability of Fire Occurrence in San Jacinto Mountains, California, USA. *Environmental Management* 17(1): 129-140.
- Fahlman S.E. and C. Lebiere. 1988. The Cascade-Correlation Learning Architecture. *Advances in Neural Information Processing Systems 2*. Morgan Kaufmann, San Francisco, CA, USA.
- Loftsgaarden, D.O. y P.L. Andrews. 1992. Constructing and testing logistic regression models for binary data: Applications to the National Fire Danger Rating System. USDA Forest Service, Intermountain Research Station, General Technical Report INT-286. 36 p.
- Martell, D. L., S. Otukol y B.J.Stocks. 1987. A logistic model for predicting daily people-caused forest fire occurrence in Ontario. *Canadian Journal of Forest Research*, 17: 394-401.
- NeuralWare 2000. NeuralSIM, the Complete Solution for neural Data Modelling. User Guide. NeuralWare, Pittsburgh, PA, USA.
- Van Wagner, C.E. 1987. Development and structure of the Canadian Forest Fire Weather Index System. Government of Canada, Canadian Forestry Service, Ottawa, Ontario, Canada. Forestry Technical Report n. 35.
- Vega-García, C., P.M. Woodard, S.J. Titus, W.L. Adamowicz y B.S. Lee. 1995. A logit model for predicting the daily occurrence of human caused forest fires. *International Journal of Wildland Fire* 5(2): 101-111.
- Vega-García, C.; P.M. Woodard; B.S.Lee y S.J.Titus.1996. Applying neural network technology to daily human-caused forest fire occurrence prediction. *AI Applications* 10(3): 9-18.
- Vega-García, C; C. Ortiz Ruiz; R. Canet Castellà; I. Sánchez Bosch y D. Queralt Creus. 2004. Aplicación operativa de un modelo diario de predicción de ocurrencia de incendios forestales por causas humanas en Cataluña. 2nd Symposium on Fire Economics, Planning and Policy: A global view. 19-22 Abril, Córdoba, España. Sociedad Española de Ciencias Forestales. CD-ROM.
- Vélez Muñoz, R. 2000. El concepto de combustibilidad. Pág. 7-1 a 7-16 en: Vélez Muñoz, R. (Coord.). 2000. La defensa contra incendios forestales. Fundamentos y Experiencias. McGraw-Hill, Madrid.