Forest Ecology and Management 275 (2012) 117–129

Contents lists available at SciVerse ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest

Sandra Oliveira [a,b,*], Friderike Oehler [a], Jesús San-Miguel-Ayanz [a], Andrea Camia [a], José M.C. Pereira [b]

[a] Institute for Environment and Sustainability, Joint Research Centre, European Commission, Via E. Fermi 2749, 21027 Ispra (VA), Italy
[b] Department of Forestry, School of Agronomy, Technical University of Lisbon, Tapada da Ajuda, 1349-017 Lisbon, Portugal

ARTICLE INFO

Article history:
Received 28 December 2011
Received in revised form 29 February 2012
Accepted 1 March 2012
Available online 18 April 2012

Keywords:
Fire occurrence
Random Forest
Multiple Linear Regression
Mediterranean Europe

ABSTRACT

Fire occurrence, which results from the presence of an ignition source and the conditions for a fire to spread, is an essential component of fire risk assessment. In this paper, we present and compare the results of the application of two different methods to identify the main structural factors that explain the likelihood of fire occurrence at European scale.

Data on the number of fires for the countries of the European Mediterranean region during the main fire season (June–September) were obtained from the European Fire Database of the European Forest Fire Information System. Fire density (number of fires/km^2) was estimated based on interpolation techniques and was used as the dependent variable in the model. As predictors, different physical, socio-economic and demographic variables were selected based on their potential influence in fire occurrence and on their availability at the European level. Two different methods were applied for the analysis: traditional Multiple Linear Regression and Random Forest, the latter being a non-parametric alternative based on an ensemble of classification and regression trees. The predictive ability of the two models, the variables selected by each method and their level of importance were compared and the potential implications to forest management and fire prevention were discussed.

The Random Forest model showed a higher predictive ability than Multiple Linear Regression. Furthermore, the analysis of the residuals also indicated a better performance of the Random Forest model, showing that this method has potentiality to be applied in the assessment of fire-related phenomena at a broad scale. Some of the variables selected are common to both models; precipitation and soil moisture seem to influence fire occurrence to a large extent. Unemployment rate, livestock density and density of local roads were also found significant by both methods. Maps of the likelihood of fire occurrence were obtained from each method at 10 km resolution, based on the selected variables. Both models show that the spatial distribution of fire occurrence likelihood is highly variable in this region: highest fire likelihood is prevalent in the northwest region of the Iberian Peninsula and southern Italy, whereas it is low in northern France, northeast Italy and north of Greece. In the most fire-prone areas, preventive measures could be implemented, associated to the factors identified by both models.

1. Introduction

Forest fires are a major hazard in Mediterranean Europe, where an average of ca. 45.000 fires occur per year (European Commission, 2011). The sustainable management of forests in Europe, which occupy nearly half of Europe's total area (FOREST EUROPE, UNECE and FAO, 2011), requires the understanding of the factors that influence fire occurrence and its environmental and socio-economic consequences. The probability of a fire to occur results from the joint combination of the existence of an ignition source and the adequate conditions for the fire to spread. The conditions for fire to occur can be assessed prior to the fire season taking into consideration the structural factors which remain relatively stable during at least one fire season (San-Miguel-Ayanz et al., 2003), such as topography, climate and infrastructures (road density etc.), providing a long-term evaluation of the most fire-prone areas. The assessment of the conditions for fire occurrence is fundamental to understand the spatial and temporal distribution of fires as well as their causes. Furthermore, fire occurrence is an essential component of fire risk, according to the terminology that considers risk as a combination of probability of occurrence and potential outcome (e.g. Bachmann and Allgöwer, 2000; Chuvieco et al., 2010; Finney, 2005). From this point of view quantitative long-term fire risk assessment is relevant for fire managers, since it supports forest

* Corresponding author. Permanent address: Estrada do Crasto, nr. 5, Casal do Arneiro, 2440-012 Batalha, Portugal. Tel.: +351 91 2856744; fax: +351 244 765515.
E-mail address: sisoliveira@gmail.com (S. Oliveira).

0378-1127/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.foreco.2012.03.003

fire prevention planning, the setting up of fire-fighting organizations and also the restoration of fire affected areas (e.g. San-Miguel-Ayanz et al., 2003).

Previous studies have investigated the factors influencing long-term fire occurrence, fire danger and risk in Europe, mainly at local and regional level. For example, at a local scale, Amatulli et al. (2006) applied classification and regression trees to assess long-term fire risk in a particular area in the southeast of Italy. Martínez et al. (2009) focused their research on the human factors of fire risk in Spain based on structural variables such as unemployment rate, average distance to roads and agrarian landscape patterns. At the European scale, Sebastián-López et al. (2002) applied the fire potential index (FPI), an integrated index which combines long-term and short-term variables to assess fire danger, developed initially by Burgan et al. (1998) for the USA. More recently, Sebastián-López et al. (2008) presented a methodology to integrate socio-economic and environmental variables to model long-term fire danger in southern Europe, using stepwise regression. Koutsias et al. (2005, 2010) describe the results obtained with Geographically Weighted Regression in Southern Europe, using structural human variables in the model.

In this study, we present a novel approach to identify the factors that influence fire occurrence at a broad scale and to model the likelihood of fire occurrence, in the European Mediterranean region (EUMed) where this phenomenon is recurrent. The assessment of fire occurrence at this scale can provide guidance for the planning and implementation of fire prevention measures, in particular the design of forest management strategies adjusted to the conditions of different ecosystems.

The proposed model considers physical and human variables. Physical variables are assumed to reflect the intrinsic natural conditions of the study area, while the importance of the human factors in forest fire occurrence assessment and risk modelling in the EUMed region is widely recognized (Catry et al., 2009; Chuvieco et al., 2010; Koutsias et al., 2005; Leone et al., 2009; Martínez et al., 2009; Chuvieco et al., 2003; Romero-Calcerrada et al., 2008, 2010; Sebastián-López et al., 2008; Vilar et al., 2010). Two different statistical methods were applied to model fire occurrence: Multiple Linear Regression (LR), a technique widely used in similar studies (e.g. Sebastián-López et al., 2008; Syphard et al., 2008) and Random Forest (RF), a non-parametric technique, which has been previously used in ecological applications (Breiman, 2001; Cutler et al., 2007). We investigated the potential applicability of the Random Forest method in the assessment of fire occurrence by comparing the results obtained from both methods. The predictive ability of the models and the level of importance of the selected variables are presented and discussed, focusing on the common variables in both models and the identification of the factors that influence fire occurrence throughout Southern Europe. Maps of the likelihood of fire occurrence in the EUMed region according to the presence of the selected variables were also derived with each method.

2. Materials and methods

2.1. Dependent variable

Data on the number of fires were obtained from the European Fire Database (Camia et al., 2010). The study area covered the five European countries most affected by fire: Portugal, Spain, France, Italy and Greece. The islands close to the mainland and with sufficient fire records were also included (Corsica, in France and Sicily and Sardinia, in Italy). We used a subset of the total dataset corresponding to the data from 2000 until 2007 for the months between June and September (here called the main fire season), which corresponds to the most recent and complete data available for all these five countries that are harmonized in the database.

In the database, records for each individual fire event include the date of the fire, the type of land cover burned, the burned-area size and the administrative region where it occurred. In most cases, the geographical location of each fire event is given as descriptive information based on the administrative region, frequently at NUTS3 level, which corresponds in most EU countries to the administrative level of provinces. In recent years, geographical coordinates have been added to the database; however this information is incomplete and lacks harmonization between countries. The number of geo-referenced fire events available is variable among and within countries and corresponds so far to a small proportion of the total number of fires. The location inaccuracy associated with the fire records made it necessary to transform the individual fire events into a continuous variable using kernel density methods. Previous authors (Allgower et al., 2005; Amatulli et al., 2007; de la Riva et al., 2004; Koutsias et al., 2004) found that these methods were suitable to estimate fire density. As a proxy of fire ignition, fire density forms the dependent variable here.

The fire events with high-resolution geographical information in the database were used for an exploratory analysis of fire distribution, in order to find out the proportion of fires that fall into Wildland and Non-Wildland areas inside each NUTS3 region. Wildland and Non-Wildland areas were defined based on Corine Land Cover (CLC) (Table 1), the only database of land cover information available for most European countries, thus comparable between countries and regions. CLC classes were grouped into three categories, depending on the characteristics of the land cover types and their potential influence on fire occurrence: Wildland, Non-Wildland and Excluded. Wildland areas were selected taking into account their potential influence in fire occurrence and spread under a conservative approach. The excluded classes are those that have a neglectable contribution to fire occurrence, either by absence of vegetation or due to their low flammability. Non-Wildland areas include the other CLC classes whose conditions are not typically related to fire occurrence but where a proportion of the fires can occur (or at least start) mainly due to human causes, such as agricultural areas (e.g. Catry et al., 2009; Leone et al., 2009; Martínez et al., 2009). The results showed that, in Greece, 72% of fires started in Wildland and 28% started in Non-Wildland areas, while in the other countries the proportions were 65% and 35%, respectively. These shares were used to proportionally assign the fire events to the Wildland and Non-Wildland areas of each NUTS3 and randomly distributing the resulting ignition points using GIS tools.

Subsequently, fire density was calculated using kernel density estimation methods. Since the spatial distribution of fires is highly irregular (clustered), adaptive kernel density estimation was applied, with the bandwidth being defined on the basis of the *k*th nearest-neighbour (KNN). This means that the size of the bandwidth is defined according to the distance between each point and its *k*th nearest-neighbour, thus varying with point concentration. The density is therefore calculated using a variable bandwidth that depends on the KNN distance for each point. The software CrimeStat III® (Levine, 2007) was used to create single density surfaces, at 1 km resolution, using bandwidths from 2 to 30 KNN, for each season in each year and per country. Afterwards, a calibration procedure was applied in order to find the most suitable bandwidth size (Amatulli et al., 2007; Breiman et al., 1977). This procedure is based on the total number of points and the distance between them to select the bandwidth that minimizes the effect of under/overestimation for that specific set of points. Amatulli et al. (2007) found that the results obtained with this method show low sensitivity to the precise point location, thus allowing for the random positioning of the fire events when their exact location is unknown. The results obtained per country were then aggregated for the entire EUMed region. The edge effect between neighbouring

Table 1
Corine Land Cover (CLC) classes aggregated in Wildland, Non-Wildland and Excluded.

CLC code	Wildland	CLC code	Non-wildland	CLC code	Excluded
2.4.3	Land principally occupied by agriculture, with significant areas of natural vegetation	1.1.2	Discontinuous urban surfaces	1.1.1	Continuous urban fabric
2.4.4	Agro-forestry areas	1.2.1	Industrial or commercial units	1.2.3	Port areas
3.1.1	Broadleaved forest	1.2.2	Road and rail networks and assoc. land	1.2.4	Airports
3.1.2	Coniferous forest	1.3.1	Mineral extraction sites	2.1.2	Permanently irrigated land
3.1.3	Mixed forest	1.3.2	Dump sites	2.1.3	Rice fields
3.2.1	Natural grasslands	1.3.3	Construction sites	3.3.2	Bare rocks
3.2.2	Moors and heathlands	1.4.1	Green urban areas	3.3.5	Glaciers and perpetual snow
3.2.3	Sclerophyllous vegetation	1.4.2	Sport and leisure facilities	4.1.1	Inland marshes
3.2.4	Transitional woodland-shrub	2.1.1	Non-irrigated arable land	4.1.2	Peat bogs
3.3.3	Sparsely vegetated areas	2.2.1	Vineyards	4.2.1	Salt marshes
3.3.4	Burnt areas	2.2.2	Fruit tree and berry plantations	4.2.2	Salines
		2.2.3	Olive groves	4.2.3	Intertidal flats
		2.3.1	Pastures	5.1.1	Water courses
		2.4.1	Annual crops associated with permanent crops	5.1.2	Water bodies
		2.4.2	Complex cultivation patterns	5.2.1	Coastal lagoons
		3.3.1	Beaches, dunes, sand	5.2.2	Estuaries

countries was reduced by averaging the density within a buffer of 2 km in each side of the border. This buffer size was found to be the best compromise between maintaining the specific density values of each country while allowing for a smoothing effect near the border, after an exploratory analysis with buffers sized 2, 5 and 10 km.

Afterwards, the average fire density for the period 2000–2007 was calculated and corrected for the fire domain extent. It was assumed that there are spatial limitations to the occurrence of fire that derive from the physical or the human characteristics of land cover. The fire domain area, i.e. the area exposed to fire, was defined based on the selection of the Corine Land Cover classes and topographic regions where the conditions for the occurrence of fire are known to exist (*Wildland* and *Non-Wildland* in Table 1), while the land cover types where fire cannot occur were excluded (*Excluded* in Table 1 and all the areas above 2000 m elevation) This is an important step for the analysis, considering the spatial nature of this phenomenon. The fire density values previously obtained from the interpolation method were corrected accordingly; for example, a grid cell where 40% of the area is urban (where no forest fire can occur) will have a higher value of fire density when corrected because fires could only occur in 60% of the area of the grid cell. The average fire density 2000–2007, corrected by the fire domain extent, was used as the dependent variable. This variable was up-scaled to a 10 km resolution by averaging the 1 km cells to a 10×10 km grid covering the EUMed region. This resolution was found to be the best compromise considering the resolution of the different available datasets. The base grid was composed of ca. 16,000 pixel cells covering the study area.

2.2. Explanatory variables: selection and pre-processing

A total of 37 variables were extracted from several databases covering physical, socio-economic and demographic aspects (Table 2). The variables were selected considering both their potential relevance for fire occurrence, based on extensive literature review and experts' knowledge, and their availability at the European level. All the variables were integrated in a Geographic Information System (GIS) and transformed to a continuous scale at 10 km resolution, following the procedures described ahead.

2.2.1. Topography

The topographic features affect vegetation distribution, composition and flammability and have also an influence on local climate variations (Syphard et al., 2008; Whelan, 1995). Topographic variables were obtained from the Digital Elevation Model (DEM) available at the European level. This DEM is based on SRTM images from NASA, further processed in order to fill in no-data voids existing in the original images (Jarvis et al., 2008; Reuter et al., 2007). Elevation values at 1 km resolution were aggregated at 10 km based on average and on maximum values, creating two layers. Slope may affect ignitions by limiting accessibility, with a threshold of 20° (36%) being mentioned as a restrictive factor for harvesting in forest conservation studies (Widayati et al., 2010). Conedera et al. (2011) also found that anthropogenic fires occurred more frequently in gentler slopes, defining a threshold below 33° (approximately 64%) for a study area in Switzerland. In view of these results and after an exploratory analysis of the frequency of slope ranges existing in the EUMed region, slope values were divided in two categories: below 30 percent and above 30 percent. The proportion of area with slope below and above 30 percent in each pixel cell was retrieved, creating two layers. Aspect was reclassified into four main directions: N (293–67°), E (67–113°), S (113–248°) and W (248–293°). Since the data were available at 100 m resolution, taking into account the circular nature of this variable and to avoid losing crucial information when upscaling to 10 km, it was preferred to retrieve the proportion of each pixel occupied by each different aspect, thus creating four other layers (proportion of aspect N, E, S, W).Topographic roughness is considered the amount of land surface variability of a particular area (Stambaugh and Guyette, 2008) and is a proxy for describing the potential of terrestrial propagation (in this case fire spread) related to topographic variability. Roughness was calculated as the ratio between the surface area and the planimetric area based on the DEM at 1 km resolution and using GIS tools. Roughness values were aggregated at 10 km based on average values and on maximum values, creating two new layers.

2.2.2. Land cover

Land cover, which represents the landscape features of the Earth's surface has been previously associated with fire occurrence (e.g. Catry et al., 2009; Martínez et al., 2009; Syphard et al., 2008; Vilar et al., 2010). Considering the specific context of the study area in relation to the causes of fire ignitions, with over 90% being anthropogenic, land cover maps were used as representative of the type of vegetation available to burn, i.e. as a proxy of fuel types, because they reflect the potential interaction with the human components. We used the land cover inventory provided by the Corine database (European Commission, 1994; EEA-ETC/TE, JRC, 2002) to define the land cover variables. From the original 44 classes (level 3) of Corine only those classes with potential influence in fire occurrence (thus matching the fire domain area) were selected and aggregated in seven larger categories (Table 3). The proportion

Table 2
Variables collected to be included as predictors in the model. The codes chosen may not include the whole description of the variable for simplicity purposes.

Variable type	Variable name	Code	Source and reference	Resolution/scale
Environmental – Topographic	Average elevation	Elev_avg	DEM Europe (Jarvis et al., 2008; Reuter et al., 2007)	1 km
	Maximum elevation	Elev_max		1 km
	Proportion of slope < 30%	Slope_below_30perc		100 m
	Proportion of slope > 30%	Slope_above_30perc		100 m
	Proportion of aspect N	Aspect_N		100 m
	Proportion of aspect E	Aspect_E		100 m
	Proportion of aspect S	Aspect_S		100 m
	Proportion of aspect W	Aspect_W		100 m
	Average roughness	Roughness_avg		1 km
	Maximum roughness	Roughness_max		1 km
Environmental – Land cover	Proportion Forest	Forest	Corine Land Cover 2000 (European Commission, 1994; EEA, 1994; EEA-ETC/TE, JRC, 2002)	100 m
	Proportion Shrubland	Shrubland		
	Proportion Grassland	Grassland		
	Proportion Other Natural Areas	Other_natur_areas		
	Proportion Land with agriculture and natural vegetation	Land_agric_natur		
	Proportion Agricultural land	Agriculture		
	Proportion Wildland-Urban Interface	WUI		
Environmental – Climatic	Average temperature (fire season)	T_{avg}	WorldClim (normals 1961–1990) (Hijmans et al., 2005)	1 km (30 arc s)
	Minimum temperature (fire season)	T_{min}		
	Maximum temperature (fire season)	T_{max}		
	Cumulative precipitation fire season	Total_prec_fireseason		
	Cumulative precipitation other season	Total_prec_nofireseason		
	Average Relative Humidity (fire season)	RH	Climate Research Unit (CRU) and Tyndall Centre (normals 1961–1990) (New et al., 2002)	18.5 km (10′)
	Average soil moisture anomaly (fire season)	Soil_moisture_anom	JRC, Soil Moisture Archive (2010)	5 km
Infrastructure	Density highways	Dens_highways	Tele Atlas (2007) (level 00)	1/100.000
	Density main roads	Dens_main_roads	Tele Atlas (2007) (levels 01–03)	1/100.000
	Density local roads	Dens_local_roads	Tele Atlas (2007) (levels 04–06)	1/100.000
	Density railways	Dens_railways	Tele Atlas (2007)	1/100.000
	Density electric stations	Dens_electric_stations	Platts, McGraw-Hill Research and Analytics, USA (2006)	Vector
	Density electric lines	Dens_electric_lines	Platts, McGraw-Hill Research and Analytics, USA (2006)	Vector
Demographic	Average population density	Popdens_avg	Gallego (2010)	100 m
	Maximum population density	Popdens_max	Gallego (2010)	100 m
	Proportion of high urban density area	High_urban_dens	EUROSTAT, GISCO (2001)	1:3 million
	Proportion of intermediate urban density area	Intermed_urban_dens	EUROSTAT, GISCO (2001)	1:3 million
	Proportion of low urban density area	Thinly_urban_dens	EUROSTAT, GISCO (2001)	1:3 million
Socio-economic	Average unemployment rate 2000–2007	Unemployment	EUROSTAT Regional Statistics (2010, annual data)	NUTS3
	Density of livestock	Livestock_dens	Farm Structure Survey, EUROSTAT (2000)	NUTS3

of the 10 km pixels occupied by each land cover type was extracted from the Corine map. Then, a new raster layer was created per land cover type, with the value of each cell representing the proportion of the pixel area occupied by that specific land cover. Seven new layers were created, one per land cover category.

2.2.3. Climate

Weather conditions are known to affect fuel accumulation and moisture (e.g. Syphard et al., 2008; Vilar et al., 2010), thus having an effect on the probability of a fire to occur. Considering the temporal scale of our study, climatic variables derived from averages of weather conditions over a period of at least 10 years, were used. Temperature, relative humidity and precipitation values were retrieved from the databases mentioned above (Table 2) and transformed at 10 km resolution. Temperature and relative humidity were considered for the main fire season months. Precipitation values, on the other hand, were divided by fire season (June–September) and off-season. It was hypothesized that also the precipitation occurring outside the fire season may affect fire occurrence by favouring seasonal growth of vegetation resulting in an increase availability of fine fuels (notably in grasslands), where fires can more easily start and spread during the main fire season; on the contrary, precipitation during the fire season may hinder fire occurrence by increasing fuel moisture content and limiting fire ignition and spread (Bravo et al., 2010; Drever et al., 2008; Moreno et al., 2011; Pausas, 2004; Pereira et al., 2005). Soil moisture anomaly was included because it is related to the plant physiological activity and is an indicator of drought (Laguardia and Niemeyer, 2008). The moisture content of live fuels is influenced by soil moisture (Chuvieco et al., 2004) and, thus, it affects the amount of heat required for plants to ignite (Bartsch et al., 2009). The soil moisture anomaly was obtained as the difference of the average value for the period 2000–2007 (for the fire season) in relation to the long-term average 1990–2004. Values below zero mean wetter than average, while values above zero mean drier than average (JRC, 2010).

2.2.4. Infrastructures

Roads represent the accessibility to the areas where fires can occur. Road density and distance to roads have been pointed out as important factors in fire occurrence studies (Catry et al., 2009; Martínez et al., 2009; Romero-Calcerrada et al., 2008; Vilar et al., 2010). The road network data from Tele Atlas (Tele Atlas, 2007) was used to calculate road density, defined as road length per unit area. Road density was calculated per grid cell at 10 km resolution, dividing Tele Atlas data in three levels: highways (00), main roads (01–03) and local roads (04–06). Railway density was also calculated as length per 10 km^2 cell. The same was done for the density of electric lines, while the density of electric stations was calculated as number of stations per grid cell.

2.2.5. Demographic variables

Population density represents the distribution of potential causative agents, considering that fires in Europe are mainly human-caused. Data on population density at European level was obtained from the dasymetric grid of population density disaggregated with Corine Land Cover and point survey data, as described in Gallego (2010), combined at 10 km resolution based both on the average and on the maximum value. The degree of urbanization was also included because it classifies each commune (NUTS5) according to its urban nature (densely, intermediate and thinly), combining population density with total population values and considering the characteristics of surrounding areas (EUROSTAT, 2001). The proportion of each grid cell occupied by each level of urban density was calculated.

Table 3
Corine classes aggregated in land cover categories.

Corine codes	Land cover categories
3.1.1, 3.1.2, 3.1.3	Forest
3.2.2, 3.2.3, 3.2.4	Shrubland
2.3.1, 3.2.1	Grassland
2.4.4, 3.3.3, 3.3.4	Other natural areas
2.1.1, 2.2.1, 2.2.2, 2.2.3, 2.4.1, 2.4.2	Agricultural land
2.4.3	Land with agriculture and natural areas
1.1.2	Wildland–Urban Interface

Based on expert's knowledge, the discontinuous urban surfaces class (code 1.1.2) was considered to represent the Wildland–Urban Interface (WUI) in Europe.

2.2.6. Socio-economic variables

Unemployment rate has been previously found in the European Mediterranean environment as a contributing factor to fire occurrence and fire risk (Ferreira de Almeida and Vilaça e Moura, 1992; Leone, 1999; Martínez et al., 2009; Sebastián-López et al., 2008; Velez, 2000), even though the reasons for this association are not clear. In some interpretations it is considered either as a generic indicator of social conflict, which in turn would increase vandalisms in country areas, or as affecting deliberate fire occurrence due to the enhanced opportunities of seasonal employments on fire prevention and fighting. Unemployment rate at NUTS3 level was obtained from the regional statistics of Eurostat (2010); when the values at NUTS3 level were inexistent, the dataset was completed with the values of the correspondent NUTS2 for the year. The average rate between 2000 and 2007 was calculated and was assigned to each cell that falls into the respective NUTS3 area. The density of livestock refers to the ratio of the number of livestock units per hectare of Utilisable Agricultural Area (UAA) and it is a proxy of agricultural intensification in animal husbandry (Eurostat, 2000). Livestock density was referred to 10 km resolution by calculating the UAA and the number of livestock units per grid cell according to the NUTS3 where it belongs to.

2.3. Models

The exploratory analysis of the data revealed unequal variances in the dependent variable. Different transformations were tested in order to obtain a normal distribution in the residuals, as required by the Multiple Linear Regression model. As a result, a square root transformation was applied to the fire density values. The predictors were assessed for multicollinearity using the Pearson's correlation coefficient; the threshold of 0.75 was applied as criteria for the removal of one of the correlated variables.

The original dataset was randomly divided into calibration (60%) and validation (the remaining 40%) samples; this procedure was repeated five times applying a sampling with replacement method, thus obtaining five random sub-samples of the data, each one with a calibration and a validation dataset. Each sub-sample was composed of a high number of observations (ca. 9500 for calibration and approximately 6500 for validation) and each grid cell of 10 × 10 km corresponds to one observation. Subsequently, two different methods of analysis were applied: Multiple Linear Regression (LR) and Random Forest (RF), using the R Statistical Software (R Development Core Team, 2010).

Regression techniques have been widely applied in fire modeling, such as linear regression (e.g. Keeley et al., 2005; Syphard et al., 2007; Sebastián-López et al., 2008) or logistic regression (e.g. Catry et al., 2009; Martínez et al., 2009). Usually these models are built with the goal of using the fewest predictors to explain the greatest variability in the response variable (Graham, 2003). There are several approaches to select the most relevant predictors in regression models, such as stepwise procedures, Akaike's information criterion, Schwarz's Bayesian information criterion or F Statistics (Murtaugh, 2009). In our study, we applied Multiple

Regression to each training sample, creating five intermediate models. In order to be included in the final model, the predictors had to fulfill the criteria of having a p value < 0. 05 and a t value higher than two in at least three of the five intermediate models. To validate the intermediate models, each of them was tested with the corresponding validation sample set aside at the beginning of the analysis. Then, the correlation coefficient was calculated between the observed and the estimated values in the validation samples, assuming that the predictive capacity of the model would be good if there would be a significant correlation between the observed and the predicted values in the independent sample.

The final model was built with the variables selected in the previous step and applied to the complete dataset. The contribution of each variable to the regression model was assessed by means of relative importance measures calculated with the package Relaimpo of the R software (Gromping, 2006). The metrics "lmg" (Lindeman et al., 1980) was selected because it represents the R^2 contribution of each variable averaged over orderings among regressors, i.e., the results of this metrics are not dependent on the order of the predictors in the model.

The exploratory analysis of the results revealed, however, nonlinear trends. Furthermore, considering the large study area, it may be expected that the same variables would operate differently depending on the location (Prasad et al., 2006), thus traditional parametric methods might not provide satisfactory results.

The second proposed method was applied to the same data: Random Forest, a nonparametric technique derived from classification and regression trees (CART). Previous studies (Amatulli et al., 2006; Lozano et al., 2008; McKenzie et al., 2000) highlighted the potential capability of CART for fire risk prediction. RF consists of a combination of many trees, where each tree is generated by bootstrap samples, leaving about a third of the overall sample for validation (the out-of-bag predictions – OOB). Each split of the tree is determined using a randomized subset of the predictors at each node. The final outcome is the average of the results of all the trees (Breiman, 2001; Cutler et al., 2007). This method has been applied in ecological studies (Cutler et al., 2007; Prasad et al., 2006), showing high accuracy and the ability to model complex interactions between variables. In addition, since it uses the OOB samples (independent observations from those used to grow the tree) to calculate error rates and variable importance, no test data or cross-validation is required. However, this method behaves as a "black box" since the individual trees cannot be examined separately (Prasad et al., 2006) and it does not calculate regression coefficients nor confidence intervals (Cutler et al., 2007). Nevertheless, it allows the computation of variable importance measures that can be compared to other regression techniques (Gromping, 2009).

In order to maintain a similar procedure between the two methods, RF was applied to each subset of samples, using the training datasets. Even though independent validation samples are not required in RF, they provide the opportunity to assess the generalization capability of this method (Cutler et al., 2007), thus they were maintained. To run the model, it was necessary to define *a priori* two essential parameters: the number of variables to try at each split (*mtry*) and the number of trees to run (*ntree*). The parameter *mtry* was found via the internal RF function TuneRF; this function computes the optimal number of variables starting from the default (total number of variables/3 for regression) and it looks below and above this threshold for the value with the minimum OOB error rate. Breiman (2001) and Liaw and Wiener (2002) mentioned that even a *mtry* of 1 can produce good accuracy, while Gromping (2009) refers the need to include at least two variables to avoid using also the weaker regressors as splitters. In our case, we realized that the increase in values of *mtry* would result in a higher predictive performance of the model and the attribution of higher importance to fewer variables. Even though the ranking of the variables in terms of importance doesn't change significantly with different *mtry*, as it was also found by Cutler et al. (2007), we used the independent samples instead to find *mtry* and applied it to the training samples to run RF, to avoid potential overfitting. The *ntree* parameter was set to 1000, to obtain stable results. The selection of the most relevant variables to include in the final model was done by ranking the variables according to their importance and excluding the least important ones in all the samples. The variable importance measure available in RF was used for this procedure, namely the mean decrease in accuracy (% IncMSE), considered a more reliable measure than the decrease in node impurity (Genuer et al., 2010); this measure corresponds to the difference between the misclassification rate for the original and the permuted out-of-bag samples, averaged over all the trees and divided by the standard deviation of the differences. The validation procedure was the same used for the Multiple Regression analysis, by calculating the correlation values between the observed and the predicted values in the independent samples.

The final models were built with the smaller set of variables previously selected. The contribution of each variable for the model was assessed via the importance measures already described for each method and the results compared based on literature. The goodness-of-fit of the models was evaluated by comparing the observed with the predicted fire density values and maps of fire probability were created with the normalized values of the predicted fire density.

2.4. Analysis of the spatial autocorrelation in the residuals

Spatial autocorrelation was analyzed with the objective to determine whether the final models take into account the spatial structure of the dependent variable. It was assumed that, if no autocorrelation remained in the residuals of the regression models, then the spatial pattern observed in the dependent variable could be explained by the spatial pattern observed in the predictors (Dormann et al., 2007; González-Megías et al., 2005; Legendre and Legendre, 1998). We tested for the presence of spatial autocorrelation in the residuals of both models by building semivariograms, which plot the semivariance as a function of distance.

3. Results

3.1. Dependent variable – fire density

The distribution of fire density is irregular in the study area (Fig. 1). An extended cluster of fire densities reaching over 1.7 fires per km^2 on average per year is prevalent in northwestern Spain, northern and central Portugal. In the northern parts of France and Italy and some parts of Greece fires occurred rarely in the study period.

3.2. Predictors

After testing for multicollinearity, the following predictors were excluded from further analysis: average temperature, minimum temperature, average elevation, slope above 30 percent, average roughness and intermediate urban density.

3.3. Multiple Linear Regression

The intermediate models created by LR selected 19 variables; the climatic variables showed generally the highest t values, followed by livestock density, density of local roads and thinly urban dense areas. In Table 4, the contribution of each variable to the model, measured by lmg metrics, is given. The adjusted R^2

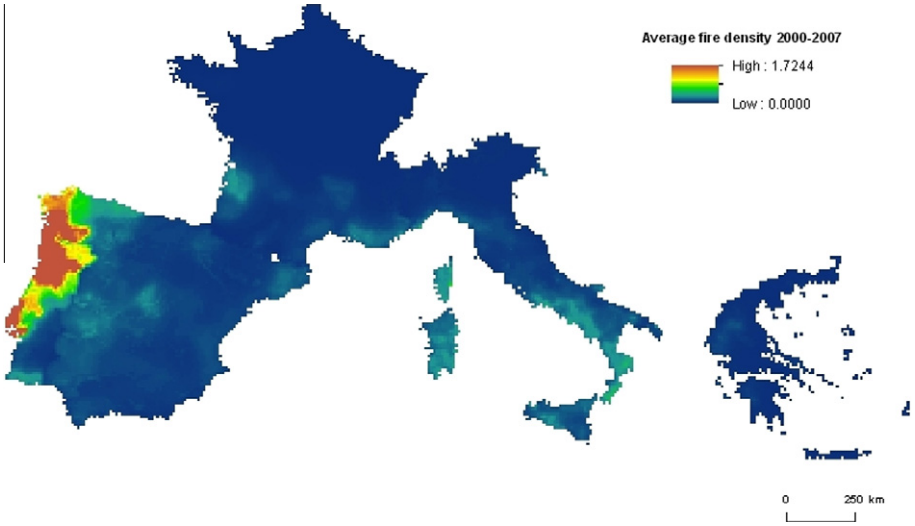

Fig. 1. Yearly average fire density in EUMed between 2000 and 2007 during the main fire season.

obtained for each testing sample and the correlation value between the observed and the predicted values are presented in Table 5. The R^2 values were between 0.45 and 0.46 and the correlation values were all above 0.68.

3.4. Final LR model

The final model was initially built with the 19 variables selected (Table 4). The percentage of variance explained with this model was 44% (adjusted $R^2 = 0.445$, residual standard error = 0.0798 on 16077 *df* and *F*-statistic = 681.3). All the variables appeared as highly significant ($p < 0.000$), except maximum elevation, significant at 0.10 level. The precipitation variables were the most relevant, corresponding to 70% of the total contribution. On the contrary, the majority of the topographic variables showed a very low contribution to the model (less than 1% each), as well as population density, density of main roads and WUI cover (Table 5). Due to the minor importance attributed to some of the variables, LR was applied again using only the highly significant variables with more than 1% relative contribution to the model (in total 11 variables), to verify if the accuracy would be maintained even when using a smaller number of variables. The adjusted R^2 obtained was similar (adjusted $R^2 = 0.439$, residual standard error = 0.08025 on 16085 *df* and *F*-statistic = 1146), showing that, in this case, a more parsimonious model does not affect its performance.

The order of the variables according to their relative contribution to the model didn't change substantially (Table 6), with the

Table 5
Adjusted R^2 and correlation values between observed (obs) and predicted (pred) values in the Linear Regression intermediate models.

	Adjusted R^2	Correlation obs *vs* pred
Sample1	0.458	0.685
Sample2	0.460	0.680
Sample3	0.451	0.691
Sample4	0.453	0.691
Sample5	0.453	0.691

Table 4
Variables selected by the intermediate models using Multiple Linear Regression, by descending order of contribution to the model measured by percentage lmg (metrics were normalized to sum 100%). The range of p values in the 5 samples and the number of samples where each variable was significant ($p < 0.05$) are presented. The direction of association between each predictor and the dependent variable is also presented (+, positive association; −, negative association).

Variables	*p* Value min	*p* Value max	No. samples signif	Direction	lmg (%)
Total_prec_nofireseason	0.0000	0.0000	5	+	48.193
Total_prec_fireseason	0.0037	0.1239	3	−	22.151
Soil_moisture	0.0000	0.0000	5	+	5.339
Livestock_dens	0.0000	0.0000	5	+	4.986
T_{max}	0.0000	0.0000	5	−	3.868
RH	0.0000	0.0000	5	−	3.283
Thinly_urban_dens	0.0000	0.0000	5	−	2.624
Dens_local_roads	0.0000	0.0000	5	+	1.888
Unemployment	0.0000	0.0010	5	−	1.293
Dens_highways	0.0001	0.0053	5	+	1.284
Elev_max	0.0000	0.0037	5	−	1.137
Aspect_W	0.0000	0.0035	5	+	1.077
WUI	0.0000	0.0014	5	−	0.728
Dens_main_roads	0.0000	0.0000	5	−	0.637
Aspect_N	0.0009	0.1133	3	+	0.463
Popdens_avg	0.0000	0.0000	5	−	0.398
Slope_below_30perc	0.0000	0.1080	3	−	0.352
Aspect_S	0.0000	0.0000	5	+	0.162
Aspect_E	0.0000	0.0001	5	+	0.139

precipitation variables being the most important ones, followed by soil moisture anomaly and livestock density.

3.5. Random Forest

The *mtry* parameter, i.e., the number of predictors used at each split, was set to 20, according to the results obtained from the tuning function of this method (section 2.3). The intermediate models created with RF show a proportion of explained variance between 93% and 95%, both in the training and testing datasets (Table 7). The correlation between the observed and the predicted values in the testing datasets is above 0.97 for all the samples.

The variable importance plots for the five samples show a threshold of ca. 20% IncMSE above which the variables assume higher importance (Fig. 2). The average % IncMSE was calculated between all the samples and the variables that showed an average IncMSE of 20% or higher were selected for the final model (Table 8).

3.6. Final RF model

The final model was built with the 12 variables shown in Table 8. The percentage of variance explained with this model was 96.3% (mean of squared residuals = 0.00043). In this model, the most important variables according to the values of % IncMSE were, by descending order: off-season precipitation, unemployment, soil moisture, fire season precipitation, density of local roads, livestock density, maximum elevation, maximum roughness, shrubland, relative humidity, maximum temperature and grassland.

Plots of the observed and the predicted values for each testing sub-sample were created; since the resulting plots are similar to all the sub-samples of each model, an example is presented in Fig. 3 for each model. In the case of LR, the distribution of both the observed and predicted values is skewed to the left due to the higher proportion of low fire density pixels. It is also evident that the predicted values of fire density are underestimated when the observed values are above the threshold of 0.4. The RF model shows a better fit. Underestimation of the higher fire density values can still be observed, but significantly less than in LR.

3.7. Likelihood of fire occurrence

Maps with the likelihood of fire occurrence were created by normalizing the predicted values in both models (Fig. 4). The map obtained with LR shows higher likelihood for fire occurrence in western Iberian Peninsula, southwestern Italy and southwestern Greece. Surprisingly, northwestern France appears with intermediate values, as well as the mountainous areas of the Pyrenees and the Alps. The map obtained with RF shows the highest values more concentrated in the northwestern Iberian Peninsula and southern Italy, including the islands. The south of France, the interior of Spain and central-western Greece show intermediate values.

In order to evaluate the goodness-of-fit of each model, maps of the residuals were created and compared (Fig. 5). In LR, the higher values of the residuals are concentrated in large areas of the northwest of the Iberian Peninsula, northeast of Italy and in parts of the islands, meaning that the model underestimated fire density for these areas. On the other hand, the LR model also overestimated fire density in southern Greece and northwest France, shown by the negative values of the residuals. In general, the RF model shows lower residuals values and a better fit in the estimation of fire density. Most of the Spanish and French territories present very low values in the residuals. Although this model also underestimates fire density in the northwest of the Iberian Peninsula, similarly to LR, the pattern is spatially more limited and covers smaller areas than in LR. In the RF model, 13% of the cells showed zero residuals, while in the LR model there were only 1% of the cells with zero value for the residuals.

Table 6
Results of the Linear Regression model with the selected 11 variables

Variables	Estimate	Std. error	*t* Value	Pr(>|t|)	Img(%)
(Intercept)	0.346	0.014	23.888	0.000	
Total_prec_nofireseason	0.000	0.000	80.009	0.000	51.191
Total_prec_fireseason	−0.001	0.000	−58.347	0.000	22.531
Soil_moisture	0.070	0.003	22.859	0.000	5.771
Livestock_dens	0.013	0.001	13.925	0.000	5.068
T_{max}	−0.008	0.000	−24.385	0.000	4.603
RH	−0.002	0.000	−16.188	0.000	3.648
Thinly_urban_dens	0.000	0.000	−11.502	0.000	2.452
Unemployment	−0.002	0.000	−9.663	0.000	1.367
Dens_local_roads	0.021	0.001	16.129	0.000	1.271
Aspect_W	0.001	0.000	8.921	0.000	1.056
Dens_highways	0.112	0.015	7.266	0.000	1.042

3.8. Spatial autocorrelation

Fig. 6 shows the semivariogram built for the dependent variable and the residuals of both models. The dependent variable shows evidence of spatial autocorrelation up to 1500 km lag. Between 2500 and 3300 km, the semivariance increases again, representing the discontinuous area of Greece, which was, for this reason, disregarded. The RF model seems to incorporate much better the effects of spatial autocorrelation than the LR model, evidenced by the flat semivariance line throughout most of the distance and by the lower values of semivariance obtained (which were multiplied by five in order to be visible at the scale of the graph).

4. Discussion

Fire density distribution in the EUMed region shows irregular patterns; the results of our study suggest that this distribution is influenced by both physical and human factors and that non-linear trends exist, thus a non-parametric method is considered suitable for modeling fire occurrence. Previous studies also evidenced the existence of non-linear relationships between fire ignition and the independent variables (Syphard et al., 2007; Vilar et al., 2010). The comparison between the results obtained with Multiple Linear Regression and Random Forest showed that RF had a much higher predictive ability, while LR showed a positive but weak relationship between the dependent variable and the predictors.

In their final form, both models in their final form included a smaller number of variables selected from the original set of 32; LR included 11 variables and RF 12. Other authors also stated that a parsimonious model would be more stable and easier to generalize (Catry et al., 2009; Vilar et al., 2010), particularly at a broad spatial scale. When studying the key factors determining fire occurrence at this extent it is crucial to bear in mind that environmental and social conditions differ from region to region and the same variables may operate differently depending on the location and the scale of analysis (Prasad et al., 2006).

There were eight variables included in both models, evidencing their importance to fire density distribution, independently of the method used. The most important variable identified in both models was the off-season precipitation (*total_prec_nofireseason*). This is most likely related to the positive influence of spring rainfall in vegetation growth and fuel accumulation. Fire season precipitation also played a crucial role. The relationship here is negative: the drier it is during the fire season, the likelier fires occur. A strong influence of precipitation variables on fire occurrence has been shown previously: Bravo et al. (2010) found that rainfall was

Table 7
Results of the intermediate models created with Random Forest, including the correlation values between observed (obs) and predicted (pred) values.

	% Variance explained		Mean squared residuals	Correlation obs *vs* pred
	Calibration sample	*Validation sample*		
Sample 1	94.68	94.38	0.0006551101	0.9716
Sample 2	94.2	95.08	0.000689778	0.9767
Sample 3	94.92	94.48	0.000578382	0.9727
Sample 4	94.29	93.96	0.0006233862	0.9732
Sample 5	94.87	94.41	0.0005946975	0.9732

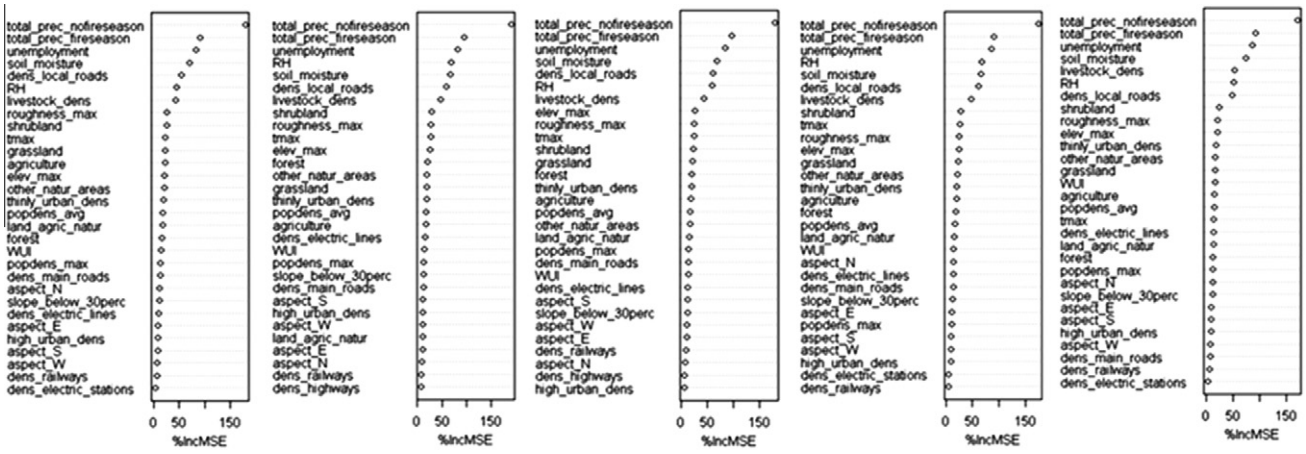

Fig. 2. Plots of the variable importance measure (% Increase MSE) for the five samples.

Table 8
Variables included in the final model using Random Forest, in descending order of importance based on average % IncMSE (mean decrease in accuracy) from the five samples.

Variables	Avg % IncMSE
Total_prec_nofireseason	179.286
Total_prec_fireseason	93.315
Unemployment	84.375
Soil_moisture	69.502
RH	58.150
Dens_local_roads	56.690
Livestock_dens	47.038
Shrubland	26.061
Roughness_max	25.050
Elev_max	23.625
T_{max}	22.917
Grassland	20.276

positively correlated to fire frequency in Argentina, because it influenced the production of fine fuels for fire. Drever et al. (2008) also found that the variables best explaining fire occurrence in the Great Lakes-St. Lawrence forest in Canada were a combination of antecedent-winter precipitation and fire season precipitation deficit/surplus. In European Mediterranean areas, Pausas (2004) analyzed the relation between fire and climate for a period of 50 years and suggested that high rainfall may increase fuel loads that burn in the successive years; Moreno et al. (2011) mentioned that the regeneration of shrubland-type vegetation was correlated with precipitation in the fall and winter immediately after the fire, which in turn contributes to fuel accumulation that may burn in successive fire seasons.

Soil moisture anomaly was also one of the most important variables, assuming the 3rd position in both models. The positive relationship suggests that an ignition is more likely to occur when the soil surface layer is drier, since the moisture content of fuels is directly affected by soil moisture (Chuvieco et al., 2004). Soil moisture is relevant because of its effect on dead fuels on the ground and because it is a proxy for drought. As pointed out also by Bartsch et al. (2009), fuel is more easily ignited under low surface wetness conditions.

Unemployment is identified in the RF model as the second most important variable, while in the LR model it comes in 8th position and with a negative and weak relationship. This may mean that a non-linear association between unemployment and fire density exists and, as such, it was better recognized by the RF model, even though it is not possible to know from the RF model the direction of the association. Unemployment may be acting as a proxy for economic depression, which, in rural areas, is often associated with land abandonment; in the Mediterranean region, where fires are mostly human-caused, it is also an indicator of potential social conflicts, which may in turn be the driver for motivations of deliberate ignitions (arson) in specific regions (Ferreira de Almeida and Vilaça e Moura, 1992; Leone, 1999; Velez, 2000). Sebastián-López et al. (2008) found that unemployment was related to fire only for the people aged below 25 years old, for which no explanation was provided. Martínez et al. (2009) included unemployment as a potential fire predictor but could not explain the association either.

Livestock density comes in 4th position in LR and in 6th position in RF. The presence of livestock has also been identified by previous authors as being positively correlated to fire occurrence (Koutsias et al., 2010; Martínez et al., 2009). Agricultural activities, such as land burning for pasture renovation, are known to cause fires that spread to shrubland and forested areas nearby. Romero-Calcerrada et al. (2008), on the other hand, found that the presence of goats and sheep was related to the absence of ignitions in Central Spain, probably because these species feed on grass and bushes and thus reduce the accumulation of fine fuels. The same was found by Sebastián-López et al. (2008), which differs from our results.

The other climatic variables, T_{max} and RH, come next in variable importance in the LR model, while in RF their position is lower but still important. All the climatic variables were included in both

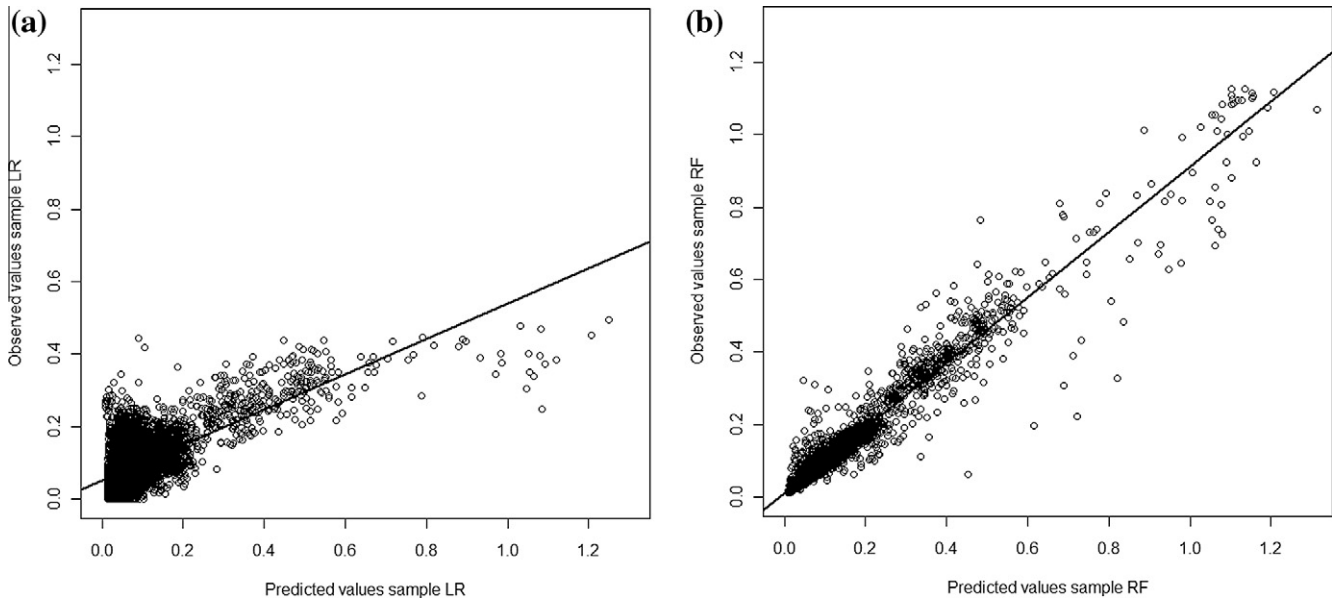

Fig. 3. (a and b) – Plots of the observed and the predicted values calculated by Multiple Linear Regression (a, on the left) and Random Forest (b, on the right) for one of the samples.

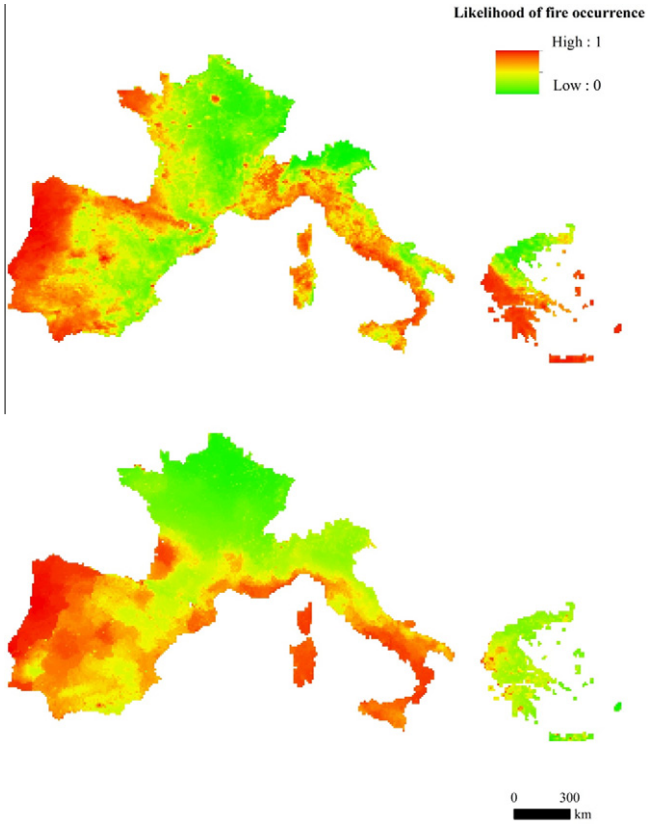

Fig. 4. Maps of the likelihood of fire occurrence obtained from Linear Regression (on top) and Random Forest (bottom) models.

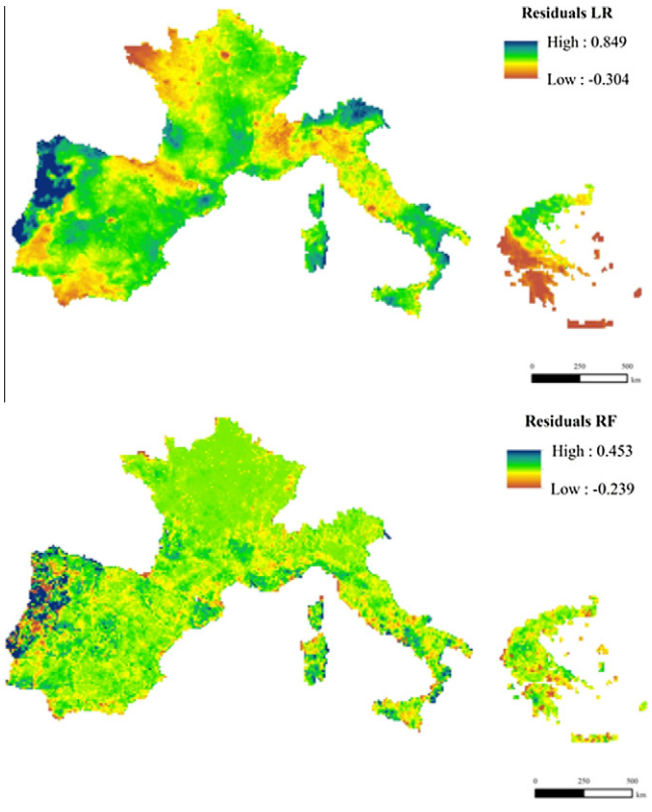

Fig. 5. Maps of residuals obtained from Linear Regression (on top) and Random Forest (bottom) models.

models and these results point out that these factors are, in general, good predictors of fire occurrence. Drever et al. (2008) also mentioned the strong influence of climate on fire occurrence, even though at a smaller scale. The majority of fires occurred with temperatures above 20 °C; the negative association that appears in the LR model between temperature and fire density may be due to the existence of a non-linear and more complex relation between these variables not revealed by a linear model, most likely associated with the human causes of fire ignitions in this region, thus not totally dependent on temperature.

Density of local roads was found to be important in both models. In the LR model, density of highways was also significant, while in RF this variable was discarded. The influence of the density of local roads is to be expected, since these roads connect peri-urban

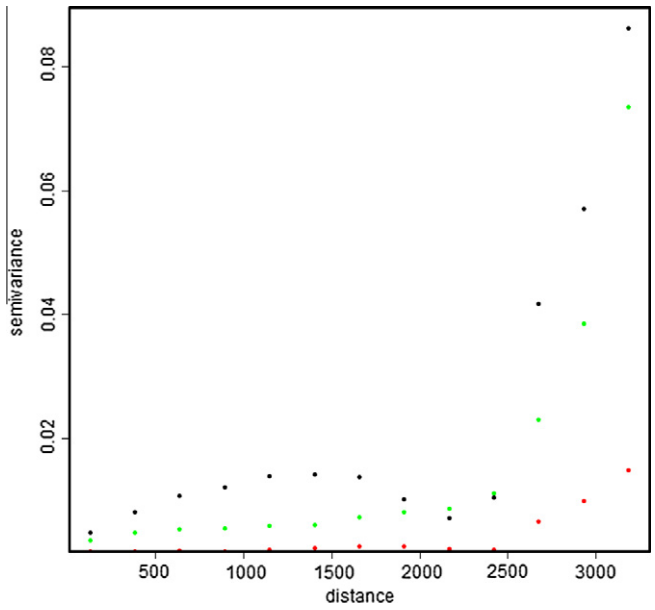

Fig. 6. Semivariance plotted as a function of distance (km) for the dependent variable (black) and the residuals of both models (LR in green and RF in red, the values of the latter multiplied by five).

areas and give access to the rural land and forested areas. Other studies mentioned the association of roads to fire occurrence; Romero-Calcerrada et al. (2008) and Syphard et al. (2008) pointed out the distance to roads to be one of the explanatory variables of fire occurrence in central Spain and southern California, respectively. In Portugal, Catry et al. (2009) and Vasconcelos et al. (2001) also found that distance to roads decreased the number of fire ignitions, the same being mentioned by Vilar et al. (2010) in relation to road density in the region of Madrid, Spain.

The LR model included also thinly populated areas and the proportion of aspect W as significant variables, but their association to fire density is not clear. Sebastián-López et al. (2008) found aspect to be very poorly related to fire occurrence, being more likely related to fire behavior and burned areas. In fact, aspect is not included in the RF model, even though this model includes other topographic variables: maximum elevation and maximum roughness. Maximum elevation was present in the first LR model (with 19 variables), having a negative association with fire density, but it was later excluded because of the low significance for the model. It is well acknowledged that elevation dampers fire occurrence, since in high altitude areas human activity and occupation are reduced, thus decreasing the likelihood of human-caused ignitions. In addition, the effects of altitude in weather conditions, vegetation cover and soil moisture are less favorable to fire occurrence as altitude increases (González et al., 2006; Sebastián-López et al., 2008; Vilar et al., 2010). The effect of roughness is not entirely clear, but the negative association found between roughness and fire density suggests that fire is more likely to occur in flatter areas.

The RF model includes as well two variables representing land cover: proportion of shrubland and grassland, which were not present in the LR model. The proportion of shrubland-type fuel was previously identified by Sebastián-López et al. (2008) as the variable with the highest predictive power in their model for Southern Europe, justified by the fact that shrubland is usually the vegetation type most affected by fire in Mediterranean areas. In studies related to the characterization of the land cover types more prone to fire, it was also found that shrublands are more fire-prone than other land cover types in Mediterranean areas (Moreira et al., 2001, 2009; Mouillot et al., 2003; Nunes et al., 2005). Likewise, grasslands are mainly composed by fine fuels and are easy to ignite, thus their relation with fire occurrence is somewhat expected.

These results provide valuable insights on the potential causes of the spatial distribution of fire events throughout the EUMed region. In spite of the strong human influence on fire occurrence in Mediterranean Europe, the physical variables – particularly those related to climatic conditions – should be included in the analysis since they form the natural setting that favors or hampers human actions. Fire management strategies should focus on the areas where the combination of specific climatic conditions, particularly precipitation before and during the main fire season, is favorable to fire occurrence. In the areas where livestock and local roads density is high and where shrubland and grassland-type vegetation are common, preventive measures should be applied. Finally, our results suggest that socio-economic problems such as unemployment should be considered in the implementation of preventive actions.

5. Conclusions

The EUMed region is the area of Europe with highest fire incidence. Inside the boundaries of the region, however, fire density has an irregular distribution in space and time. The probability of a fire to occur depends on the interactions between the physical and human variables that affect the ignition and spread of a fire. In this study, the likelihood of fire occurrence was modeled using two different methods: Multiple Linear Regression and Random Forest. The comparison of the results obtained with these two methods allowed for the examination of non-linear relationships between the variables, not assumed in Multiple Linear Regression, and the investigation of the potentialities of the RF method in fire occurrence modeling. Moreover, both methods ranked the variables according to their relative contribution to the model, allowing for the identification of the common factors in both models and, thus, emphasizing their significance in explaining fire density distribution.

The two models showed distinct results; the RF model showed higher predictive accuracy than LR, reflecting the existence of non-linear trends. Moreover, spatial autocorrelation in model residuals was reduced to a much higher degree with the RF model. In spite of these differences, both models identified northwestern Iberia and southern Italy as areas with high fire density, while northern France, northeastern Italy and northern Greece were identified as low fire density areas. Furthermore, it was also possible to identify common significant variables, providing important insights to better understand the factors affecting fire occurrence in this region, during the fire season.

At the European level, the lack of comparable data and the complexity of the factors that influence fire occurrence and risk have been pointed out as limitations for the systematic investigation of long-term fire occurrence and risk at this broad scale. Nevertheless, the importance of this assessment for fire prevention and management encourages the further development of research in this area. Future work could focus on the variations in the predictors at different spatial and temporal scales, in order to assess the consistency of the explanatory power of the variables throughout the whole study area. The same methods could be applied to other European regions with the same type of variables, to understand the implications of regional differences in the level of importance of the variables selected and in the overall predictive ability of the models.

References

Allgower, B., Camia, A., Francesetti, A., Koutsias, N., 2005. Fire hot spot areas in Southern Europe – detection of large-scale wildland fire occurrence patterns by

adaptive kernel density interpolation. In: de la Riva, J., Perez-Cabello, F., Chuvieco, E. (Eds.), Proceedings of the 5th International Workshop on Remote Sensing and GIS Applications to Forest Fire Management: Fire Effects Assessment. Univ. Zaragoza, Spain.

Amatulli, G., Rodrigues, M.-J., Trombetti, M., Lovreglio, R., 2006. Assessing long-term fire risk at local scale by means of decision tree technique. Journal of Geophysical Research 111 (G04S05), 15.

Amatulli, G., Perez-Cabello, F., de la Riva, J., 2007. Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. Ecological Modelling 200, 321–333.

Bachmann, A., Allgöwer, B., 2000. The need for a consistent wildfire risk terminology. In: Neuenschwander, L.F., Ryan, K.C., Gollberg, G.E., Greer, J.D. (Eds.), Proceedings of the Joint Fire Science Conference and Workshop: Crossing the Millenium: Integrating Spatial Technologies and Ecological Principles for a New Age in Fire Management, Boise Idaho, June 15–17, 1999, University of Idaho (2000), pp. 67–77.

Bartsch, A., Baltzer, H., George, C., 2009. The influence of regional surface soil moisture anomalies on forest fires in Siberia observed from satellites. Environmental Research Letters 4, 045021.

Bravo, S., Kunst, C., Grau, R., Araoz, E., 2010. Fire-rainfall relationships in Argentine Chaco savannas. Journal of Arid Environments 74, 1319–1323.

Breiman, L., Meisel, W., Purcell, E., 1977. Variable kernel estimates of multivariate densities. Technometrics 19 (2), 135–144.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Burgan, R.E., Klaver, R.W., Klaver, J.M., 1998. Fuel models and fire potential from satellite and surface observations. International Journal of Wildland Fire 8 (3), 159–170.

Camia, A., Durrant Houston, T., San-Miguel-Ayanz, J., 2010. The European fire database: development, structure and implementation. In: Viegas, D.X. (Ed.), Proceedings of the VI International Conference on Forest Fire Research, Coimbra, Portugal.

Catry, F.X., Rego, F.C., Bação, F.L., Moreira, F., 2009. Modelling and mapping the occurrence of wildfire ignitions in Portugal. International Journal of Wildland Fire 18, 921–931.

Chuvieco, E., Allgöwer, B., Salas, J., 2003. Integration of physical and human factors in fire danger assessment. In: Chuvieco, E. (Ed.), Wildland Fire Danger Estimation and Mapping. The Role of Remote Sensing Data 4, 197–218.

Chuvieco, E., Aguado, I., Dimitrakopoulos, P., 2004. Conversion of fuel moisture content values to ignition potential for integrated fire danger assessment. Canadian Journal of Forest Research 34, 2284–2293.

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

Conedera, M., Torriani, D., Neff, C., Ricotta, C., Bajocco, S., Pezzatti, G.B., 2011. Using Monte Carlo simulations to estimate relative fire ignition danger in a low-to-medium fire-prone region. Forest Ecology and Management 261 (12), 2179–2187.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in Ecology. Ecology 88 (11), 2783–2792.

de la Riva, J., Perez-Cabello, F., Lana-Renault, N., Koutsias, N., 2004. Mapping wildfire occurrence at regional level. Remote Sensing of Environment 92, 288–294.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, C., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres'Neto, P.R., Reineking, B., Schroder, B., Schurr, F.M., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628.

Drever, C.R., Drever, M.C., Messier, C., Bergeron, Y., Flannigan, M., 2008. Fire and the relative roles of weather, climate and landscape characteristics in the Great Lakes-St. Lawrence forest of Canada. Journal of Vegetation Science 19, 57–66.

EEA, European Environmental Agency, 1994. Corine Land Cover report. Available from: <http://www.eea.europa.eu/publications/COR0-landcover>. (accessed in April 2010).

European Commission, 1994. CORINE land Cover Technical Guide. EUR 12585 EN, OPOCE Luxembourg.

European Commission, 2011. Forest Fires in Europe 2010. JRC Scientific and Technical Reports, Report nr. 11. EUR 24910 EN, Italy.

EEA-ETC/TE, JRC, 2002. CORINE Land Cover update: I&CLC2000 Project Technical Guidelines, August 2002.

EUROSTAT, 2000. Farm Structure Survey. Available from: <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>. (accessed in April 2010).

EUROSTAT, 2001. Degree of urbanisation. GISCO database (Geographic Information System of the European Commission). Available from: <http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/popups/references/Population%20Distribution%20-%20Demography>. (accessed in April 2010).

EUROSTAT, 2010. Regional Statistics. Available from: <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>. (accessed in April 2010).

Ferreira de Almeida, A.M.S., Vilaça e Moura, P.V.S., 1992. The relationship of forest fires to agro-forestry and socio-economic parameters in Portugal. International Journal of Wildland Fire 2, 37–40.

Finney, M.A., 2005. The challenge of quantitative risk analysis for wildland fire. Forest Ecology and Management 211 (1-2), 97–108.

FOREST EUROPE, UNECE and FAO, 2011. State of Europe's Forests 2011. Status and Trends in Sustainable Forest Management in Europe.

Gallego, F.J., 2010. A population density grid of the European Union. Population and Environment 31 (6), 460–473.

Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2010. Variable selection using random forests. Pattern Recognition Letters 31, 2225–2236.

González, J.R., Palahí, M., Trasobares, A., Pukkala, T., 2006. A fire probability model for forest stands in Catalonia (north-east Spain). Annals of Forest Science 63, 169–176.

Gonzalez-Megías, A., Gómez, J.M., Sanchez-Piñero, F., 2005. Consequences of spatial autocorrelation for the analysis of metapopulation autodynamics. Ecology 86 (12), 3264–3271.

Graham, M.H., 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84 (11), 2809–2815.

Gromping, U., 2006. Relative importance for linear regression in R: the package relaimpo. Journal of Statistical Software 17 (1).

Gromping, U., 2009. Variable importance assessment in regression: linear regression versus random forest. The American Statistician 63 (4), 308–319.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978.

Jarvis A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture, CIAT. Available from: <http://srtm.csi.cgiar.org>. (accessed in May 2010).

JRC – Joint Research Center of the European Commission, 2010. Soil Moisture Archive, Joint Research Centre, European Commission. Available from: <http://desert.jrc.ec.europa.eu/action/php/index.php?action=view&id=19>. (accessed in May 2010).

Keeley, J.E., Fotheringham, C.J., Baer-Keeley, M., 2005. Determinants of postfire recovery and succession in mediterranean-climate shrublands of California. Ecological Applications 15 (5), 1515–1534.

Koutsias, N., Kalabokidis, K., Allgöwer, B., 2004. Fire occurrence patterns at landscape level: beyond positional accuracy of ignition points with kernel density estimation methods. Natural Resource Modeling 17 (4), 359–375.

Koutsias, N., Martínez, J., Chuvieco, E., Allgöwer, B., 2005. Modeling wildland fire occurrence in Southern Europe by a geographically weighted regression approach. In: De la Riva, J., Pérez-Cabello, F., Chuvieco, E. (Eds.), Proceedings of the 5th International Workshop on Remote Sensing and GIS Applications to Forest Fire Management: Fire Effects Assessment, pp. 57–60.

Koutsias, N., Martínez, J., Allgöwer, B., 2010. Do factors causing wildfires vary in space? Evidence from geographically weighted regression. GIScience & Remote Sensing 47 (2), 1548–1603.

Laguardia, G., Niemeyer, S., 2008. On the comparison between the LISFLOOD modelled and the ERS/SCAT derived soil moisture estimates. Hydrology and Earth System Sciences 12 (6), 1339–1351.

Legendre, P., Legendre, L., 1998. Numerical Ecology. Elsevier, Amsterdam, 853 pp.

Leone, V., 1999. Los incendios en el Mediodıa Italiano. In: Araque Jimenez, E. (Ed.), Incendioshistoricos: una aproximacion multidisciplinar. Universidad Internacional de Andalucıa, Seville.

Leone, V., Lovreglio, R., Pilar Martín, M., Martínez, J., Vilar, L., 2009. Human factors of fire occurrence in the Mediterranean. In: Chuvieco, E. (Ed.), Earth Observation of Wildland Fires in Mediterranean Ecosystems. Springer, p. 251.

Levine, N., 2007. CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations, version 3.1. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington DC.

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2 (3), 18–22.

Lindeman, R.H., Merenda, P.F., Gold, R.Z., 1980. Introduction to Bivariate and Multivariate Analysis. Scott, Foresman, Glenview, IL.

Lozano, F.J., Suarez-Seoane, S., Luis, E., 2008. A multi-scale approach for modeling fire occurrence probability using satellite data and classification trees: a case study in a mountainous Mediterranean region. Remote Sensing of Environment 112 (3), 708–719.

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.

McKenzie, D., Peterson, D.L., Agee, J.K., 2000. Fire frequency in the interior Columbia River Basin: building regional models from fire history data. Ecological Applications 10 (5), 1497–1516.

Moreira, F., Rego, F.C., Ferreira, P.G., 2001. Temporal (1958–1995) pattern of change in a cultural landscape of northwestern Portugal: implications for fire occurrence. Landscape Ecology 16, 557–567.

Moreira, F., Vaz, P., Catry, F., Silva, J.S., 2009. Regional variations in wildfire susceptibility of land-cover types in Portugal: implications for landscape management to minimize fire hazard. International Journal of Wildland Fire 18, 563–574.

Moreno, J.M., Zuazua, E., Pérez, B., Luna, B., Velasco, A., Resco de Dios, V., 2011. Rainfall patterns after fire differentially affect the recruitment of three Mediterranean shrubs. Biogeosciences 8, 3721–3732.

Mouillot, F., Ratte, J.-P., Joffre, R., Moreno, J.M., Rambal, S., 2003. Some determinants of the spatio-temporal fire cycle in a mediterranean landscape, Corsica, France. Landscape Ecology 18, 665–674.

Murtaugh, P.A., 2009. Performance of several variable-selection methods applied to real ecological data. Ecology Letters 12, 1061–1068.

New, M., Lister, D., Hulme, M., Makin, I., 2002. A high-resolution data set of surface climate over global land areas. Climate Research 21, 1–25.

Nunes, M.C.S., Vasconcelos, M.J., Pereira, J.M.C., Dasgupta, N., Alldredge, R.J., Rego, F.J., 2005. Land cover type and fire in Portugal: do fires burn land cover selectively? Landscape Ecology 20, 661–673.

Pausas, J.G., 2004. Changes in fire and climate in the eastern Iberian Peninsula (Mediterranean Basin). Climatic Change 63, 337–350.

Pereira, M.G., Trigo, R.M., da Câmara, C.C., Pereira, J.M.C., Leite, S.M., 2005. Synoptic patterns associated with large summer forest fires in Portugal. Agricultural and Forest Meteorology 129 (1–2), 11–25.

Platts, McGraw-Hill Research and Analytics, USA, 2006. Available from: <http://www.platts.com/Products/gisdata>. (accessed in May 2010).

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199.

R Development Core Team, 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. Available from: <http://www.R-project.org/>. (Accessed in May 2010).

Reuter, H.I., Nelson, A., Jarvis, A., 2007. An evaluation of void-filling interpolation methods for SRTM data. International Journal of Geographical Information Science 21 (9), 983–1008.

Romero-Calcerrada, R., Novillo, C.J., Millington, J.D.A., Gomez-Jimenez, I., 2008. GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid, Central Spain. Landscape Ecology 23, 341–354.

Romero-Calcerrada, R., Barrio-Parra, F., Millington, J.D.A., Novillo, C.J., 2010. Spatial modelling of socioeconomic data to understand patterns of human-caused wildfire ignition risk in the SW of Madrid, Central Spain. Ecological Modelling 221, 34–45.

San-Miguel-Ayanz, J., Carlson, J.D., Alexander, M., Tolhurst, K., Morgan, G., Sneeuwjagt, R., 2003. Current methods to assess fire danger potential. In: Chuvieco, E. (Ed.), Wildland Fire Danger Estimation and Mapping. The role of remote sensing data. World Scientific, Singapore, pp. 21–61.

Sebastián-López, A., San-Miguel-Ayanz, J., Burgan, R.E., 2002. Integration of satellite sensor data, fuel type maps and meteorological observations for evaluation of forest fire risk at the pan-European scale. International Journal of Remote Sensing 23 (13), 2713–2719.

Sebastián-López, A., Salvador-Civil, R., Gonzalo-Jiménez, J., San-Miguel Ayanz, J., 2008. Integration of socio-economic and environmental variables for modelling long-term fire danger in Southern Europe. European Journal of Forest Research 127, 149–163.

Stambaugh, M.C., Guyette, R.P., 2008. Predicting spatio-temporal variability in fire return intervals using a topographic roughness index. Forest Ecology and Management 254, 463–473.

Syphard, A.D., Radeloff, V.C., Keely, J.E., Hawbaker, R.J., Clayton, M.K., Stewart, S.I., Hammer, R.B., 2007. Human influence on California Fire Regimes. Ecological Applications 17, 1388–1402.

Syphard, A.D., Radeloff, V.C., Keuler, N.S., Taylor, R.S., Hawbaker, T.J., Stewart, S.I., Clayton, M.K., 2008. Predicting spatial patterns of fire on a southern California landscape. International Journal of Wildland Fire 17 (5), 602–613.

Tele Atlas NV and Tele Atlas North America, 2007. Tele Atlas MultiNet® Version 3.4.2.1 Data Specification.

Tyndall Centre. Available from: <http://www.cru.uea.ac.uk/~timm/grid/CRU_TS_1_2.html>. (accessed in May 2010).

Vasconcelos, M.J.P., Silva, S., Tomé, M., Alvim, M., Pereira, J.M.C., 2001. Spatial prediction of fire ignition probabilities. Photogrammetric Engineering and Remote Sensing 67 (1), 73–82.

Velez, R., 2000. La prevencion. In: Garcıa-Brage, A. (Ed.), La defensa contra incendios forestales fundamentos y experiencias. McGraw-Hill/Interamericana de Espana, Madrid.

Vilar, L., Woolford, D.G., Martell, D.L., Pilar Martín, M., 2010. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. International Journal of Wildland Fire 19 (3), 325–337.

Whelan, Robert.J., 1995. The ecology of fire. Cambridge University Press, New York, NY, 346 pp.

Widayati, A., Jones, S., Carlisle, B., 2010. Accessibility factors and conservation forest designation affecting rattan cane harvesting in Lambusango forest, Buton, Indonesia. Human Ecology 38, 731–746.