A new wildland fire danger index for a Mediterranean region and some validation aspects

Javier de Vicente[A] and Fortunato Crespo[B,C]

[A]Departamento Forestal, Vaersa, Generalitat Valenciana, Spain.

[B]Dto. de Estadística e Investigació Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia - Spain.

[C]Corresponding author. Email: fcrespo@eio.upv.es

Abstract. Wildland fires are the main cause of tree mortality in Mediterranean Europe and a major threat to Spanish forests. This paper focuses on the design and validation of a new wildland fire index especially adapted to a Mediterranean Spanish region. The index considers ignition and spread danger components. Indicators of natural and human ignition agents, historical occurrence, fuel conditions and fire spread make up the hierarchical structure of the index. Multi-criteria methods were used to incorporate experts' opinion in the process of weighting the indicators and to carry out the aggregation of components into the final index, which is used to map the probability of daily fire occurrence on a 0.5-km grid.

Generalized estimating equations models, which account for possible correlated responses, were used to validate the index, accommodating its values onto a larger scale because historical records of daily fire occurrence, which constitute the dependent variable, are referred to cells on a 10-km grid. Validation results showed good index performance, good fit of the logistic model and acceptable discrimination power. Therefore, the index will improve the ability of fire prevention services in daily allocation of resources.

Additional keywords: fire risk, ignition occurrence, generalized estimating equations, logistic regression, odds ratio

Introduction

An average of 49838 forest fires affect an area of 471644 ha each year in Mediterranean Europe (JRC 2011). Forest fires are the main cause of tree mortality in this region and one of the major threats to the Spanish forest ecosystems (Castedo-Dorado *et al.* 2011). Due to climate change, higher frequency and longer duration of extreme conditions, such as droughts, are expected to increase the risk of wildland fire and with it the demand on the resources needed to prevent and fight forest fires (Climent *et al.* 2008).

The forest fire phenomenon in Spain is mainly related to socioeconomic and meteorological factors (Vilar *et al.* 2007). Between 1998 and 2009, in the Community of Valencia (CV) (Fig. 1), the Mediterranean Spanish autonomous region where this study was carried out, each year an average of 500 fires affected 14,000 ha, 50% of which was woodland (DGMNPF 2006, 2010). Although the causes of ignition are very diverse, human activity and lightning have emerged as the main agents of fire ignition in this area (Romero-Calcerrada *et al.* 2008). Large fires, despite only accounting for 2% of the total number of fires, have been responsible for more than 90% of the affected area (DGMNPF 2006). This highlights the importance that early detection and accurate location of wildfire have on fire suppression and on prevention of large fire occurrences (Sahin and Turker 2009). Thus, fire risk assessment is a critical part in fire prevention and one of the main concerns of the Spanish Forestry Administration.

Fig. 1. Study region: forested area on a 10-km grid for the three provinces of the Community of Valencia.

Since 1994, the CV Forest Service has been using a wildland fire danger (WFD) index developed by the National Institute of Meteorology (Mediavilla *et al.* 1994) to daily deploy a mobile fleet of 97 vehicles in the forest land. This index does not take into account important risk factors related to fire occurrence, such as human activities or lightning, nor does it consider potential effects of fire. This is a common problem in Fire Danger Rating (FDR) systems, mainly because daily registers of human activities do not exist or are rarely available (Martell *et al.* 1987) and because traditional approaches do not put a strong emphasis on potential damage of fire (Chuvieco *et al.* 2010). It is therefore imperative to develop better forecasting tools to support fire prevention services in the efficient allocation of resources.

In the last decade new approaches to the concept of fire risk have been established (Bachman and Allgöwer 2001; Blanchi *et al.* 2002; Fairbrother and Turnley 2005). These include two risk components: WFD and vulnerability. WFD represents the probability a fire ignites and the potential hazard of fire propagation or spread danger (Finney, 2005), while vulnerability accounts for potential effects of fire. New FDR systems which incorporated these components can be found in Sebastián-López *et al.* (2002), Blanchi *et al.* (2002), Yebra *et al.* (2008), Chuvieco *et al.* (2010) and Verde and Zêzere (2010). According to this approach we propose a new WFD index for the CV. Indicators of ignition and spread danger have been identified, including human and natural occurrence agents, fuel conditions, historical occurrence and spread rate. These indicators make up the hierarchical structure for the index, which, following the criteria of the European Commission (San-Miguel-Ayanz *et al.* 2003) should include both short and long-term indicators.

Human variables have traditionally been incorporated into prediction models trough indirect indicators adapted to local conditions. Examples of these indicators are: distance to urban and recreational areas (Romero-Calcerrada 2008; Padilla and Vega-García 2011), distance to roads (Pew and Larsen 2001; Hernández-Leal *et al.* 2006), distance to agriculture land (Vasconcelos *et al.* 2001), distance to power lines (Vasilakos *et al.* 2007), agroforestry interface area (Martínez *et al.* 2009) and unemployment rate (Maingi and Henry 2007, Martínez *et al.* 2009). We have also used some of these indicators to assess fire danger derived from human behavior.

Regarding fire occurrences due to lightning, indirect indicators related to topography (Podur *et al.* 2003), fuel characteristics (Chuvieco *et al.* 2010), polarity of lightning strikes (Wotton and Martell 2005) or historical data (Castelo-Dorado *et al.* 2011) have been reported. Since these are long-term indicators and we are seeking for a daily WFD index, we propose a new indicator based on weather forecasts to obtain the probability of storm calculated according to Buizza and Hollingsworth (2002).

Traditionally, fuel conditions and their relationship with meteorological variables have also been incorporated into WFD indexes (Aguado *et al.* 2007; Padilla and Vega-García 2011). Considering this aspect, we use two indicators of fuel conditions: a long-term indicator which measures the species flammability and its influence on the ignition process, and a short-term indicator that measures the probability of ignition according to the methodology by Andrews (1986). This indicator is based on dead fuel moisture content (DFMC) which is considered one of the most important variables in the fire ignition component (Yebra *et al.* 2008; Nieto *et al.* 2010).

Following the proposal of Stocks *et al.* (1989), since forest fires are a complex phenomenon with many variables involved which are difficult to predict, and where it is not possible to model every ignition agent, we also included an indicator of historical fire occurrence in our index.

FDR systems already in use, such as the Canadian or the United States systems (Stocks *et al.* 1989) include a fire propagation component. Spread rate (Rothermel 1983) is considered a good indicator to estimate the probability of an outbreaking fire to turn into a wildfire (Chuvieco and Salas 1996). Thus we include this component as a spread danger indicator.

Integration of the index components has been commonly carried out by different techniques, such as qualitative methods, which use classification tables for pairs of components (Gouma and Chronopoulou-Sereli 1998), logistic regression (Preisler *et al.* 2004; Chuvieco *et al.* 2004), neural networks (Li *et al.* 2009), or multi-criteria analysis (MCA) (Vadrevu *et al.* 2010). We decided to apply MCA techniques in the integration of the components of our index because of its potential to aggregate qualitative and quantitative variables and its capability for taking into account experts' opinions which have been incorporated into the index through a weighting of the different components.

Regarding the index validation, this has usually been accomplished through an analysis of the relationship between WFD indexes and historical fire activity (Preisler *et al.* 2004; Catry *et al.* 2009; Bradstock *et al.* 2009). Logistic regression has been broadly used in the development of danger indexes (Martínez *et al.* 2009; Chuvieco *et al.* 2010) as well as a validation tool (Andrews *et al.* 2003). We used logistic regression techniques, more specifically Generalized Estimating Equations (GEE) models to validate our index, using daily historical fire observations as truth terrain data to contrast the index.

The purpose of this paper is to present the structure of a new WFD index specially adapted to the CV, and the methodology and difficulties encountered in its validation.

Methods

The wildland fire danger index

The WFD index is structured at four hierarchical levels (Fig. 2), integrating all its components into a single value. At the second level, the index is considered a combination of two main components: ignition and spread danger. Following Verde and Zêzere (2010), the ignition danger component was divided into ignition agents (human and natural causes) and fuel conditions, but we decided to add a historical fire occurrence factor at this level to better account for lurking human factors which could be responsible of fire (Castedo-Dorado *et al.* 2007).

Fig. 2. Hierarchical structure of the WFD index.

(i) Ignition agents. Modeling fire danger related to human activity is very complex (Sturtevant and Cleland 2007), this explains why human factors are rarely included in fire danger models (Martínez *et al.* 2009). We have selected a range of human risk ignition indicators representative of different specific human risk activities. All of them are long-term risk indicators:

- Roads: distance to roads was used as a danger indicator because the presence and distribution of ignition agents are closely related to road accessibility to forest land. Distances between 100 and 500 m have been considered as relevant in previous studies (González-Calvo *et al.* 2008). We decided to assign a 2 danger score to points in the territory which are less than 250 m apart from roads, 1 to points between 250 and 500 m and 0 to those points whose distance to roads is larger than 500 m.
- Railroads: we have defined a 100 m wide buffer around railroads. Points in this area get a danger value of 1. Previous data on forest fires (DGMNPF 2010) showed absence of fire ignitions for points farther than this distance, so they receive a 0 risk value.

- Power lines: distance to power lines has been also used as a measurement of ignition danger due to photovoltaic arc (Vasilakos *et al.* 2007). Early reports (DGMNPF 2010) showed that the arc capable of causing a fire reaches a maximum of 50 m, so we decided to assign a value of 1 to a 50 m buffer either side of the line, and 0 to the rest of the territory.
- Agricultural interface: in 2010, 9% of the wildland fires in the CV were caused by agricultural practices (DGMNPF 2010). Following Milani *et al.* (2002), we have defined four buffers of risk around agricultural areas. Distances to these areas are provided by the Preventive Silvicultural Plan of the CV (CMA 1995). Points in the territory are classified as being up to 100, 200, 500 or more than 500 m away from agricultural areas, and receive a decreasing risk value from 4 to 0 as their distance to these areas increase.

Regarding the natural ignition agents, being storm lightning the most efficient cause of fire ignition in the CV in terms of total area affected per number of fires (DGMNPF, 2010), instead of associating the storm lightning ignition to factors related to vegetation and orography, the forecasted probability of electric storm, provided by the Ensemble Prediction System of the European Center for Medium-term Weather Forecast (ECMWF) is used as a short-term risk indicator (Buizza and Hollingsworth, 2002). This indicator ranges from 0 to 1 and it is doubled when the storm is a dry one.

(ii) Fuel conditions. With respect to the ease at which flammable materials may ignite, we have included two of the elements which have a bearing on the same: FMC through its relationship to the probability of ignition, and species flammability.

As a short-term indicator we focused on moisture content of fine dead fuels (FDFMC) because they are very dependent on atmospheric changes (Chuvieco *et al.* 2004) and their FMC is inversely related to the probability of ignition (Danson and Bowyer 2004). Therefore, we have obtained the probability of ignition using the methodology of BEHAVE which takes into account fuel shading, FDFMC and air temperature (Andrews 2009). FDFMC calculations were performed assuming the maximum weather forecasted risk scenario, i.e., the worst predicted combination of high temperatures and very low humidity conditions.

Although FMC is a relevant indicator for the ease of ignition, different species show different ignition times and heating values under the same FMC because of their flammability (Núñez-Regueira *et al.* 1997), which is directly influenced by the chemical composition of its flammable gases, resins, terpenes, etc. To evaluate this long-term danger factor, we have employed an indicator based on the flammability studies of INIA (Martín and Hernando 1989) and INRA (Valette *et al.* 1979) specially adapted to Mediterranean species. Using the classification carried out by these authors, we turned the CV forest vegetation map (DGMNPF 2005) into a flammability danger map, assigning a flammability value between 1 and 4 to each vegetation type.

(iii) Historical occurrence. Historical fire occurrence databases acquire great importance in order to detect hidden human factors which are responsible for fire and to discover differences in fire danger between regions (Castedo-Dorado *et al.* 2007). Therefore, we decided to add the historical wildfire occurrence risk index of ICONA (Vélez 2000) as a long-term danger indicator. This factor, besides the affected area, analyses the number and causes of fires.

(iv) Spread danger. We estimated the spread danger component by means of a short-term indicator: the spread rate. This indicator is calculated using the BEHAVE program (Andrews 1986) and implementing the corresponding weighted algorithms for each standard fuel model (Anderson 1982) derived from the cartography (DGMNPF 2005) assuming, once more, the maximum weather forecasted risk scenario.

Study area and data

Our study area is the forested land in the CV, a Spanish autonomous region that lies in the Mediterranean basin. This forested land is made up of woodland, grassland and shrubland areas, encompassing an overall area of 1247090 ha which represents 53% of the total extent of land in the CV. The climate is temperate, mean annual temperature ranges from 11ºC to 26ºC with average annual rainfall varying from 300 to 700 mm. This region is divided into three administrative provinces: Castellón, Valencia and Alicante (Fig. 1), which present a distinct degree of intensity in urban and agricultural activities, and are characterized by a north–south vegetation gradient. Furthermore, some factors that affect the WFD index have a distinct influence in each province (e.g. lightning is responsible for most of forest fires in Castellón). As these provinces may present different responses to fire, we are interested in checking if the proposed WFD index accounts for those differences.

The data needed to carry out calculations of the index indicators are very heterogeneous. Some of these come in cartographic format, while others, such as weather forecast data, come from alphanumeric databases which need to be adapted to cartographic format before calculations. A digital terrain model and data on land use, roads, and power lines have been provided by the MA10 and the CV10 cartographical series of the Valencian Cartographic Institute on a 1:10000 scale. Data related to land cover, species flammability, type of vegetation and fuel model, adapted to the 13 standard fuel models of Anderson (1982) were accommodated using the cartography from the third National Forest Inventory (DGMNPF 2005) on a 1:50000 scale.

Meteorological data were supplied by the company *Meteogrid S.L.* Daily forecasts on a 0.5 km grid that covers the CV are based on reports of the ECMWF. Fire records and data used to derive the historical occurrence indicator and to validate the index came from the forest fire database of the Ministry of Environment (DGMNPF 2006). Wildfire data are related to a 10-km grid cell system anchored to UTM (Universal Transverse Mercator) coordinates. A summary of all the referenced datasets is shown in Table 1.

Table 1. Datasets

Name, source, year of publication and scale/resolution for all the datasets used to build the index.

Dataset	Source	Year	Scale / Resolution
Digital terrain model	Valencian Cartographic Institute	2000	1:10000
Land uses			
Power lines			
Roads			
Types of vegetation	Ministry of Environment of Spanish Government	2005	1:50000
Fuel model			
Land cover (vegetation)			
Flammability (species)			
Weather forecast	Meteogrid. Ltda. / European Centre for Medium-Range Weather Forecasts		0.5 km
Forest fire records	Ministry of Environment of Spanish Government	2006	10 km

The initial data for the model were first mapped (vegetation, roads, agricultural interface, weather forecasts, etc.) onto a 0.5-km grid covering the whole territory of the CV. This scale is consistent with other studies on regional indexes (Sebastián-López *et al.* 2002; Martínez *et al.* 2009; Padilla and Vega-García 2011).

Index integration

In order to use a homogeneous numerical scale for the different components, prior to integration these were standardized using the score range method (Malczewski 1999). Integration consists of the progressive aggregation of the components that make up every level in the WFD index and, after that, the integration of these levels to obtain a unique value for the index. As MCA methods are oriented to deal with hierarchically structured problems and with situations in which conflicting goals prevail, this methodology, which has been applied extensively in the management of environmental resources (Noble and Chirstmas 2008), has been the one adopted for the aggregation of components.

We used the TOPSIS method (Hwang and Yoon 1981) in the integration of components. This method is applied to obtain a single value at each hierarchical level of the WFD index. This value depends on its distance, in a geometric sense, to the ideal and anti-ideal points (Zavadskas *et al*. 2006).

The opinions of a group of experts are taken into account by weighting each component of the index. An Analytic Hierarchy Process (AHP) (Saaty, 1987) is used to aggregate experts' opinions on each component into a single value. The benefits afforded by this method, due to its capability to measure and control inconsistency of individual opinions, have been key in adopting this methodology – one of the weighting methods most widely used in environmental studies (Moffet *et al.* 2006). We conducted a poll among a group of experts (e.g. technical managers of forest administration, firefighters, surveillance coordinators, etc.) involved in wildfire prevention in the CV. The opinion of nine experts about the relative contribution of each factor to the corresponding hierarchical danger level was consulted and integrated in the index structure through pair-wise comparisons at every index level. Weights obtained applying this method are shown in Table 2.

Table 2. Hierarchical levels, names and weights for the WFD components.

Weights obtained integrating the opinion of a group of experts through an AHP

Hierarchical level	Component	Weight
FWD	Ignition danger	0.63
	Spread danger	0.37
Ignition danger	Historical occurrence	0.11
	Ignition agents	0.45
	Fuel conditions	0.44
Ignition agents	Roads	0.18
	Power lines	0.06
	Railroads	0.06
	Agriculture	0.34
	Lightning	0.36
Fuel conditions	Probability of ignition	0.63
	Flammability	0.37

Index validation

Before introducing our WFD index to fire prevention units, it should undergo rigorous evaluation and validation to fully assess its relationship to fire activity and to detect any possible limitation. Most authors use historical fire activity to check the performance of fire danger indexes, analyzing the relationship between the values taken by the index and fire occurrence (Chuvieco *et al.* 2010). Others compare their indexes to a well known and established set of indexes (Dasgupta *et al.* 2006; Sharples *et al.* 2009).

Since fires are the unique available truth terrain data, we used the historical fire occurrences compiled by the Spanish official statistics on wildfires (DGMNPF 2006) recorded on a grid of 275 10-km grid cells as the data to assess our index performance.

(i) Sample data. The data consisted of historical records for the period from 1994 to 2003. This study period is consistent with previous studies (Preisler *et al.* 2008; Carmel *et al.* 2009) and with the usual time-frame for fire prevention planning in Spain (Vélez 2000). A 10 year period like this, involves a sample with a large number of observations with no fire and a very small number of cases with fire. To ensure a minimum number of cases with fire in the sample, some authors have drawn their samples retrospectively by selecting those days with high probability of fire activity (e.g. summer days or days with at least one fire) (Dasgupta *et al.* 2006; Chuvieco *et al.* 2010).

In our case, data collection were oriented to include all large fires that took place during the study period. In total, 60 days with at least one fire affecting more than 100 ha were selected. As data referred to a grid of 275 10-km cells, our sample is integrated by a total of 275 x 60 = 16500 observations (day-cells).

Therefore, although the index resolution is 0.5 km, since the historical fire database resolution is 10 km, we were forced to assign a single WFD index value to each 10-km day-cell by summarizing WFD index values in the 0.5-km day-cells that make them up (see details in Fig. 3). Each 10-km cell is made up of 400 0.5-km cells, however as the index is only calculated in those cells with forested vegetation (where a fuel model applies), the number of forested cells that make up the 10-km cells is variable, ranging from 3 to 394.

Fig. 3. Detailed view of the different grids used: our WFD index is calculated for all the 0.5-km forested cells, but historical data on fire occurrence are related to 10-km cells.

Since wildland fire is a contagious process (Chou *et al.* 1990), the probability of fire occurrence in a location is influenced not only by local conditions but also by conditions in surrounding areas (Bachmann and Allgöwer 2001). Therefore, on a given day, the probability of fire occurrence in a 10-km cell is strongly related to the maximum probability of fire occurrence in the set of 0.5-km forested cells that make it up. In fact, other summary measures as the average or the median could hide extreme conditions. So, on a daily basis, the summarizing measure for the index in each of the 10-km cells was calculated as the maximum value of the index in the forested cells that make them up.

Figure 4 shows a map with the probability of fire occurrence (WFD index values) estimated on a given date in each 0.5-km forested cell, highlighting the 10-km cells where a forest fire took place. The right side of this picture represents the aggregated values of the index in each 10-km cell.

Fig. 4. Probability of fire occurrence map on a 0.5-km grid (left), and its aggregated values on a 10-km grid (right), including the location of forest fires (10/08/1994).

(ii) Validation methodology

To validate our index, we want to analyze if a positive significant relationship with the observed fire occurrences exists, and also, if the index is capable of capturing the effect of the potential differences between provinces and the different number of forested cells that compose each 10-km cell.

A wide range of statistical methods have been used in the validation of WFD indexes. Some nonparametric tests as the Wilcoxon and the Kruskall-Wallis tests are used to see if significant differences on the index values exist between cells/days with and without fire activity (Wasserman 2007). Logistic regression is one of the techniques most widely used because of its capability for modeling binary data (Preisler *et al.* 2008; Catry *et al.* 2009; Bradstock *et al.* 2009; Padilla and Vega-García 2011). Andrews *et al.* (2003) demonstrated the reliability of using logistic regression to validate fire danger indexes. Following these authors, we aimed to use a logistic regression model to validate our index, however, observations in our sample may not be independent because they conform a set of 275 clusters, each a 10-km cell, with repeated observations on 60 different days. Responses in each cluster might be correlated due to the impact that some geographical characteristics and time considerations have on the different risk indicators, so the assumption of independence of responses made by standard logistic regression models may not hold in our setting. We needed to use an analytic approach that explicitly took into account possible correlated binary responses: the generalized estimating equations (GEE) models. A review of methods for clustered binary data can be found in Pendergast *et al*. (1996).

We have tested two within cluster correlation structures: first order autoregressive (AR1) and independent (Kleinbaum and Klein 2002). If responses were correlated, the AR1 structure accounts for situations where the chances of a fire occurring at a certain cell on a given time period are dependent on the situation encountered in that cell on the previous time period considered. This is clearly the case when the time interval between two observations is small and for cells where the forested area is not too large (if a fire burned all the available fuel, the chances of having another fire in the next observation period are less than if there were no fire in that cell in the near past).

To perform an assessment of fit via external validation, we decided to randomly exclude 25% of our observations and develop a model based on the remaining cases. To assure a minimum number of day-cells in the validation sample, as our data constitute a set of clusters (275 10-km cells) with repeated observations (60 days), we randomly selected 15 days and excluded all the related cell-days from the original sample to obtain and fit the models.

Results and discussion

Exploratory analysis

Before proceeding with the GEE models, we carried out an exploratory analysis of our index using only the sample for calibration – 45 observations (days) in each of the 275 10-km cells. The total number of day-cells without fires is 12176 while the number of day-cells with fires is 199, representing 1.63% of the whole set. The distribution of the index values is similar in all 275 cells, ranging from 0.23 to 0.76. As it might be expected, there is a positive correlation (r = 0.4) between the index values and the number of forested cells that make up each 10-km cell (*NumCells)* ($P < 2.2\cdot10^{-16}$ from the Spearman's rank correlation test). This result suggests that the variable *NumCells* could be an explanatory variable in the GEE model.

Although the number of fires and their geographical distribution is similar for the three provinces, Valencia is the province with the largest area, while Castellón is the one with the largest mean number of forested cells per 10-km cell (193.2) compared to Valencia (161.9) and Alicante (122.2), these differences are clearly significant (P = 0.0007 from a Kruskal-Wallis rank sum test). This outcome and the different geographic and demographic characteristics of the three provinces convert *Province* into another eligible explanatory variable.

A first approach to assess the performance of the WFD index is to compare the distributions of the index in the samples of day-cells with and without fires. Comparisons of theses distributions through box plots and the empirical cumulative distribution functions (ECDF) (Fig. 5) showed the expected behavior of the index: day-cells where a fire took place present larger values of the index than day-cells with no fire. This conclusion was supported by a Wilcoxon rank sum test with continuity correction applied to two samples of equal sizes (each with 199 observations), where a one-sided *p-value* as small as $8.47 \cdot 10^{-7}$ shows a clearly significant and positive shift location for the distribution of the index in the population of day-cells with fires.

Fig. 5. Comparison of the WFD index distribution in the samples of cell-days with and without fires: boxplots and Empirical Cumulative Distribution Funtions (ECDF).

GEE models fitting

Taking into account these results, apart from the index, two additional explanatory variables were considered to account for possible sources of variation associated with the *Province* and to the number of 0.5-km forested cells that make up each cluster (*NumCells*). The data structure is presented in Table 3.

Table 3. Data structure.

For each observation (day-cell): cell identification, province, number of forested cells that make up each 10-km cell, index value and the binary response variable.

Cell. Id	Province	NumCells	Index	Fire
1	Valencia	17	0.29	0
1	Valencia	17	0.34	0
...	...	...	...	...
177	Castellón	31	0.51	1
177	Castellón	31	0.62	0
...	...	...	...	...
275	Alicante	14	0.38	0
275	Alicante	14	0.34	0

The initial GEE logistic models proposed to fit our data, considered the effect of these three explanatory variables and all second and third order interaction terms. Table 4 shows results for the terms that could not be eliminated from the initial model using a drop-in-deviance test, for a GEE model with an AR1 correlation structure. A backward elimination procedure was used to eliminate non significant variables (significance level α=0.05).

Table 4. GEE model that assumes an AR1 correlation structure

Estimates for the coefficients, standard errors, Wald statistics and associated *P* values for the explanatory variables that could not be eliminated using a backward selection procedure.

Parameter	Estimate	Std. Error	Wald	P(>|W|)

Intercept	-8.06447	0.53490	227.31	$< 2\cdot10^{-16}$
Index	7.52329	0.91535	67.55	$2\cdot10^{-16}$
Castellón	-0.13906	0.47117	0.09	0.768
Valencia	0.37385	0.36953	1.02	0.312
NumCells	0.00155	0.00157	0.97	0.324
Cast*NumCells	-0.00217	0.00213	1.04	0.309
Val*NumCells	-0.00441	0.00192	5.52	0.019

The estimated correlation parameter for this model is r = 0.0439 which is not significant (P =0.31 two-sided Wald test). Although not shown here, to avoid replication, we obtained similar estimated parameters and standard errors with the GEE model that use an independence correlation structure. Furthermore results are almost identical to the ones obtained from scratch, with a standard logistic regression model using a backward selection method (Table 5). All these results reveal that responses inside the 10-km cells are independent.

Table 5. Standard logistic regression model

Estimates for the coefficients, standard errors, z statistics and associated P values for the explanatory variables that could not be eliminated using a backward selection procedure.

Parameter	Estimate	Std. Error	z-value	P(>|z|)
Intercept	-8.16273	0.53724	-15.19	$< 2\cdot10^{-16}$
Index	7.72658	0.95357	8.1	$5.4\cdot10^{-16}$
Castellón	-0.12215	0.40976	-0.3	0.766
Valencia	0.36408	0.32096	1.13	0.257
NumCells	0.00151	0.00153	0.99	0.324
Cast*NumCells	-0.00225	0.00201	-1.12	0.262
Val*NumCells	-0.00449	0.00178	-2.53	0.011

In these models, the only significant effects are: *Index* which is clearly significant, and the interaction term *Val*NumCells.* As results obtained with the three models are consistent, using as an example, the standard logistic regression model, the estimated coefficient associated to *Index* is 7.727 – 95% confidence interval (CI): 5.86 to 9.60 – showing a clear positive association between the values of our index and the probability of a fire being present.

The estimated coefficient associated to the interaction term *Val*NumCells* is -0.004489 (95% CI: -0.00798 to -0.001), so it seems that the larger the number of forested cells that make up each 10-km cell in Valencia, the lower the probability of having a fire. Although this upshot needs further investigation, a possible explanation of this result could be related to the existence of a distinctive distribution of the forested cells in the province of Valencia, as could be the case if, those 10-km cells which are integrated with a lower number of 0.5-km forested cells were the ones with a higher risk of fire because of, for example, being closer to railways or roads, or being the ones that are immersed into agricultural areas.

Assessing the fit of the model

Using the validation sample, we followed the directions of Hosmer and Lemeshow (2000), and used a combination of three tests to assess the fit of the model: the Hosmer-Lemeshow decile of risk test yield a $\hat{C}$ statistic of 7.76 (P=0.54), hence we conclude that the model fits, the Osius and Rojek normal approximation to the distribution of the Pearson chi-square statistic, another overall measure of the model fit, was z =-0.85 (P=0.39), so again we cannot reject the hypothesis that the model fits, and finally, the Stukel's test that determines if the tails of the proposed model are either longer or shorter that the standard logistic regression model. This contrast provides a test of the basic logistic regression model assumption and in that sense it is a useful adjunct to the previous tests. The partial likelihood ratio for this test yield a value of 1.56 (P=0.21), concluding that we cannot reject the hypothesis that the logistic regression model is the correct model.

Another measure of model performance which could be a useful supplement to the previous overall tests of fit is the ROC curve which plots sensitivity (probability of detecting true fire) and 1-specificity (probability of detecting a false fire). The area under the ROC curve provides a measure of the model's ability to discriminate between those day-cells with an actual fire, and those day-cells with no fire. The area under the ROC curve in our model is 0.73, pointing an adequate performance of our model, as values between 0.7 and 0.9 are considered as useful discrimination (Swets 1988). Similar values are reported by Modugno *et al.* (2008) for their risk index in Catalunya (Spain) and Padilla and Vega-García (2011) for the logistic models they propose to model fire risk in 53 Spanish ecoregions.

Model applicability

The sample used to carry out all the analyses was selected retrospectively from the population of days with fires to assure a minimum number of observation with fire. As this is a case-control study, prospective probabilities cannot be estimated because the intercept in these models cannot be validly estimated without knowledge of the sampling fractions within cases and controls, being the *risk odds ratio* (ROR) – which compares the odds of having a fire in two different locations – the only valid estimate that can be used to compare two groups of binary responses (Hosmer and Lemeshow 2000).

Using the estimated coefficients from the proposed logistic regression model (Table 5), the ROR comparing two cells in Castellón and/or Alicante, on a specific day where, as an example, the WFD index takes the values 0.5 and 0.6 would be: ROR = exp(7.72658 x (0.6-0.5)) = 2.17. So, if we compare cells where the WFD index differs in 0.1 units, the odds of fire for the cells with the largest values are estimated to be 2.17 times as large as the odds of fire for the cells with the lowest values. A 95% CI for the odds of fire for cells where the WFD index increase 0.1 points relative to the other cells is 1.80 to 2.61.

When considering cells in the province of Valencia, the effect of increasing the values of the WFD index on the odds ratio depends on the number of forested cells that make up each 10-km cell. Figure 6 presents point and confidence estimates of the odds ratio for 0.1 and 0.25 increase in the WFD index. The point-wise 95% limits are indicated by the vertical bars. The graph indicates that 10-km cells in Valencia whose WFD index values increase (by 0.1 or 0.25 points) are progressively more likely to remain fire free as the number of 0.5-km forested cells that make up each 10-km cell increases. The decrease in the estimates of the odds ratio are more important for large increments of the WFD index values (e.g. the odds ratios gradually decrease from 6.89 to 4.43 for a 0.25 increase in the index, while the odds ratios change from 2.16 to 1.81 for a 0.1 increase).

Fig. 6. Estimated odds ratio and 95% confidence limits for a 0.1 and 0.25 increase in Index based on the model in Table 4.

The proposed models (Tables 4 and 5) which were use for validating our WFD index, could also be exploited this way to obtain a risk map on a 10-km scale using the odd ratios derived from them.

Conclusions

In this paper, a new WFD index to obtain the daily probability of fire occurrence and specially adapted to a Mediterranean Spanish region has been presented. The index is structured in four hierarchical levels, including dynamic and structural risk ignition indicators and can be used to improve our ability to target resource protection efforts and manage fire risk on a local scale.

The index uses the TOPSIS method in the aggregation of components and the AHP methodology to integrate the opinions of a group of experts on the importance of the different risk and danger factors, obtaining this way a better fit to local conditions in the study area.

GEE models used to validate the index showed that responses inside the different cells on a 10-km grid were independent. This result denotes the appropriateness of the standard logistic models for this type of studies.

The proposed WFD index, and the second order interaction term *Val*NumCells* are clearly significant to predict fire occurrence on a 10-km cell grid. Validation results showed good index performance, good fit of the logistic model and acceptable discrimination power.

Apart from validating the index, the proposed GEE models can be used to derive risk odd ratios, which can be used to obtain a new map representation of fire risk. This idea could be used to obtain risk maps in those case-control analysis that use logistic regression or other binomial regression techniques to predict the probability of fire occurrence, where the prospective probabilities cannot be properly estimated.

The proposed WFD index could be integrated into a more general FDR model which considers also vulnerability, accounting this way, for potential damage caused by fire.

Acknowledgements

The authors acknowledge the support received from the Ministry of Science and Innovation through the research project *Modelling and Optimization Techniques for a Sustainable Development*, Ref. EC02008-05895-C02-01/ECON.

References

Aguado I, Chuvieco E, Boren R, Nieto H (2007) Estimation of dead fuel moisture content from meteorological data in Mediterranean areas, applications in fire danger assessment. *International Journal of Wildland Fire* **16**, 390–397.

Anderson H (1982) Aids to determining fuel models for estimating fire behavior. USDA Forest Service, Intermountain Forest and Range Experiment Station, General Technical Report INT-122. (Ogden, UT)

Andrews PL (1986) BEHAVE: Fire behaviour prediction and fuel modelling system - BURN subsystem, part 1. USDA Forest Service, Intermountain Forest and Range Experiment Station, General Technical Report INT-194. (Ogden, UT)

Andrews PL, Loftsgaarden DO, Bradshaw LS (2003) Evaluation of fire danger rating indexes using logistic regression and percentile analysis. *International Journal of Wildland Fire* **12**, 213–226. doi:10.1071/WF02059.

Andrews PL (2009) BehavePlus Fire Modeling System, Version 5.0: Variables. General Technical Report RMRS-GTR-213WWW Revised. Department of Agriculture, Forest Service, Rocky Mountain Research Station. (Fort Collins, CO)

Bachmann A, Allgöwer B (2001) A consistent wildland fire risk terminology is needed! *Fire Management Today* **61**(4), 28–33.

Benvenuti M, Testi C, Conese C, Romani M (2002) Forest Fires Ignition Probability Model: test and validation on FORFAIT Italian pilot site. In: 'Proceedings of IV International Conference on Forest Fire Research & Wildland Fire Safety'. Luso, Coimbra, Portugal

Blanchi R, Jappiot M, Alexandrian D (2002) Forest fire risk assessment and cartography, a methodological approach. In: 'Proceedings of IV International Conference on Forest Fire Research & Wildland Fire Safety'. Luso, Coimbra, Portugal

Bradshaw L, Deeming J, Burgan RE, Cohen J (1983) The 1978 National Fire-Danger Rating System: technical documentation. USDA Forest Service, Intermountain Forest and Range Experiment Station, General Technical

Report INT-169. (Ogden, UT)

Bradstock RA, Cohn JS, Gill AM, Bedward M, Lucas C (2009) Prediction of the probability of large fires in the Sydney region of south-eastern Australia using fire weather. *International Journal of Wildland Fire* **18**, 932–943.

Buizza R, Hollingsworth A (2002) Storm Prediction Over Europe Using the ECMWF Ensemble Prediction System. *Meteorological Applications* **9**(3), 289–305.

Castedo-Dorado F, Juarez I, Ramírez J, Ruiz I, Rodríguez C, Vélez L (2007) Utilidad del análisis de la estadística de incendios en las estrategias de prevención y extinción. In 'Proceedings 4th International Wildland Fire Conference', Sevilla, Spain. (Ministerio de Medio Ambiente)

Castedo-Dorado F, Rodríguez-Pérez JR, Marcos-Menéndez JL, Álvarez-Taboada MF (2011) Modelling the probability of lightning-induced forest fire occurrence in the province of León (NW Spain). *Forest Systems* **20**(1), 95-107.

Carmel Y, Paz S, Jahashan F, Shoshany M (2009) Assessing fire risk using Monte Carlo simulations of fire spread. *Forest Ecology Management,* **257**:370–377.

Catry FX, Rego FC, Bação F, Moreira F (2009) Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire* **18**, 921-931.

Climent J, Prada MA, Calama R, Sánchez de Ron D, Chambel MR, Alía R (2008) To grow or to seed: ecotypic variation in reproductive allocation and cone production by young female Aleppo pine (Pinus halepensis, Pinaceae) *American Journal of Botany* **94,** 1316–1320.

CMA (1995) Plan de selvicultura preventiva de incendios en los sistemas forestales de la Comunidad Valenciana, Valencia, Spain. (Conselleria de Medio Ambiente)

Chou YH, Minnich R, Salazar L, Power J, Dezzani R (1990) Spatial autocorrelation of wildfire distribution in the Idyllwild Quadrangle, San Jacinto Mountain, California. *Photogrammetric Engineering and Remote Sensing* **56**, 1507–1513.

Chuvieco E, Congalton RG (1989) Application of remote sensing and Geographic Information Systems to Forest fire hazard mapping. *Remote Sensing of Environment* **29,** 147–159.

Chuvieco E, Cocero D, Riaño D, Martín P, Martínez-Vega J, Riva J, Pérez F (2004) Combining NDVI and surface temperature for estimation of live fuel moisture content in forest fire danger rating. *Remote Sensing of Environment* **92**, 322–331.

Chuvieco E, Salas J (1996) Mapping the spatial distribution of forest fire danger using GIS. *International Journal*

GIS **10(**3), 333–345.

Chuvieco E, Englefield PB, Trishchenko AP, Yi Luo C (2008) Generation of long time series of burn area maps of the Boreal forest from NOAA-AVHRR composite data. *Remote Sensing of Environment* **112**, 2381–2396.

Chuvieco E, Aguado I, Yebra M, Nieto H, Salas J, Martín MP, Vilar L, Martínez J, Martín S, Ibarra P, de la Riva J, Baeza J, Rodríguez F, Molina JR, Herrera MA, Zamora R (2010) Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling* **221**, 46–58.

Danson FM, Bowyer P (2004) Estimating live fuel moisture content from remotely sensed reflectance. *Remote Sensing of Environment* **92**, 309–321.

Dasgupta S, Qu JJ, Hao X (2006) Design of a susceptibility index for fire risk monitoring. IEEE *Geoscience and Remote Sensing Society Newslett*er, **3**(1), 140−144.

DGMNPF (2005) Mapa forestal de España. Escala 1:50.000. Comunitat Valenciana. Serie técnica. Formato digital. Banco de datos de la naturaleza, Madrid, Spain, (Ministerio de Medio Ambiente)

DGMNPF (2006) Los incendios forestales en España. Decenio 1996–2005. Área de Defensa Contra Incendios Forestales, Madrid, Spain. (Ministerio de Medio Ambiente)

DGMNPF (2010) Los incendios forestales en España. Año 2010. Área de Defensa Contra Incendios Forestales, Madrid, Spain. (Ministerio de Medio Ambiente)

Fairbrother A, Turnley JG (2005) Predicting risks of uncharacteristic wildfires: application of the risk assessment process. *Forest Ecology and Management* **211**, 28–35.

Finney MA (2005) The challenge of quantitative risk assessment for wildland fire. *Forest Ecology Management* **211**, 97–108.

González-Calvo A, Hernández-Leal PA, Alonso-Benito A, Arbelo M, Arvelo-Valencia L (2008) Modelado del riesgo de incendios forestales en las Islas Canarias usando datos de satélite y aplicaciones SIG. In 'Tecnologías de la Información Geográfica para el Desarrollo Territorial. Servicio de Publicaciones y Difusión Científica de la ULPGC', Las Palmas de Gran Canaria, Spain, 588–596.

Gouma V, Chronopoulou-Sereli A (1998) Wildland fire danger zoning—a methodology. *International Journal of Wildland Fire* **8**(1), 37–43.

Hernández-Leal PA, Arbelo M, Gonzalez-Calvo A (2006) Fire Risk assessment using satellite data. *Advances in Space research* **37**(4), 741–746.

Hosmer DW, Lemeshow S (2000) 'Applied logistic regression.' (John Wiley & Sons: New York)

Hwang CL, Yoon KL (1981) 'Multiple attribute decision making: Methods and applications.' (Springer-Verlag: New York)

JRC, Joint Research Center (2011) Forest Fires in Europe 2010. Report nº 11. EUR 24910 EN, Luxembourg. (Office for Official Publications of the European Communities 2009)

Kleinbaum DG, Klein M (2002) 'Logistic Regression: A Self-Learning Text.' (Springer-Verlag: New York)

Li LM, Song WG, Ma J, Satoh K (2009) Artificial neural network approach for modeling the impact of population density and weather parameters on forest fire risk. *International Journal of Wildland Fire* **18**, 640–647.

Maingi JK, Henry MC (2007) Factors influencing wildfire occurrence and distribution in eastern Kentucky, USA. *International Journal of Wildland Fire* **16**, 23–33.

Malczewski J (1999) 'GIS and Multicriteria Decision Analysis.' (John Wiley & Sons: New York)

Martell DL, Otukol S, Stocks BJ (1987) A logistic model for predicting daily people-caused forest fire occurrence in Ontario. *Canadian Journal of Forest Research* **17**, 394–401.

Martín E, Hernando L (1989) Inflamabilidad y Energía de las Especies de Sotobosque. Monografía INIA nº 68, Madrid, Spain. (Ministerio de Agricultura Pesca y Alimentación)

Martínez J, Vega-García C, Chuvieco E (2009) Human-caused wildfire risk rating for prevention planning in Spain. *Journal of Environmental Management,* **90**, 1241–1252.

Mediavilla J, Alcover V, Tamayo J, Correa B (1994) Índice Meteorológico de Peligrosidad de los Incendios Forestales. Centro Meteorológico Territorial de Valencia. Grupo de Predicción y Vigilancia de Valencia. Campaña de apoyo meteorológico al PREVIFCO. Nota técnica nº 2, 15 de marzo de 1994, Spain. (Instituto Nacional de Meteorología)

Milani D, Ferraz SFB, Ferraz FFB, Moraes GF (2002) Forest fire risk evaluation system - mapping of preventive action at International Paper do Brasil. In 'Proccedings of IV International Conference on Forest Fire Research & Wildland Fire Safety'. 18–23 November 2002, Luso, Coimbra, Portugal

Modugno S, Serra P, Badia A (2008) Dinámica del riesgo de ignición en un área de interfase urbano-forestal. In 'Tecnologías de la Información Geográfica para el Desarrollo Territorial. Servicio de Publicaciones y Difusión Científica de la ULPGC'. Las Palmas de Gran Canaria, Spain, 650–659.

Moffett A, Garson J, Sarkar S (2005) MULTCSYNC: a software package for incorporating multiple criteria in conservation planning. *Environmental Modelling and Software* **20**, 1315–1322.

Nieto H, Aguado I, Chuvieco E, Sandholt I. (2010) Dead fuel moisture estimation with MSG–SEVIRI data. Retrieval of meteorological data for the calculation of the equilibrium moisture content. *Agricultural and Forest*

Meteorology **150**, 861–870.

Noble BF, Christmas LM (2008) Strategic environmental assessment of greenhouse gas mitigation options in the Canadian Agricultural Sector. *Environmental Management* **41,** 64–78.

Núñez-Regueira L, Rodríguez Añón J, Porpullón J (1997) Calorific values and flammability of forest species in Galicia. Continental high mountainous and humid Atlantic zones. *Bioresource Technology* **61**, 111–1109.

Padilla M, Vega-García C (2011) On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human-caused fire occurrences in Spain. *International Journal of Wildland Fire* **20**, 46–58.

Pendergast JF, Gange SJ, Newton MA, Lindstrom MJ, Palta M, Fisher MR (1996). A survey of methods for analyzing clustered binary response data. *International Statistical Review* **64**, 89–118.

Pew KL, Larsen CP (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.

Podur J, Martell DL, Csillag F (2003) Spatial patterns of lightning-caused forest fires in Ontario, 1976–1998. *Ecological Modelling* **164**(1), 1–20.

Preisler HK, Brillinger DR, Burgan RE, Benoit JW (2004) Probability-based models for estimation of wildfire risk. *International Journal of Wildland Fire,* **13**, 133–142.

Preisler HK, Chen S, Fujioka F, Benoit, JW, Westerling AL (2008) Wildland fire probabilities estimated from weather model-deduced monthly mean fire danger indices. *International Journal of Wildland Fire* **17**, 305–316.

Romero-Calcerrada R, Novillo J, Millington D, 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**(3), 341–354.

Rothermel RC (1983) How to predict the spread and intensity of forest and range fires. USA Forest Service, Intermountain Research Station, General Technical Report INT-143. (Ogden, UT)

Saaty TL (1987) Rank generation, preservation, and reversal in the Analytic Hierarchy Decision Process. *Journal of the Decision Sciences Institute* **18**(2), 157–177.

Sahin YG, Turker S (2009) Early Forest Fire Detection Using Radio-Acoustic Sounding System. *Sensors* **9**, 1485–1498.

San-Miguel-Ayanz J, Barbosa P, Schmuck G, Libertà G, Meyer-Roux J (2003) The European Forest Fire Information System (EFFIS). In 'Proceedings of the 6th AGILE', 24–26 April 2003, Lyon, France, 27–30.

Sebastián-López A, San-Miguel-Ayanz J, Burgan RE (2002) Integration of satellite sensor data, fuel type maps and meteorological observation for evaluation of forest fire risk at the pan-European scale. *International Journal of*

Remote Sensing **23**, 2713–271.

Sharples JJ, McRae RHD, Weber RO, Gill AM (2009) A simple index for assessing fire danger rating. *Environmental Modelling and Software* **24**, 764–774.

Stocks BJ, Lawson BD, Alexander ME, Van Wagner CE, Mcalpine RS, Lynham TJ, Dube DE (1989) Canadian Forest Fire Danger Rating System: an overview. *Forestry Chronicle* **65**, 450–457.

Sturtevant BR, Cleland DT (2007) Human and biophysical factors influencing modern fire disturbance in northern Wisconsin. *International Journal of Wildland Fire* **16**, 389–413.

Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* **240**, 1285–1293.

Vadrevu KP, Eaturu A, Badarinath KV (2010) Fire risk evaluation using multicriteria analysis - a case study. *Environmental monitoring and assessment*, **166(**1), 223-239.

Valette JC, Clement A, Delabraze P (1979) Inflammabilité d'espèces méditerranéennes, test rapides. Document PIF 197901, 28 p. INRA. Avignon, France

Vasconcelos MJP, Silva S, Tomé M, Alvim M, Pereira JMC (2001) Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. *Photogrammetric Engineering and Remote Sensing* **67**, 73–83.

Vasilakos C, Kalabokidis K, Hatzopoulos J, Kallos G, Matsinos, Y (2007) Integrating new methods and tools in fire danger rating. *International Journal of Wildland Fire* **16**, 306–316.

Vélez R (2000) Estrategias defensivas. In 'La defensa contra incendios forestales: fundamentos y experiencias'. (McGraw Hill: Madrid), 10.3–10.51.

Verde JC, Zêzere JL (2010) Assessment and validation of wildfire susceptibility and hazard in Portugal. *Natural Hazards and Earth System Sciences* **10**, 485–497.

Vilar L, Gómez Nieto I, Martín Isabel MP, Martínez Vega FJ (2007) Análisis comparativo de diferentes métodos para la obtención de modelos de riesgo humano de incendios forestales. In 'Proceedings 4th International Wildland Fire Conference', 14–18 May 2007, Sevilla, Spain. (Ministerio de Medio Ambiente)

Wasserman L (2007) 'All of Nonparametric Statistics' (Springer: New York)

Wotton BM, Martell DL (2005) A lightning fire occurrence model for Ontario. *Canadian Journal of Forest Research* **35,** 1389–1401.

Yebra M, Aguado I, García M, Nieto H, Chuvieco E, Salas J (2007) Fuel moisture estimation for fire ignition mapping. In 'Proceedings 4th International Wildland Fire Conference', 14–18th May 2007, Sevilla, Spain. (Ministerio de Medio Ambiente)

Yebra M, Chuvieco E, Riaño D. (2008) Estimation of live fuel moisture content from MODIS images for fire risk assessment. *Agricultural and forest meteorology* **148,** 523–536.

Zavadskas EK, Zakarevicius A, Antucheviciene J (2006) Evaluation of ranking accuracy in multi-criteria decisions. *Informatica* **17**(4), 601–618.