

**ADVANCES IN
FOREST FIRE RESEARCH
2018**

EDITED BY

DOMINGOS XAVIER VIEGAS
ADAI/CEIF, UNIVERSITY OF COIMBRA, PORTUGAL

Wildfire susceptibility mapping in Liguria (Italy). Comparison of Statistical Driven Partitioning and Machine Learning approach

Guido Biondi¹, Lorenzo Campo¹; Mirko D'Andrea¹; Silvia Degli Esposti¹; Paolo Fiorucci^{1*}; Marj Tonini ^{*2}

¹*CIMA Research Foundation. Via A. Magliotto 2, 17100, Savona, {paolo.fiorucci@cimafoundation.org*}*

²*Institute of Earth Surface Dynamics (IDYST), Faculty of Geosciences and Environment, University of Lausanne, Lausanne, Switzerland, {marj.tonini@unil.ch*}*

Abstract

The identification of areas most vulnerable to fire risk is a key tool in wildfire management, particularly in view of the limited availability of fire risk management resources, most of which are used for national and regional air services. The few resources available can thus be used on a yearly basis to mitigate problems in the areas at highest risk by defining a program of interventions.

In this contest, the main objective of the present study is to elaborate wildfire susceptibility mapping for Liguria region (Italy), and to compare results obtained by a statistical driven partitioning models with the ones obtaining by applying a machine learning approach, based on Random Forest (RF) algorithm.

The availability of a mapping of fire perimeters spans almost 20 years (1996-2016), and this, combined with a detailed knowledge of topography, climate and land cover, allowed to understand the main features involved in wildfire occurrences and their behavior. The seasonality of the fire regime was also considered, partitioning the analysis in two macro season (November-April and May- October).

The analysis was based on a recursive-quantiles subdivision of the territory in classes based on the different available information layers: elevation, slope, aspect, rainfall height, temperature (the latter subdivided in winter and summer periods). The algorithm was designed in order to assure the equal representation of each class, in which the number of fires occurred in the period of analysis is considered, in order to have an estimation of the fire hazard with a constant statistical confidence.

The resulting wildfire susceptibility map was then compared with the map obtained applying a machine learning (ML) approach. ML explores the conception of algorithms capable of learning from and make predictions on data, through the modeling of the hidden relationships between a set of input and output variables (i.e. the predisposing factors and the occurrence of the phenomenon). The main benefit is that these approaches are data driven, meaning that they do not need a priori knowledge of the process. Moreover, RF directly provides the measurement of the importance of each variable.

The results show that both the approaches identify shrubs as the vegetation most affected by wildfire both in summer and winter season. RF algorithm seems to perform better, detecting 90% of the of the BA within 10% of the entire territory instead of 60% of burned area within the 20% of the territory for the statistical driven partitioning. These results are preliminary and need to be further evaluated in detail by establishing a working plan accounting for a common validation dataset for the two approaches.

Keywords: Wildfire susceptibility mapping, Machine learning, Random Forest, Statistical partitioning, forest management plan

1. Introduction

Susceptibility maps are a modern tool to support forest protection plans and to address fuel management strategies in order to reduce fires' consequences. Wildfire susceptibility mapping allows to identify areas subject to fires, assessed defining a rank from low to high. These are elaborated taking into account two aspects: where the wildfires occur and which are the predisposing factors. Distance to transport networks and distance to urban or recreation areas are the most frequently used human

factors. As regards environmental variables, those related to weather, fuel and topography result to be the most significant drivers of ignition of wildfires, especially in Mediterranean-type regions (Ganteaume et al., 2013). A number of models for wildfire hazard mapping and risk analyses were recently developed. These are often GIS oriented, including statistical analyses and modeling to assess the variables' importance (Eugenio et al., 2016; Parente and Pereira, 2016; Pourghasemi et al., 2016; Pourtaghi et al., 2016; Rodrigues et al., 2014; Fiorucci et al., 2008), and often rely on expert knowledge to estimate predisposing factors. Comparison of physically based and stochastic approach highlight the benefit of using a data driven method (Leuenberger et al., 2018), which do not need a priori knowledge of the process.

Wildfire risk is particularly significant in Italy, both in summer and winter season due to the high topographic and vegetation heterogeneity of the territory. Liguria is one of the few regions in Italy affected by wildfires both in summer and winter. Most of the fires in Italy occur in summer season and the burned area is largely greater than in winter season. The seasonal fire regime in Liguria differs from this behaviour and is characterized by a number of wildfires and burned area higher in winter than in summer (Fig.1). Winter fire regime is mainly due to frequent extremely dry winds from the north in condition of curing for most of the herbaceous species.

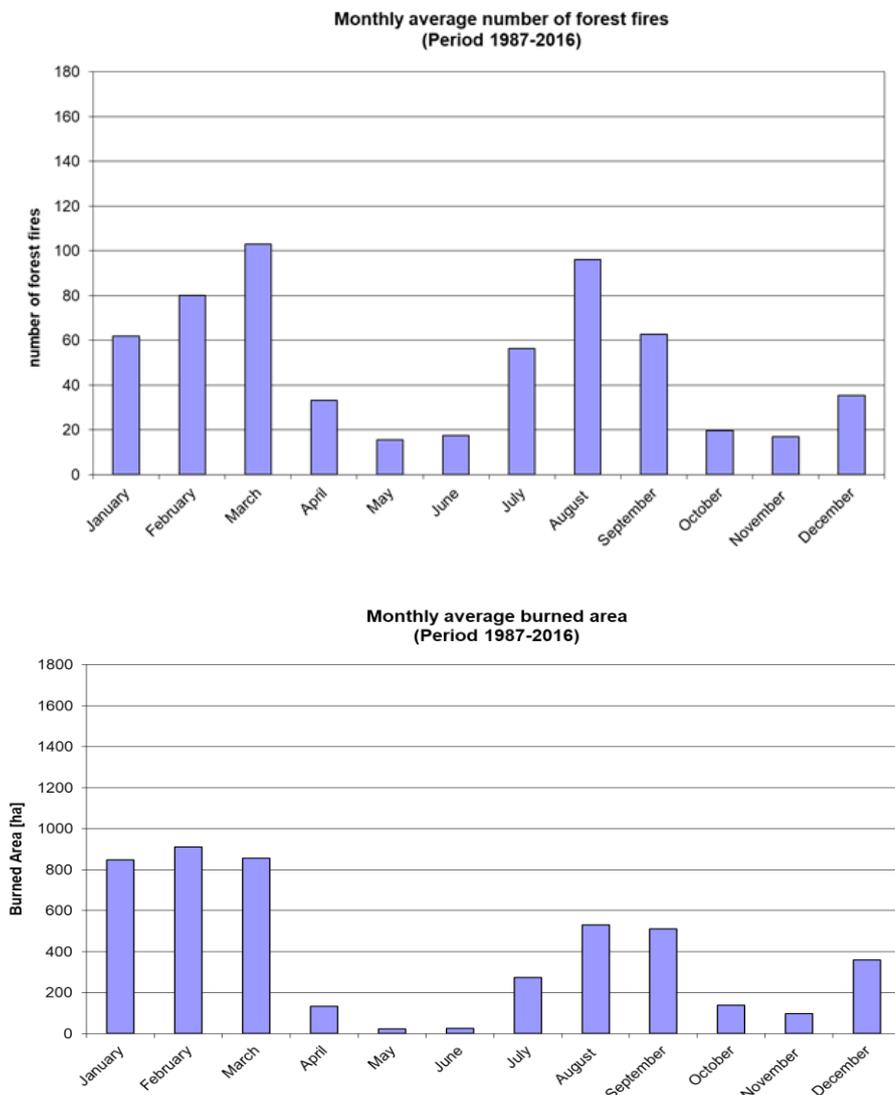


Figure 1 - Monthly frequency of forest fires and burned area in Liguria region (Italy) in the last two decades (1987 – 2016)

In view of the limited availability of fire risk management resources, most of which are used in the management of national and regional air services, it is necessary to precisely identify the areas most vulnerable to fire risk. The few resources available can thus be used on a yearly basis to mitigate problems in the areas at highest risk by defining a program of forest management interventions.

The main objective of the present study is to elaborate wildfire susceptibility mapping for Liguria region using two approaches, the first needing of a priori knowledge of the fires behaviour within the investigated area, and the second based on a machine learning stochastic algorithm. Namely, we compared results obtained by a statistical driven partitioning models with the ones obtaining by applying Random Forest.

2. Materials and methods

The availability of a mapping of fire perimeters spans almost 20 years (1996-2016), and this, combined with a detailed knowledge of topography, climate and land cover, allowed to understand the main features involved in wildfire occurrences and their behavior. The seasonality of the fire regime was also considered, partitioning the analysis in two macro season (November-April and May-October). Total precipitation and average air temperature obtained from the interpolation of 30 years-long time series from 164 raingauges and 127 thermometers series were considered. The analysis was carried out at a spatial resolution of 20 m on the Liguria region territory (5,400 km²) by using a dataset of fires occurrences that spans from 1996 to 2016.

The statistical driven partitioning analysis relied on a recursive-quantiles subdivision of the territory in classes based on the different available information layers: elevation, slope, aspect, type of vegetation (Table 1), rainfall height, temperature (these two latter subdivided in winter and summer periods). The algorithm was designed in order to assure the equal representation of each class, in which the number of fires occurred in the period of analysis is considered, in order to have an estimation of the fire hazard with a constant statistical confidence.

Table 1 - Type of vegetation cover considered in the models

Code	Type of Vegetation
10	Non-vegetated area
21	Other cultivated area
22	Grasslands
23	Olive groves
32	Pine forests
34	Other deciduous forest
35	Chestnut groves
37	Shrub vegetation
333	Less-burnable forest

The wildfire susceptibility map obtained with the described approach has been compared with the map obtained with a machine learning (ML) method. ML include a class of algorithm for the analysis, modelling and visualization of environmental data, and perform particularly well to model environmental and anthropogenic hazard, which naturally present a complex and non-linear behavior. ML explores the conception of algorithms capable of learning from and make predictions on data, through the modeling of the hidden relationships between a set of input and output variables (i.e. the predisposing factors and the occurrence of the phenomenon). After a training procedure, allowing to calibrate the parameters of the model, susceptibility maps can be displayed. In this study, we used Random Forest (RF) (Breiman, 2001), an ensemble ML algorithm based on decision trees (Fig.2). The

number of decision trees generated (n_{tree}), and the number of variables randomly sampled as candidates at each split (m_{try}) are the only hyperparameters that need to be specified. Then the algorithm generates n_{tree} subsets of the training dataset (counting about two-third of the observations) by bootstrapping (i.e. random sampling with replacement). For each subset, a decision tree is generated by iterating the following rules up to the maximum level, when each node contains less than a fixed number of data points: at each split, the algorithm select randomly m_{try} variables and computes the Gini index to select the best variable. The prediction of new data points is finally computed taking the average value of all decision tress for regression and the maximum voting for classification. The prediction error is assessed by evaluating prediction on those observations which were not used in the subset, that is the remaining one-third defined “out-of-bag” (OOB)

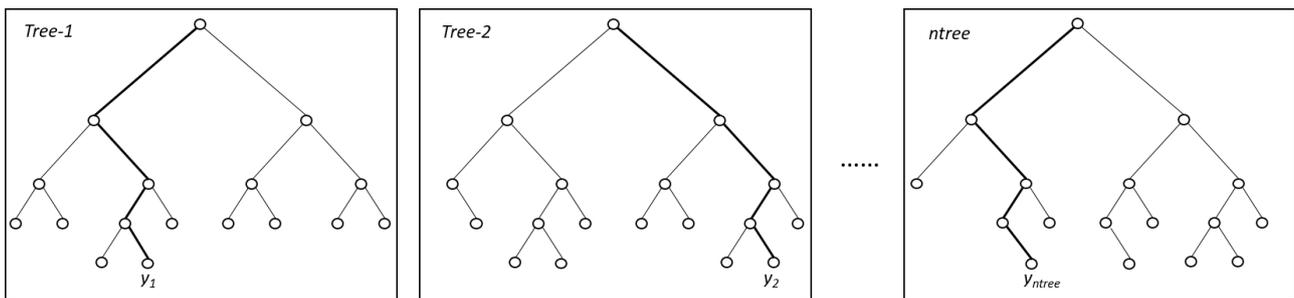


Figure 2 - Structure of Random Forest based on an ensemble of single decision trees

The main benefit of using ML algorithm in general is that these approaches are data driven, meaning that no a priori knowledge of the process is required. Another specific advantage of RF is that it allows to measure the relative importance of each variable on the prediction. This is obtained by looking at how much the tree nodes, which use that variable, reduce impurity across all the trees in the forest.

3. Results

Results of statistical driven partitioning method show a very high correlation with the topographic aspects, both in winter and summer. Rainfall is almost uncorrelated in both season. Shrubs results the vegetation most affected by fires both in summer and winter. In winter grassland burn quite often in the internal part characterized by high elevation and slope.

Random forest gives as output a prediction value, expressed as the probability for each pixel of burning under the assumption of a set af predisposing variables. This value is then projected on the training and testing datasets, allowing to produce a susceptibility maps and to validate the procedure. The testing and the OOB mean squared error are in our case both equal to about 0.2 for winter season and to 0.17 for summer summer. This let us to conclude that the RF setted model is robust and good for generalisation, since it performs well on testing data which were never used during the learing process.

As regards the variable importance measurement, it results that shrub vegetation is by far the most important variable in predicting burnig area (Fig.3), both in summer and in winter season. Grassland is the second most important type of vegetation. Nevertheless this one and all the other type of vegetation are less important than raifall, altitute and slope and temperature.

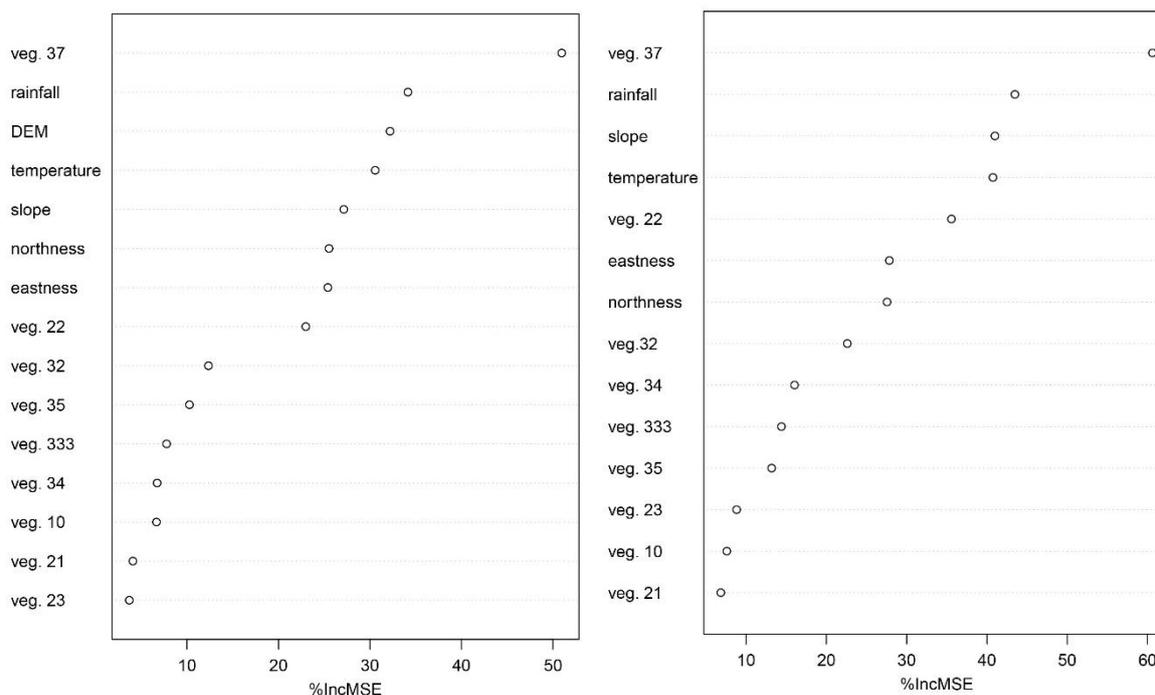


Figure 3 - Variable importance diagrams resulting from Random Forest, evaluated by computing the percentage increase of mean square error (%IncMSE). Left side: summer season; right side: winter season

Susceptibility maps for winter and summer were elaborated as resulting from the two approach (Fig.4). In Tab. 2 a numerical comparison of the statistical driven partitioning (PA) and machine learning approach, namely random forest (RF) is reported. Results are classified in 5 classes (CI) of susceptibility to wildfires, using the same thresholds for both approaches. The total area (TA) classified in each class and the burned area (BA) occurred in each class is reported considering the two approaches in summer and winter.

Table 2 - Comparison of statistical driven partitioning (PA) and random forest algorithm (RF)

	Summer				Winter			
	PA		RF		PA		RF	
	TA [%]	BA [%]	TA [%]	BA[%]	TA [%]	BA[%]	TA [%]	BA[%]
CI1	0,26	0,007	0,50	0,020	0,14	0,006	0,13	0,003
CI2	0,11	0,011	0,09	0,009	0,11	0,015	0,11	0,004
CI3	0,33	0,131	0,25	0,048	0,43	0,191	0,41	0,044
CI4	0,14	0,147	0,06	0,032	0,15	0,175	0,14	0,045
CI5	0,15	0,703	0,10	0,890	0,17	0,614	0,20	0,904

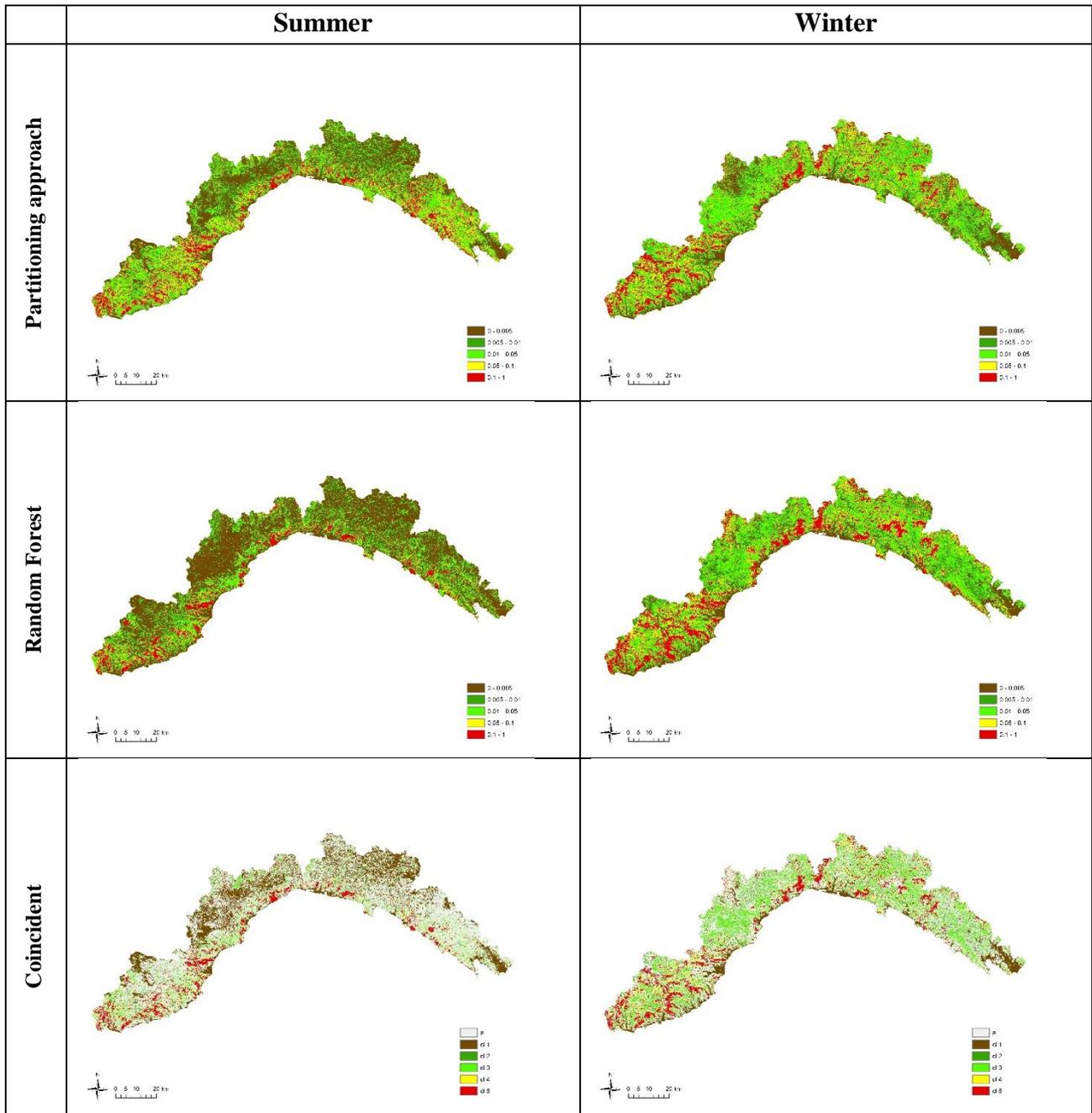


Figure 4 - Wildfires susceptibility maps for the two methods based on 5 generated classes from heavy high (red) to very low (brown). The map of coincidences is also shown on the bottom.

Susceptibility maps obtained by applying the two methods are broadly similar and show higher class matching with observed burned area, which are closer to the coast in summer and develop along the interior in winter. The RF algorithm seems to perform better, detecting about 90% of the BA within 10% of the entire territory, in summer, and 90% within 20% in winter VS 70% of BA within the 15% of the territory detected by the statistical driven partitioning in summer, and about 60% of BA within 17 % of the territory in winter. These results are preliminary and need to be further evaluated in detail by establishing a working plan accounting for a common validation dataset for the two approaches.

In summary, RF seems to be a promising alternative to deterministic or statistical expert- based method for wildfire susceptibility mapping.

4. References

- Breiman L, (2001). Random forests. *Machine Learning*, **45**, 5-32
- Eugenio F C, dos Santos A R, Fiedler N C, Ribeiro G A, da Silva A G, dos Santos Á B, Paneto G G, Schettino V R, (2016). Applying GIS to develop a model for forest fire risk: a case study in Espírito Santo, Brazil. *Journal of Environmental Management*, **173**, 65-71
- Fiorucci P., Gaetani F., Minciardi R., (2008). Regional partitioning for wildfire regime characterization. *Journal of Geophysical Research: Earth Sciences*, V. 113, Issue F2. <https://doi.org/10.1029/2007JF000771>.
- Ganteaume A, Camia A, Jappiot M, San-Miguel-Ayanz J, Long-Fournel M, Lampin C, (2013). A review of the main driving factors of forest fire ignition over Europe. *Environmental Management*, **51**, 651-662
- Leuenberger M, Parente J, Tonini M, Pereirac M G, Kanevski M, (2018) Wildfire susceptibility mapping: Deterministic vs. stochastic approaches. *Environmental Modelling & Software*, **101**, 194-203
- Parente J, Pereira M G, (2016). Structural fire risk: the case of Portugal. *Science of the Total Environment*, **573**, 883-893
- Pourghasemi H R, (2016). GIS-based forest fire susceptibility mapping in Iran: a comparison between evidential belief function and binary logistic regression models. *Scandinavian Journal of Forest Research*, **31**, 80-98
- Pourtaghi Z S, Pourghasemi H R, Aretano R, Semeraro T, (2016). Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecological Indicators*, **64**, 72-84
- Rodrigues M, de la Riva J, Fotheringham S, (2014). Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Applied Geography*, **48**, 52-63