A multivariate approach to assess the structural determinants of large wildfires: evidence from a Mediterranean country

David Sousa[A], *Frederico Cruz-Jesus* (iD) [A,C], *André Sousa*[B] *and Marco Painho*[A]

[A]NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal.
[B]ISEG Lisbon School of Economics and Management, Universidade de Lisboa, Rua do Quelhas 6, 1200-781 Lisboa, Portugal.
[C]Corresponding author. Email: fjesus@novaims.unl.pt

Abstract. This paper analyses the factors behind wildfire propagation in a Mediterranean European country, Portugal, using a set of variables related to vegetation and climatic, topography and human aspects. Spatial cluster analysis was used to find homogeneous regions, and two-part regression models were used to model the contribution of the different elements driving extensive fire propagation. Our findings confirm the presence of spatial variability in the contribution exerted by most structural factors driving large wildfire spread. Additionally, the results of this study show that vegetation types, in particular the presence of shrubs, and a lack of human activities, such as agriculture, represent the main factors facilitating fire spread in this region, corroborating information from previous work. This research provides relevant input for implementation in different fields, from large fire awareness and prevention to the development of wildfire policies, as well as addressing methodological concerns in fire danger and fire risk analyses.

Introduction

Fire is perceived as an important agent for both ecological development and deterioration of forest ecosystems around the world (Verde and Zêzere 2010; Ferreira-Leite *et al.* 2013*b*). In fact, even though fire may have always been present as a landscape transformation and renewal factor (Goudie 2006), fire events can have disastrous human, environmental and economic consequences, especially when they develop into large wildfires (Tedim *et al.* 2013).

Mediterranean Europe is particularly vulnerable to these destructive phenomena. This region is home to the second-most diverse community of species worldwide after the tropics, and even though Mediterranean ecosystems are historically fire resilient, intense human pressure on the environment has favoured fire recurrence and increased fire size in many areas of the Mediterranean basin in the second half of the 20th century (Moreno *et al.* 2013). Large fires play a significant role in this context because they are responsible for the majority of the total burned area while representing only a small share of all fire events (San-Miguel-Ayanz *et al.* 2013). Trends of increased fire incidence and intensity have been observable across European Mediterranean countries, although at different rates. In Portugal, there has been evidence of higher fire incidence than in other Mediterranean countries (Rego and Silva 2014), a steady increase in the frequency of large wildfires and an increase in

the extent of burned area during the second half of the 20th century (Ferreira-Leite *et al.* 2013*a*). Climatic characteristics, land abandonment and other socioeconomic transformations have left the Portuguese inland territory susceptible to the occurrence of large wildfires (Oliveira *et al.* 2012), and climate change is expected to significantly increase fire danger (Rego and Silva 2014).

According to Álvarez-Díaz *et al.* (2015), the four decisive conditions for wildfire occurrence and spread are favourable meteorological conditions, the presence of fuel, its spatial continuity and a source of ignition. All significant driving factors of wildfires are discussed from these perspectives in the present study. This work follows an adapted categorisation from Mhawej *et al.* (2015) and Ganteaume *et al.* (2013), in which the driving factors of wildfires are divided into (i) vegetation; (ii) climatic; (iii) topographic; and (iv) human.

Vegetation conditions represent the decisive component of any fire. Fuel is the main requirement for fire ignition and spread (Cao *et al.* 2013; Holsinger *et al.* 2016), and different vegetation patterns promote the fire susceptibility of landscapes. Even though many fires are human-caused, specifically in southern Europe, the main features of local vegetation remain a determinant factor driving wildfire risk, as they control the success of the ignition event and – most importantly – fire behaviour (Calviño-Cancela *et al.* 2016).

Evidence from several studies, particularly in Mediterranean regions, has established the importance of climatic factors in the analysis of wildfire patterns. Climate is well noted for its impact in shaping fire regimes in those areas (Ganteaume *et al.* 2013). Moreover, the impact weather exerts during fire events is deemed very relevant (Hernandez *et al.* 2015). Among these features, we count temperature, precipitation and wind. It is also important to note that climatic aspects are subject to significant spatial and temporal variation (Keeley and Syphard 2016).

Topography has been found to be associated with wildfire risk by many authors, with Ganteaume *et al.* (2013) considering it one of the most elementary environmental factors driving wildfire occurrence in Mediterranean Europe. Although the significance and direction of this association have varied among studies, several authors have mentioned the effect of different topographic features on burned area and ignition density, as shown by Nunes *et al.* (2016). Among these features, we include slope, aspect and elevation.

Finally, human factors are widely recognised among the most important drivers of wildfires. These human factors include population density and dynamics, socioeconomic characterisics, changes in land cover, infrastructure and human activities, such as agriculture and keeping livestock (Balsa Barreiro and Hermosilla 2013; Nunes *et al.* 2013). It should be noted that in contrast to non-human factors, human factors are predominantly non-stationary in time and space (Rodrigues *et al.* 2016; Keeley and Syphard 2018).

The relative influence of the factors driving large wildfire spread in Portugal needs to be further studied for a more in-depth understanding of fire dynamics. As Tedim *et al.* (2013), Calviño-Cancela *et al.* (2017) and San-Miguel-Ayanz *et al.* (2013) note, improved knowledge regarding the various interactions among the main factors driving wildfire risk and fire behaviour and their arrangement and patterns should translate into an improvement in risk management strategies, such as more efficient prevention measures, the development of existing regulations and the creation of campaigns aimed at increasing the awareness of different stakeholders in connection with specific activities and settings. Additionally, a better understanding of large wildfire occurrence is essential for prevention efforts and firefighting planning, as well as for legislative frameworks (Moreira *et al.* 2010).

The results of empirical studies focusing on the drivers of wildland fire propagation are still considered valuable input for model parameterisation from the perspective of risk analysis (Miller and Ager 2013), and there is a high degree of uncertainty involved in the study of wildfires (Rodrigues *et al.* 2016). Furthermore, research studying large wildfires in Portugal has not fully investigated their underlying causes, specifically concerning human-related elements (see e.g. Costa *et al.* 2011; Fernandes *et al.* 2016; Turco *et al.* 2019), even though large fires are responsible for most wildfire damage (Ganteaume and Jappiot 2013).

The present research aims to explore the main determinants of fire spread in central Portugal between 2005 and 2015, specifically evaluating their spatial patterns and dynamics as well as their individual contributions to fire propagation. The focus of this study is on medium to large events in terms of burned area, and all occurrences with a burned area larger than 100 ha were selected for this purpose (Ganteaume and Jappiot 2013). This research should result in beneficial developments in awareness, prevention and wildfire policies and advances in the effective mitigation of large wildfires (Grala *et al.* 2017), specifically in the Portuguese context.

Materials and methods

Study area

The study area of this research is the central region of Portugal (see Fig. 1), which, in 2011, comprised ~1.6 million inhabitants. This area is characterised by a Mediterranean climate (Köppen-Geiger classification Csa and Csb) (Kottek *et al.* 2006), much like the rest of the Portuguese territory, although there is a noticeable contrast in annual precipitation totals between the wetter northern coastal areas and the drier southern inland areas. According to the national land cover and land use cartography, in 2010, ~65% of the study region was forest area. The main species comprising these forests were pine trees (38.5%), followed by eucalyptus trees (24.5%). Whereas pine tree forests were relatively abundant throughout the study area, particularly in the centre, the north-eastern part of the study region had virtually no eucalyptus. However, shrublands were very common in the north-eastern area (14% overall), and forests with other species (predominantly oak trees and sclerophyllous vegetation) were extensive and found mainly along the border (23% overall). Protected areas were also frequent over the inland extent of this region (in total, 12 protected areas intersect the territory under analysis). Rangelands, which comprised ~20% of the entire area, were mostly restricted to the south-eastern portion of the study region.

As in the rest of the country, the ignition density and burned area are spatially dissociated in the study area (Moreira *et al.* 2010), with ignitions occurring primarily in densely populated areas along the coastline and larger burned areas being mostly restricted to inland areas, coinciding with higher elevations and shrub vegetation (Mateus and Fernandes 2014). The fire regime in the study area is characterised by short fire-return intervals (Oliveira *et al.* 2012). There is a marked seasonal pattern in fire activity, which is concentrated in the summer months (90% of the entire area burned between 1996 and 2012 occurred from June to September), and most fires are stand-replacement crown fires (Mateus and Fernandes 2014).

Data collection and processing

The spatial framework of this research is the European ETRS89-LAEA 1×1 km Inspire grid (European Forum for Geography and Statistics and Eurostat 2019), and we used evidence from previous analyses in this field of study (e.g. see Rodrigues 2015; Vilar *et al.* 2010). The study area corresponds to 21 570 1-km^2 grid cells. Variables have been transposed to the grid framework following different methodologies (see Figs S1–S6 in the Supplementary material). Data processing and integration were performed with the help of *ArcMap*® 10.5.1 geographic information system (GIS) software, *Microsoft Excel*® and *Microsoft Access*®. The target variable, the percentage of burned area within each cell, was calculated from the original data by the National Nature Conservation and Forestry Institute (Instituto da Conservação da Natureza e das Florestas, I.P.; ICNF, IP). The
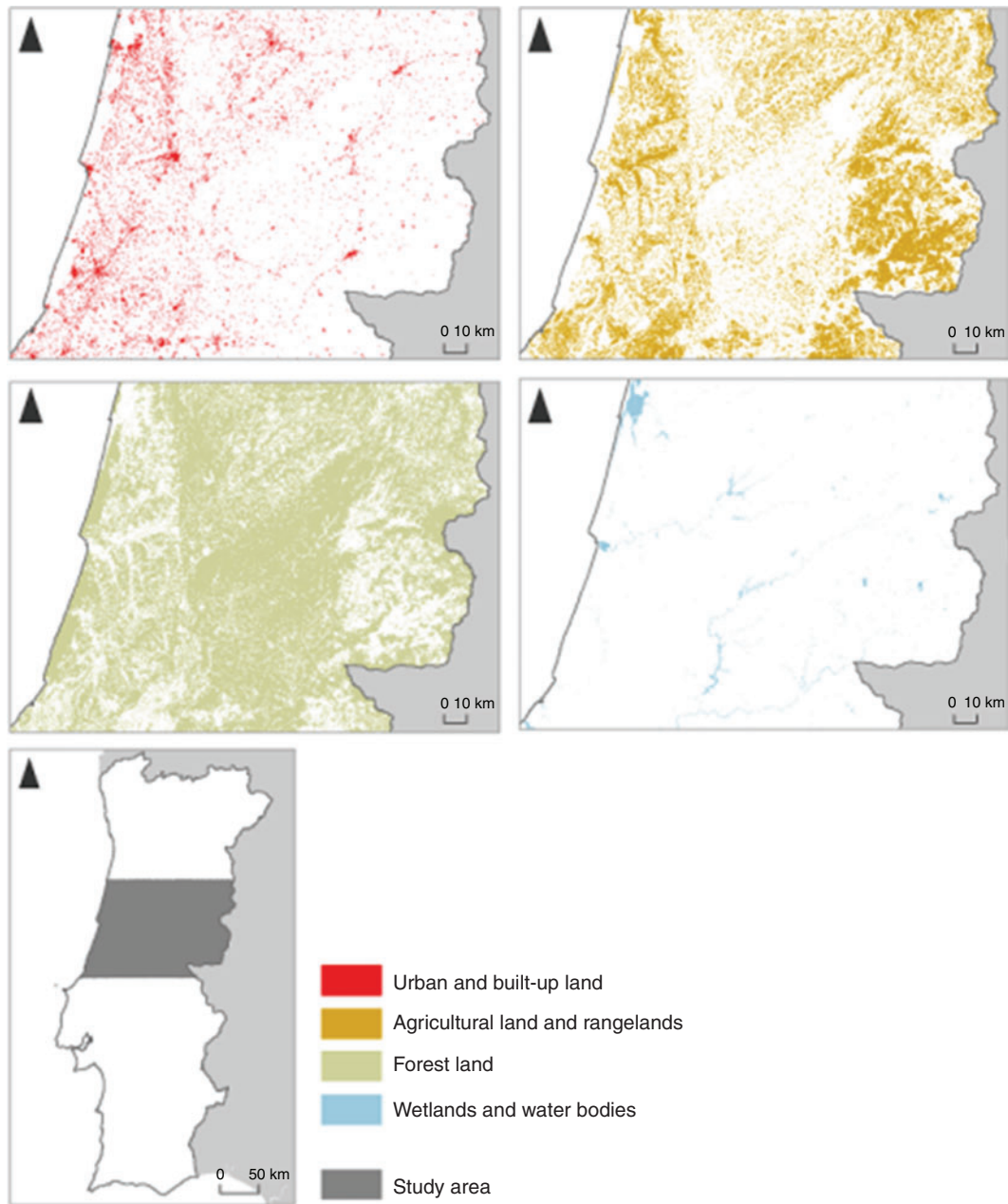
Fig. 1. Study area. This strip of land covers most of the area corresponding to mainland central Portugal and is composed of 65% forest areas.

burned area corresponded to 383 different fire perimeters, equivalent to 5343 grid cells or ~25% of the study area (see Fig. 2). Initially, data pertaining to 37 variables were collected based on the results of previous works in this field as well as data availability (see Table S1 in the Supplementary material).

Data analysis

Preliminary steps were conducted to successfully apply quantitative analysis (see Fig. 3). The first step consisted of an initial examination of the data to ensure its reliability and suitability for cluster analysis and two-part models. As the data included a large number of variables, data reduction was employed. To this end, we used Pearson's correlation coefficient to measure the (linear) association between every pair of metric variables. For each pair with an absolute value greater than 0.5, we kept only the one with the highest value in predicting fire presence, making use of a measure of entropy similar to a decision tree of depth one. After this procedure, we standardised the data (Sharma 1996; Cleff and Cleff 2014), and two standardised geographical variables denoting cell x and y coordinates were added to the analysis, as is usually advisable in geographic data analysis.
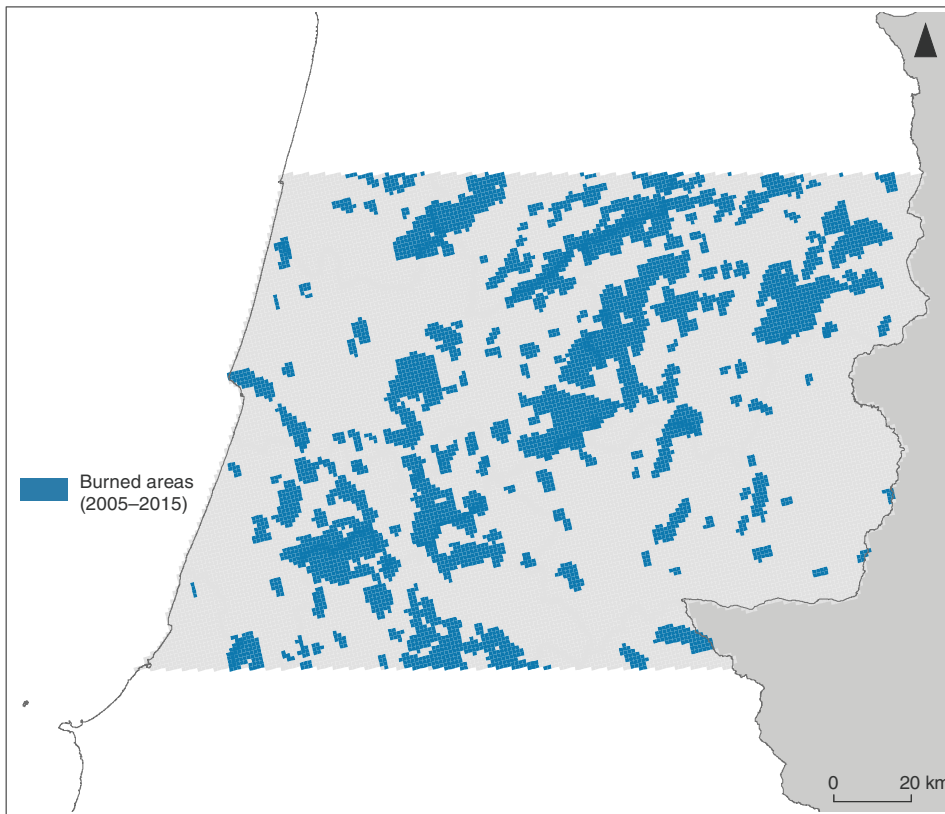
Fig. 2. Burned area cells (2005–15). Large fires are concentrated in the north-eastern and south-western portions of the study area.

The second step, after the initial examination of the data and their dimensional reduction, consisted of using cluster analysis to form homogeneous groups of cells with respect to the independent variables kept in the first step. The goal of cluster analysis is to form groups of cells that are as homogeneous as possible within clusters and as heterogeneous as possible between clusters. Following Sharma's (1996) recommendation, a hierarchical clustering analysis with five different algorithms (single, complete, average, centroid and Ward's) was performed. Based on an analysis of the R^2 and dendrograms of the five methods, Ward's algorithm was chosen to define the number of clusters and the initial seeds to the non-hierarchical algorithm, k-means. This approach has been used before in the literature with evidence of providing better results than using only one type of method (Sharma 1996; Cruz-Jesus *et al.* 2012). The cluster analysis was performed with SAS *Enterprise Guide*® statistical software. According to Kaiser's criterion, the optimal number of clusters was five. Ward's method was used to generate the initial seeds, which were subsequently optimised by the k-means algorithm.

The third and last step was the application of the two-part models, which enabled the ultimate goal of this study: to understand the structural determinants of large wildfires. A short review of wildfire danger assessment methods confirmed the suitability of generalised regression techniques, such as logistic regression, in this field of study. These regression methods were chosen to explore the contribution of the main

factors of large wildfire spread, as well as to recognise the variability of each factor's importance among regions. The percentage of burned area in each cell follows a highly skewed distribution with zero inflation, which prevents the use of the classic multivariate ordinary least-squares (OLS) linear regression model. The applicability of non-linear two-stage estimation procedures, such as two-part models (2PM), has been successfully demonstrated in a variety of fields where observed data (either count or continuous data) are characterised by a heavy presence of zeros in the response variable (Farewell *et al.* 2017), including in the context of wildland fires. For example, Parisien *et al.* (2014) used zero-inflated negative binomial models to describe burned area count data. In this study, we implement a 2PM. This type of model is composed of two distinct stages, the first predicting the probability of occurrence (0,1) through a binary response model, and the second predicting the target variable conditional on non-zero outcomes using the linear regression model (Buntin and Zaslavsky 2004). For the first stage, the probit model was chosen, and this approach uses a specific case of binary response models following the standard normal cumulative distribution (Wooldridge 2012).

Specifically, the first part of the model, the probit model, can be presented as follows:

$$
\begin{aligned}
P(y = 1|X) &= P(y = 1|x_1, x_2, \ldots, x_k) \\
&= G(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik})
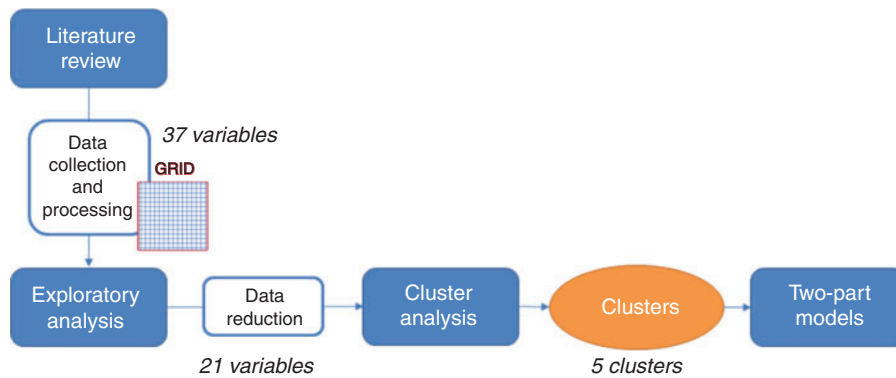\end{aligned}
\tag{1}
$$

Fig. 3. Methodological workflow.

where y is the binary variable, which equals 1 when a wildfire occurs in grid cell i and 0 otherwise, β_0 is a constant, X is the full set of explanatory variables, $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients, and G is a function taking on values strictly between zero and one: $0 < G(z) < 1$, for all real numbers z. The explanatory variables for the global model and for each of the five clusters are listed in Supplementary Table S2.

The second part of the model, the linear regression model, takes the following form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i, (i = 1, 2, \ldots, n) \quad (2)$$

where y is the dependent variable (percentage of burned area), β_0 is a constant, $x_{i1}, x_{i2}, \ldots, x_{ik}$ are the explanatory variables of grid cell i, $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients, and u_i is the error term of grid cell i. The explanatory variables for the global model and for each of the five clusters are detailed in the Results section.

For the purpose of this study, the percentage burned area was first generalised to a binary target variable that could be understood as fire presence. To evaluate the results from the probit models, the average partial effects (APE), which represent the average of the non-linear function measuring the effect of x_k on $P(y = 1|X)$, were estimated, and the probability of burning associated with the entire range of population values was plotted for each variable. Subsequently, an OLS regression was fitted to the subpopulation of cells displaying burning activity during the decade. All models were fitted with the forward-stepwise selection procedure and using *Stata®*. Fig. 3 presents the condensed methodological workflow of this study.

Results

After the previously described process of data reduction, 21 variables were kept for further analysis (see Table 1).

The hierarchical cluster analysis resulted in a five-cluster solution as the most balanced partition, corresponding to an R^2 value of 0.298. The k-means method performed after needed a total of five iterations for completion and improved the value of R^2 to 0.324. The spatial representation of the clustering solution can be found in Fig. 4.

Table 1. List of variables kept for statistical analysis, considering the measures of variable correlation and variable worth in predicting fire presence

Variable	Acronym
Aging index	AGE_INDEX
Agricultural area in cell (%)	AGR_COS
Area of eucalyptus forests in cell (%)	EUC_COS
Area of other types of forests in cell (%)	OTHER_COS
Area of shrubland in cell (%)	SHRUB_COS
Aspect (°)	ASPECT
Average number of livestock per farm (no.)	LVSTK_NFARM
Average utilised agricultural area per farm (ha)	SAUFARM_HA
Distance to ignition locations (km)	IGN_DIST
Distance to primary roads (km)	PROAD_DIST
Distance to protected sites (km)	AP2015_DIST
Elevation (m)	ELEVATION
Farm density (no./km^2)	FARMDEN_KM
Rangelands in cell (%)	GRZ_COS
Livestock units per utilised agricultural area (no./ha)	LUNITS_NSAU
Number of dry months (Gaussen index)	DRYMONTH
Population density (no./km^2)	POP_GRID
Population employed in agriculture, livestock, fishing, forestry and hunting (%)	PRIM_PERC
Potentiality index	POTENT_INDEX
Seasonal, secondary use and empty housing (%)	SSEHOUS_PERC
Slope (°)	SLOPE

Table 2 summarises the clustering solution regarding group dimension, cluster location and specific fire spread-related characteristics. A great contrast between groups is visible overall, although the Southern Urban Coastline and Northern Urban Coastline seem very close with respect to their main characteristics, exhibiting a distinct urban profile, with differences limited to the number of dry months (extended drought season in the south) and livestock activity and animal density (more intense in the south). The North-eastern Highlands and Central Forests also share several resemblances, including an aging population and a high proportion of empty or secondary-use housing, although fire incidence patterns differ significantly. The North-eastern Highlands region is mainly defined by biophysical aspects, including high elevations and large
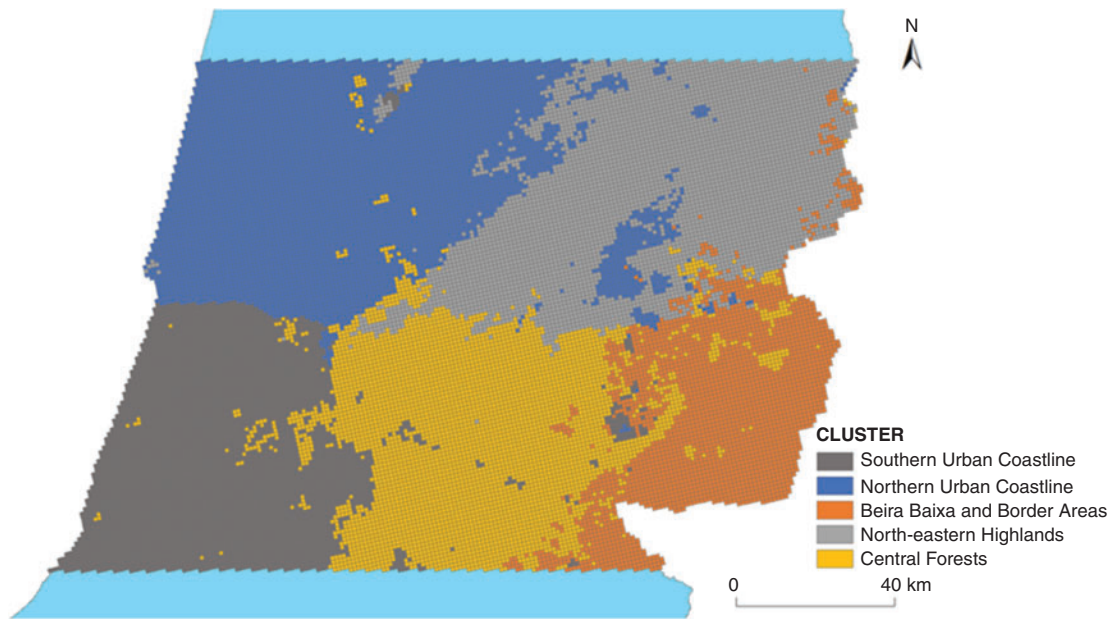
Fig. 4. Spatial representation of the clustering solution (five clusters) based on the 21 variables selected to explain wildfire propagation.

Table 2. Description of the clustering solution regarding the number of cells and the 21 variables selected to explain wildfire propagation

Features	Southern Urban Coastline	Northern Urban Coastline	Beira Baixa and Border Areas	North-eastern Highlands	Central Forests
No. cells	3899	5546	2726	4853	4546
Population	Average to high population density, young population with a medium potentiality index	Medium to high population density, young population with an average to high potentiality index	Average to low population density, very old population with a medium potentiality index	Medium to low population density, relatively old population with a very low potentiality index	Average to low population density, moderately old population with medium to high potentiality index
Roads and housing	Average to high density of primary roads and high ongoing housing occupation	Medium to high road density and high ongoing housing occupation	Very low density of primary roads and high amount of not permanently occupied housing	Average road density and high amount of empty housing	Average density of primary road and considerable to high amount of not permanently occupied housing
Agriculture and livestock	Very few people employed in primary sector activities, agricultural activity is still present, and rangelands are less dominant, average to high density of small to medium farms, high livestock activity and very high animal density per utilised agricultural area	Average to low proportion of population employed in the primary sector, average extent of agricultural areas but smaller rangelands, very high density of small to medium farms, moderate livestock activity and animal density	High number of people employed in primary sector activities, large agricultural areas, especially for grazing, low density of high dimension farms, average to low livestock activity and low animal density	Considerable number of people employed in primary sector activities and medium to low extents of agriculture and rangelands, moderate to low density of regular-sized farms, medium to low livestock activity and low animal density	Average number of people employed in primary sector activities and very small extents of agriculture and rangelands, medium to low density of moderate to small farms, average to low livestock activity and animal density
Biophysical aspects	Flat and low-altitude landscape, moderate presence of eucalyptus and forests of other types, average to small shrub extents, extended drought season	Terrain is relatively flat and at low elevation, medium to high presence of eucalyptus trees, average to low extents of shrublands and forests of other species, short drought season	Flat landscape at medium–low altitude, very few eucalyptus trees and shrubs, large forests of other types, extended drought season	Rugged surface at high elevation, very few eucalyptus trees, average to high presence of other types of forest and very large shrublands, short drought season	Rugged surface at medium elevation, high eucalyptus presence, medium to low extents of other species, moderate shrub presence, extended drought season
Ignition density	Medium-high	Medium-high	Very low	High	Medium

Table 3. Quality assessment measures for the fire presence models (probit models): global model and for each of the five clusters

Notes: n equals 21 750 in the global model (number of grid cells) and the dimension of the clusters for each of the other five models. Significance of the Pearson χ^2 statistics: ***, $P < 0.01$; **, $P < 0.05$; *, $P < 0.1$

Model assessment	Global	Southern Urban Coastline	Northern Urban Coastline	Beira Baixa and Border Areas	North-eastern Highlands	Central Forests
Pearson χ^2	272 234***	4525***	4862	2594	18 346***	3733
Sensitivity	42.58%	47.76%	39.33%	2.08%	64.14%	50.86%
Correct classifications	80.06%	82.92%	83.01%	92.81%	72.88%	81.74%
Area under the ROC curve	84.44%	85.39%	86.33%	84.07%	81.51%	86.40%
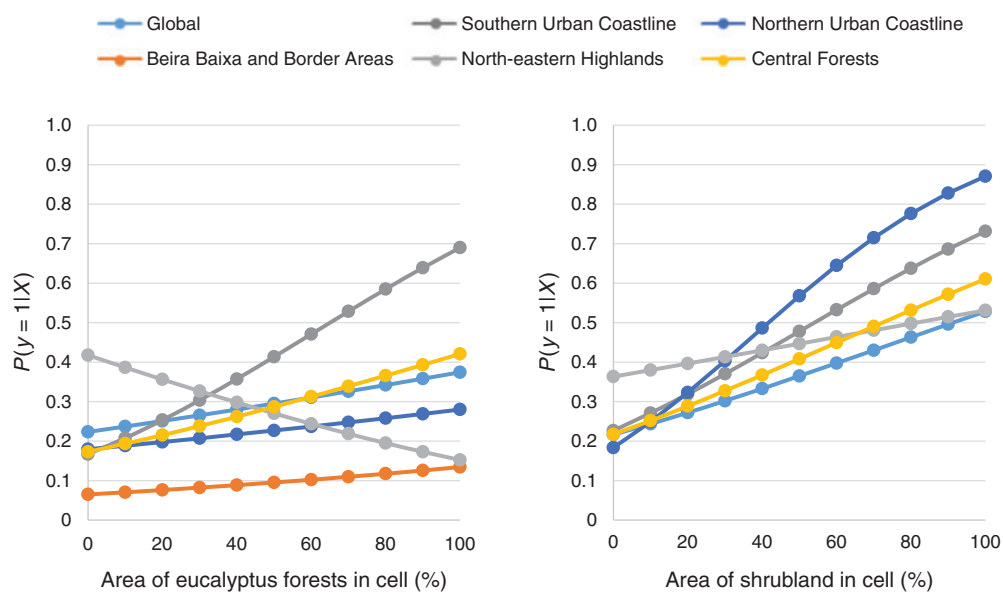
Fig. 5. Burning probability associated with the fire presence models (probit models) plotted for the entire range of population values: selected vegetation factors.

shrub extents. The Central Forests region has very few agricultural areas and rangelands and is instead dominated by eucalyptus forests.

Probit models, which in this case can be understood as modelling fire presence, were then developed for the entire study area and each region according to the clustering solution. The results of this analysis are presented, including the estimated models' quality assessment measures, in Table 3, as well as plots of burning probability for the entire range of population values (Figs 5–7). The APE and the statistical significance of the estimates can be found in Table S2.

From the goodness-of-fit measures (Table 3), it is clear the fire presence models performed adequately, especially when considering the percentage of correct classifications and the area under the receiver operating characteristic (ROC) curve. Nevertheless, three models displayed non-significant Pearson χ^2 values, and the percentage of accurate positive classifications for Beira Baixa and Border Areas (both the smallest and the least affected by fire) was very small.

The model specification left out the number of farms per kilometre squared (FARMDEN_KM) and the percentage of rangelands (GRZ_COS), suggesting that these two factors were not determinants of the propagation of large wildfires across the study area. Three other variables were deemed significant both in the global model and in all five regions: the percentage of the resident population employed in primary sector activities (PRIM_PERC); the percentage of land covered by eucalyptus forests (EUC_COS); and the distance to ignition locations (IGN_DIST).

The results of the models estimating the percentage of burned area (area-burned models) provide further insights about the impact of each driving factor on fire propagation in the case of large wildfire events, complementing the findings from the binary response models. Table 4 presents the quality assessment measures of the six models, as well as the number of cells in each cluster ($y = 1$). The models' coefficients and associated statistical significance are summarised in Table 5. The residuals from all models were found to be approximately normally distributed, which confirmed the suitability of this technique.
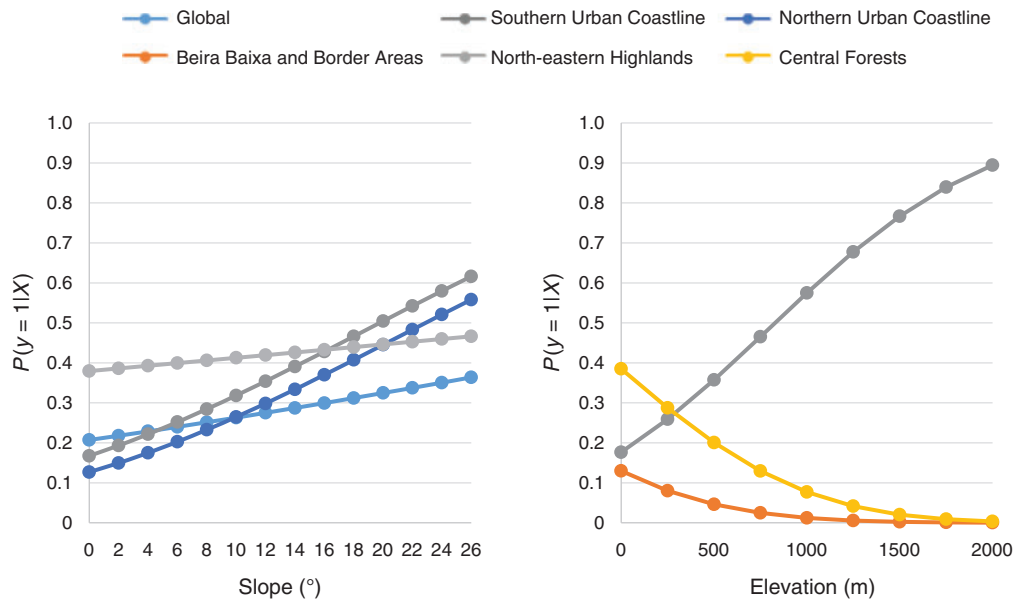
Fig. 6.   Burning probability associated with the fire presence models (probit models) plotted for the entire range of population values: selected topographic factors.

It can be seen that the North-eastern Highlands region is by far the region with the highest proportion of burned area cells (approximately 41%), while in the Beira Baixa and Border Areas, only 7% of the territory experienced burning by large wildfire events during the 2005–15 period.

Specific driving factors have been highlighted for discussion because of their impact on large fire spread, predictive ability or inconsistency across clusters. Fig. 5 summarises the main results of the six fire presence models concerning the considered vegetation factors, displaying the associated burning probability for specific values within the variables' ranges.

Fuel availability represents one of the most relevant contributing factors to large wildfire spread, as gathered from the models' results. The extent of eucalyptus tree cover (EUC_COS) displayed noticeable positive associations with burning probability in all regions, except for the North-eastern Highlands (Fig. 5). The Southern Urban Coastline shows a very pronounced rise in burning probability in connection with eucalyptus forests, though the relationship is also prominent in the North-eastern Highlands. The results from the area-burned models follow the same behaviour (i.e. a negative association in the North-eastern Highlands), with the Southern Urban Coastline and Central Forests presenting higher coefficient values.

Shrublands (SHRUB_COS), on the other hand, were found to increase burned area probability in all regions. This pattern was observable in the fire presence and area-burned models, with only North-eastern Highlands having a weaker association. The 90% burning probability corresponding to 100% shrub cover in the Northern Urban Coastline is of note (Fig. 5). Among the OLS coefficient estimates, this same region stands out with a higher coefficient (Table 5), meaning that the percentage of shrub cover increases the percentage burned area in a cell that does burn in roughly half its proportion (0.546).

Fig. 6 summarises the main results of the six fire presence models concerning the topographic factors considered.

The terrain slope (SLOPE) seems to display a positive relationship with burning likelihood, which is particularly noticeable for the Southern Urban Coastline and Northern Urban Coastline (Fig. 6). The probabilities associated with the different slope values range ~40 percentage points in these regions, with the maximum gradient corresponding to more than 60% burning probability in the Southern Urban Coastline. The estimated OLS coefficients also exhibit positive associations for this variable, with the highest coefficient in the Beira Baixa and Border Areas, although the value was barely significant (Table 5).
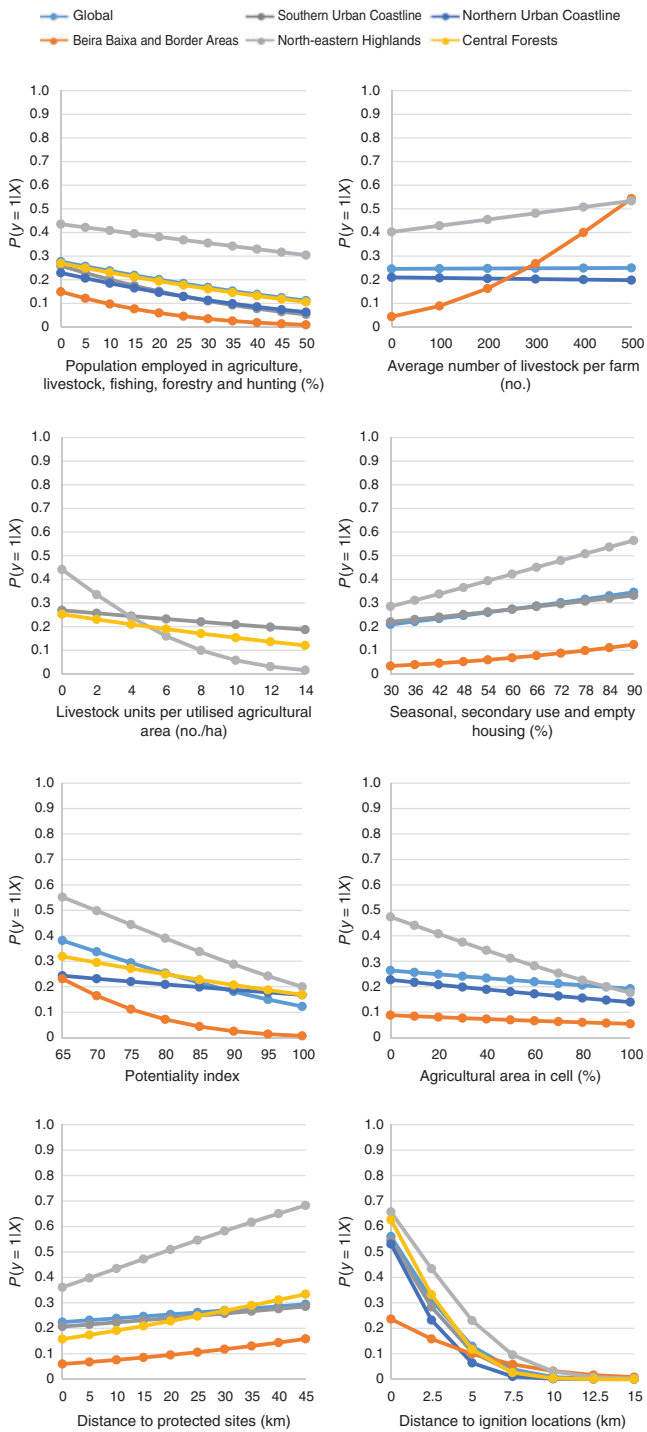
The variable plot for elevation (ELEVATION) shows a very steep rise in burning likelihood for the Southern Urban Coastline with an increase in altitude (Fig. 6), while the other regions considered exhibit the opposite trend. The results of the area-burned models show similarly mixed results, if not exactly for the same clusters (Table 5). The Southern Urban Coastline presents a moderate positive association with burned area (an increase of ~2.3 percentage points for each 100 m rise), whereas for the North-eastern Highlands and Central Forests, the relationship is negative.

Fig. 7 summarises the main results of the six fire presence models concerning the human factors considered.

The results from the fire presence models suggest that population working in the primary sector (PRIM_PERC) impacts fire spread as a deterrent, regardless of geographical location, with all clusters displaying a negative relationship between this variable and the likelihood of burned area. The maximum burning probability (higher than 40%) is listed at 0% primary sector working force for the North-eastern Highlands (Fig. 7).

The results from the area-burned models also support this evidence, with the Southern Urban Coastline, Beira Baixa and Border Areas and Central Forests having a negative relationship between the PRIM_PERC and the percentage of burned area per cell (Table 5). This relationship can be considered stronger in the

Fig. 7. Burning probability associated with the fire presence models (probit models) plotted for the entire range of population values: selected human factors.

be considered more pronounced in the case of the North-eastern Highlands (Fig. 7). This region exhibits the highest burning probability (~50%) at 0% agricultural land cover, while the negative behaviour of all other regions considered is weak.

When considering the percentage of burned area per cell, the effect of agricultural land cover is very similar, with all models displaying negative associations between both variables (Table 5). The weight of this variable is again most relevant in the case of the North-eastern Highlands, where agricultural areas represent a decrease of ~65 percentage points in cell burned area.

The effect of livestock production on burning probability is mixed. A strong positive association is observable for the average number of animals per farm (LVSTK_NFARM) in the Beira Baixa and Border Areas and North-eastern Highlands, with 500 animals corresponding to a burning probability of over 50% in both regions (Fig. 7). Animal density (LUNITS_NSAU), on the other hand, exhibits a negative relationship with burning probability for the Southern Urban Coastline, Central Forests and, particularly, North-eastern Highlands. In the latter case, burning probability declines from ~45% at no livestock units per hectare to virtually 0% when animal density is highest. The outcomes of the area-burned models differ, with the Northern Urban Coastline and North-eastern Highlands displaying a negative association, and the global model and Southern Urban Coastline presenting the opposite trend.

All models presented a clear positive association between the distance to protected sites (AP2015_DIST) and the probability of burning, meaning that areas further from these locations are expected to burn in wildfire events larger than 100 ha. This statement is particularly true for the North-eastern Highlands, where the range of associated probabilities is larger (Fig. 7). The majority of the estimated OLS coefficients support these results, with the North-eastern Highlands displaying a very high positive coefficient (Table 5). The Northern Urban Coastline, on the other hand, exhibits a negative connection to percentage burned area, which suggests that the impact of this variable fluctuates across space.

Demographic growth potential (POTENT_INDEX) seems to be negatively associated with burning probability throughout the study area, with the Beira Baixa and Border Areas and North-eastern Highlands displaying the most pronounced drops in burning probability with an increase in the index (Fig. 7). These same results are also observable in the area-burned models, where estimated coefficients support these negative relationships and where the North-eastern Highlands region stands out with a decrease in burned area percentage in approximately the same proportion as the drop in the index (Table 5). The population density (POP_GRID) is also associated with burned area in the same direction, with the North-eastern Highlands showing a decrease of ~4 percentage points in the percentage burned area with an increase of 100 inhabitants.

The distance to ignition points shows significant negative associations with a large wildfire propagation likelihood everywhere in the study area (Fig. 7). This effect on probability follows the same behaviour in all regions, with a sharp decline in burning probability with an increase in distance to ignition locations. This variable is also strongly associated with the extent of burned area, given the high coefficients displayed by

case of the Southern Urban Coastline, where an increase of 1% of people employed in primary sector activities means a decrease in burned area of ~1.29 percentage points.

The influence of agricultural areas (AGR_COS) on burning probability is similar to that displayed by PRIM_PERC and can

Table 4.   Dimension and quality assessment measures for the area-burned models (OLS models): global model and for each of the five clusters
Significance of the F statistics: ***, $P < 0.01$; **, $P < 0.05$; *, $P < 0.1$

Model assessment	Global	Southern Urban Coastline	Northern Urban Coastline	Beira Baixa and Border Areas	North-eastern Highlands	Central Forests
Number of cells (n)	5343	917	1139	192	1994	1101
F statistic	89.34***	36.31***	44.10***	6.06***	28.50***	31.47***
Adjusted R^2	0.188	0.316	0.294	0.096	0.181	0.200

Table 5.   Coefficients of the area-burned models (OLS models) and associated statistical significance: global model and for each of the five clusters
Significance of the coefficients: ***, $P < 0.01$; **, $P < 0.05$; *, $P < 0.1$. Blank cells indicate the variables were excluded by the forward-stepwise selection procedure

Variables	Global	Southern Urban Coastline	Northern Urban Coastline	Beira Baixa and Border Areas	North-eastern Highlands	Central Forests
CONSTANT ($\hat{\beta}_0$)	103***	29***	71***	40***	153***	106***
PRIM_PERC		−1.29375***	0.71687***	−0.60788**		−0.47366**
FARMDEN_KM		−1.03295**				−1.59624**
SAUFARM_HA					−0.23557**	
LVSTK_NFARM	0.00107***	0.00069**				
LUNITS_NSAU	0.39146*	0.76439***	−1.69891**		−7.00184***	
SSEHOUS_PERC	0.13319***	0.56645***				
POTENT_INDEX	−0.67417***		−0.39672*		−0.97300***	−0.48580**
AGE_INDEX					−0.00324**	
POP_GRID	−0.01909***		−0.01133***		−0.03855***	
AGR_COS	−0.44250***	−0.35647***	−0.43232***	−0.23051 *	−0.64592***	
EUC_COS	0.22741***	0.48727***	0.26358 ***		−0.34679***	0.44333***
GRZ_COS	0.55225***			0.41351 ***	0.55249***	
OUTR_COS		0.19743***			−0.36986***	
SHRUB_COS	0.24486***	0.50703***	0.54626***		0.08008***	0.49115***
AP2015_DIST	0.10367**		−0.35040***		1.10158***	0.28249***
PROAD_DIST			1.14231*		1.18853***	
DRYMONTH	−2.57044**					−8.75216**
SLOPE	0.61038***		1.52079***	2.24835*	0.57521***	1.33822***
ASPECT	−0.02087*	−0.07253***			−0.03574**	
ELEVATION		0.02316**			−0.01151***	−0.03843***
IGN_DIST	−1.16798***	−2.40021***	−2.83019***		−1.38465***	

the urban coastline clusters in the second part of the models (Table 5). For the northern region, a cell affected by a fire over 100 ha and located 15 km from an ignition location would have a reduction in its expected burned area percentage of ~42 percentage points.

Discussion

The results of this study have shown that central Portugal is composed of five different clusters with respect to the distribution of the main structural factors driving the propagation of large fires. This study also found that there is spatial variability in the contribution exerted by most structural factors to burning probability, with different results across clusters; additionally, the influence of a given variable on cell burning is not necessarily similar to the influence of the same variable on burning extent, which in turn may also vary depending on location.

Additionally, some of the most striking drivers of fire presence and burned area were identified as the type of vegetation, terrain slope and human presence, specifically activities connected to agriculture and livestock.

The overall clustering partition is coherent with previous knowledge on the characteristics of central Portugal, namely, the urban coastline–rural inland dichotomy. Larger burned areas in Portugal are concentrated in the regions covered by the North-eastern Highlands and Central Forests clusters (Tedim *et al.* 2013; Benali *et al.* 2016). Nunes *et al.* (2013) discussed the essential role of socioeconomic transformations, such as the rural exodus, and favourable climatic conditions (higher precipitation) for the proliferation of available fuel in uncultivated areas in these regions.

The characteristics of the coastal areas contrast with these patterns, particularly in the north-western pocket of the study

area (Northern Urban Coastline). Although ignitions are concentrated along the coast and around urban areas, higher population density, fragmented settlements, intense agricultural activities, higher accessibility and fire suppression efforts prevent the occurrence of large wildfire events (Barros and Pereira 2014).

Forests are considered more fire-prone than farms, with eucalyptus plantations displaying the same fire hazard as pine stands (Moreira *et al.* 2011), a pattern observed across clusters in our study. Nevertheless, effects connected to the economic value and active management of paper and pulp production in eucalyptus forests may mitigate the influence of fuel availability and other favourable burning conditions (Barros and Pereira 2014). Additionally, it is interesting to note that according to both of these authors, shrublands are more susceptible to fire than forests, which may explain why the effect of eucalyptus stands on burning probability was negative in the shrub-dominated region of the North-eastern Highlands. Shrublands are mentioned as one of the most significant wildland cover varieties in the mountainous areas of northern and central Portugal (North-eastern Highlands and Central Forests), specifically in connection to their high flammability and preferential burning (Fernandes *et al.* 2010). These vegetation features are coupled with a rugged landscape, higher fuel connectivity and lower population density, favouring fire spread (Mateus and Fernandes 2014).

Regarding climate factors, and contrary to previous knowledge on this subject, which states that a longer dry season increases the chances of fire spread (Ganteaume and Guerra 2018), only in the Northern Urban Coastline should a higher number of dry months increase burning probability, suggesting that the influence of drought on large burned area should be assessed using different data. In fact, according to Parisien *et al.* (2014), models making use of temporal averages in climate variables fail to capture time-related variability and tend to highlight the overall importance of land cover factors.

Slope is known to influence fire spread both directly and indirectly because of fire dynamics (flames closer to ground fuel) and fuel moisture and density patterns (Holsinger *et al.* 2016). The findings of the present study are consistent with this information.

However, the pronounced relationship found between an increase in elevation and an increase in large fire spread probability, such as that in the Southern Urban Coastline, can be explained by the concentration of urban areas at low altitudes. Evidence from Australia has found that fires tend to become larger at higher altitudes, this fact being connected to limited fire suppression activities at such elevations (Price *et al.* 2016). The same pattern has also been confirmed in the European Mediterranean context (Sande Silva *et al.* 2010). The results obtained from the models seem to indicate that elevation promotes large fire spread in densely populated regions (such as the Southern Urban Coastline), while in inland rural regions, large wildfire events occur at lower altitudes, where there is more available fuel. These findings are not entirely consistent with previous knowledge on this subject, meaning further investigation into the effects of this topographic factor is needed.

Many arguments support the findings from this study, which generally agree that agricultural activities work as a deterrent factor for large burned areas. Wildfires occurring in agricultural land cover are not expected to develop into high-intensity events, mainly owing to a small fuel load of predominantly dry fine fuels (Mitsopoulos *et al.* 2015). Several authors link the process of agriculture, pasture and forestry abandonment that occurred in European Mediterranean regions during the second half of the 20th century to an increase in the size and intensity of wildfires as a result of an accumulation of flammable materials (Moreira *et al.* 2011; Viedma *et al.* 2015; Calviño-Cancela *et al.* 2016).

Another reason explaining the low probability of fire spreading into agricultural spaces, or a higher burning probability in wildland areas, has to do with the interaction between agriculture and topography. Farms are usually located in flatlands, and slope is known to strongly affect fire spread (Calviño-Cancela *et al.* 2017). Agricultural activities are also associated with the presence of humans in rural areas, which is associated with earlier fire detection and more efficient firefighting (Moreira *et al.* 2011), as farm management requires that holders be vigilant and mindful of their property (Calviño-Cancela *et al.* 2016). Nunes *et al.* (2016) found the same overall patterns as the present work regarding the relationship between agricultural activities and reduced burned area in a study conducted at a municipal level in Portugal.

Livestock activities and grazing have been found by Moreira *et al.* (2011) to contribute to reduced fire hazard because these activities naturally control fuel availability and density. This fact may help justify the negative effect displayed by animal density in some clusters. In this context, it is essential to consider animal type for the influence exerted by different species, as previously mentioned by Oliveira *et al.* (2014). This said, evidence from the literature highlights the different impacts that pastures and livestock have on large wildfire occurrence.

Oliveira *et al.* (2017) found that Portuguese parishes affected by larger burned areas still had a considerable presence of livestock and grazing. In a way, higher burned area percentages in rangelands might be explained by fuel availability. Additionally, many authors have highlighted the role of fire in clearing land for grazing purposes in Mediterranean areas and specifically in Portugal (Ferreira-Leite *et al.* 2013a; Ganteaume and Jappiot 2013; Álvarez-Díaz *et al.* 2015; Vilar *et al.* 2016).

Protected areas are generally assumed to influence fire occurrence, either as a deterrent factor connected to landscape protection (Rodrigues *et al.* 2014, 2016) or as a factor related to an increase in ignitions resulting from conflicts about the establishment of these protected sites (Fuentes-Santos *et al.* 2013; Calviño-Cancela *et al.* 2017). The results from our study are consistent with the views of Srivastava *et al.* (2014) on this topic, which associate protected areas with increased success in efficient wildfire suppression.

The outcomes of this study suggest that large wildfires tend to spread into demographically depressed regions, with other authors supporting this claim. Communities affected by larger burned areas in Portugal are known to suffer from depopulation, as young and educated people migrate from inland and mountainous areas to coastal regions, which in turn renders these locations increasingly vulnerable to fire (Oliveira *et al.* 2017). The influence of population growth potential, which is generally connected to human presence and density of human activities,

has been found to reduce the probability of larger burned areas in Spain, although it increases the probability of fire ignition (Martínez-Fernández *et al.* 2013; Rodrigues and De la Riva 2014).

Similarly to the findings of this study, ignition locations have been found to impact fire spread in Portugal, although it has also been suggested that it mostly depends on complex interactions with biophysical elements and that it varies considerably across regions (Benali *et al.* 2016).

Conclusions

Overall, this study shows that, between 2005 and 2015, central Portugal displayed a prevalence of large fire events in inland central areas, with larger burned areas rarely occurring near the coastline. It confirms the existence of spatial variability in the impact of most structural factors driving the propagation of large fires in this area while also emphasising that the contributions of these factors may vary depending on whether we are modelling the probability of cell burning or the percentage of burned area. Moreover, vegetation type, slope and human activities such as agriculture and keeping livestock are found to impact large fire propagation the most across clusters. It is vital that upcoming studies in the field of large wildfires, specifically in the Portuguese context, account for these crucial features.

Based on the main conclusions of this work and some of the methodological issues faced, two research topics can be proposed for future development. One is to compare the results of this study with those stemming from the use of data mining techniques, especially regarding the development of predictive models. This methodology may be more suitable given the high data volume (both variables and observations). Another possibility is to select a condensed set of variables for a more detailed assessment of the effect of different elements in large wildfire spread, namely, those factors whose results seem to contradict previous knowledge in this field.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

References

Álvarez-Díaz M, González-Gómez M, Otero-Giraldez MS (2015) Detecting the socioeconomic driving forces of the fire catastrophe in NW Spain. *European Journal of Forest Research* **134**, 1087–1094. doi:10.1007/S10342-015-0911-1

Balsa Barreiro J, Hermosilla T (2013) Socio-geographic analysis of wildland fires: CAUSES of the 2006's wildfires in Galicia (Spain). *Forest Systems* **22**, 497–509. doi:10.5424/FS/2013223-04165

Barros AMG, Pereira JMC (2014) Wildfire selectivity for land cover type: does size matter? *PLoS One* **9**. doi:10.1371/JOURNAL.PONE.0084760

Benali A, Ervilha AR, Sá ACL, Fernandes PM, Pinto RMS, Trigo RM, Pereira JMC (2016) Deciphering the impact of uncertainty on the accuracy of large wildfire spread simulations. *The Science of the Total Environment* **569–570**, 73–85. doi:10.1016/J.SCITOTENV.2016.06.112

Buntin MB, Zaslavsky AM (2004) Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. *Journal of Health Economics* **23**, 525–542. doi:10.1016/J.JHEALECO.2003.10.005

Calviño-Cancela M, Chas-Amil ML, García-Martínez ED, Touza J (2016) Wildfire risk associated with different vegetation types within and outside wildland–urban interfaces. *Forest Ecology and Management* **372**, 1–9. doi:10.1016/J.FORECO.2016.04.002

Calviño-Cancela M, Chas-Amil ML, García-Martínez ED, Touza J (2017) Interacting effects of topography, vegetation, human activities and wildland-urban interfaces on wildfire ignition risk. *Forest Ecology and Management* **397**, 10–17. doi:10.1016/J.FORECO.2017.04.033

Cao X, Cui X, Yue M, Chen J, Tanikawa H, Ye Y (2013) Evaluation of wildfire propagation susceptibility in grasslands using burned areas and multivariate logistic regression. *International Journal of Remote Sensing* **34**, 6679–6700. doi:10.1080/01431161.2013.805280

Cleff T, Cleff T (2014) 'Univariate data analysis.' (Pearson Education Limited). doi:10.1007/978-3-319-01517-0_3

Costa L, Thonicke K, Poulter B, Badeck F (2011) Sensitivity of Portuguese forest fires to climatic, human, and landscape variables: SUBNATIONAL differences between fire drivers in extreme fire years and decadal averages. *Regional Environmental Change* **11**, 543–551. doi:10.1007/S10113-010-0169-6

Cruz-Jesus F, Oliveira T, Bacao F (2012) Digital divide across the European Union. *Information & Management* **49**, 278–291. doi:10.1016/J.IM.2012.09.003

European Forum for Geography and Statistics, Eurostat (2019) GSGF Europe – Implementation guide for the Global Statistical Geospatial Framework in Europe – Proposal from the GEOSTAT 3 project. (Stockholm)

Farewell VT, Long DL, Tom BDM, Yiu S, Su L (2017) Two-part and related regression models for longitudinal data. *Annual Review of Statistics and Its Application* **4**, 283–315. doi:10.1146/ANNUREV-STATISTICS-060116-054131

Fernandes PM, Luz A, Loureiro C (2010) Changes in wildfire severity from maritime pine woodland to contiguous forest types in the mountains of north-western Portugal. *Forest Ecology and Management* **260**, 883–892. doi:10.1016/J.FORECO.2010.06.008

Fernandes PM, Monteiro-Henriques T, Guiomar N, Loureiro C, Barros AMG (2016) Bottom–up variables govern large-fire size in Portugal. *Ecosystems* **19**, 1362–1375. doi:10.1007/S10021-016-0010-2

Ferreira-Leite F, Bento-Gonçalves A, Lourenço L, Úbeda X, Vieira A (2013*a*) Grandes incêndios florestais em Portugal continental como resultado das perturbações nos regimes de fogo no mundo Mediterrâneo. *Silva Lusitana* **21**, 127–142.

Ferreira-Leite F, Lourenço L, Bento-Gonçalves A (2013*b*) Large forest fires in mainland Portugal, brief characterization. *Méditerranée* **121**, 53–65. doi:10.4000/MEDITERRANEE.6863

Fuentes-Santos I, Marey-Pérez MF, González-Manteiga W (2013) Forest fire spatial pattern analysis in Galicia (NW Spain). *Journal of Environmental Management* **128**, 30–42. doi:10.1016/J.JENVMAN.2013.04.020

Ganteaume A, Guerra F (2018) Explaining the spatio-seasonal variation of fires by their CAUSES: The case of south-eastern France. *Applied Geography (Sevenoaks, England)* **90**, 69–81. doi:10.1016/J.APGEOG.2017.11.012

Ganteaume A, Jappiot M (2013) What causes large fires in southern France. *Forest Ecology and Management* **294**, 76–85. doi:10.1016/J.FORECO.2012.06.055

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. doi:10.1007/S00267-012-9961-Z

Goudie A (2006) 'The human impact on the natural environment – past, present and future.' (Blackwell Publishing: Oxford)

Grala K, Grala RK, Hussain A, Cooke WH, Varner JM (2017) Impact of human factors on wildfire occurrence in Mississippi, United States. *Forest Policy and Economics* **81**, 38–47. doi:10.1016/J.FORPOL.2017.04.011

Hernandez C, Drobinski P, Turquety S (2015) How much does weather control fire size and intensity in the Mediterranean region? *Annales Geophysicae* **33**, 931–939. doi:10.5194/ANGEO-33-931-2015

Holsinger L, Parks SA, Miller C (2016) Weather, fuels, and topography impede wildland fire spread in western US landscapes. *Forest Ecology and Management* **380**, 59–69. doi:10.1016/J.FORECO.2016.08.035

Keeley JE, Syphard AD (2016) Climate change and future fire regimes: examples from California. *Geosciences* **6**, 37. doi:10.3390/GEOSCIENCES6030037

Keeley JE, Syphard AD (2018) Historical patterns of wildfire ignition sources in California ecosystems. *International Journal of Wildland Fire* **27**, 781–799. doi:10.1071/WF18026

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen–Geiger climate classification updated. *Meteorologische Zeitschrift (Berlin)* **15**, 259–263. doi:10.1127/0941-2948/2006/0130

Martínez-Fernández J, Chuvieco E, Koutsias N (2013) Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. *Natural Hazards and Earth System Sciences* **13**, 311–327. doi:10.5194/NHESS-13-311-2013

Mateus P, Fernandes PM (2014) Forest fires in Portugal: dynamics, causes and policies. 'Forest context and policies in Portugal. Present and future challenges'. (Ed F Reboredo) pp. 117–154. (Springer). doi:10.1007/978-3-319-08455-8.

Mhawej M, Faour G, Adjizian-Gerard J (2015) Wildfire likelihood's elements: a literature review. *Challenges* **6**, 282–293. doi:10.3390/CHALLE6020282

Miller C, Ager AA (2013) A review of recent advances in risk analysis for wildfire management. *International Journal of Wildland Fire* **22**, 1–14. doi:10.1071/WF11114

Mitsopoulos I, Mallinis G, Arianoutsou M (2015) Wildfire risk assessment in a typical Mediterranean wildland–urban interface of Greece. *Environmental Management* **55**, 900–915. doi:10.1007/S00267-014-0432-6

Moreira F, Catry FX, Rego F, Bacao F (2010) Size-dependent pattern of wildfire ignitions in Portugal: when do ignitions turn into big fires? *Landscape Ecology* **25**, 1405–1417. doi:10.1007/S10980-010-9491-0

Moreira F, Viedma O, Arianoutsou M, Curt T, Koutsias N, Rigolot E, Barbati A, Corona P, Vaz P, Xanthopoulos G, Mouillot F, Bilgili E (2011) Landscape – wildfire interactions in southern Europe: implications for landscape management. *Journal of Environmental Management* **92**, 2389–2402. doi:10.1016/J.JENVMAN.2011.06.028

Moreno JM, Vallejo VR, Chuvieco E (2013) Current fire regimes, impacts and the likely changes – VI: Euro Mediterranean. In 'Vegetation fires and global change. Challenges for concerted international action. A White Paper directed to the United Nations and International Organizations'. (Ed JG Goldammer) pp. 115–131. (Kessel Publishing House: Remagen-Oberwinter)

Nunes A, Lourenço L, Bento-Gonçalves A, Vieira A (2013) Três décadas de incêndios florestais em Portugal: incidência regional e principais fatores responsáveis. *Caderno de Geografia* **32**, 133–143. doi:10.14195/0871-1623_32_12

Nunes AN, Lourenço L, Meira ACC (2016) Exploring spatial patterns and drivers of forest fires in Portugal (1980–2014). *The Science of the Total Environment* **573**, 1190–1202. doi:10.1016/J.SCITOTENV.2016.03.121

Oliveira SLJ, Pereira JMC, Carreiras JMB (2012) Fire frequency analysis in Portugal (1975–2005), using Landsat-based burnt area maps. *International Journal of Wildland Fire* **21**, 48–60. doi:10.1071/WF10131

Oliveira S, Pereira JMC, San-Miguel-Ayanz J, Lourenço L (2014) Exploring the spatial patterns of fire density in southern Europe using Geographically Weighted Regression. *Applied Geography (Sevenoaks, England)* **51**, 143–157. doi:10.1016/J.APGEOG.2014.04.002

Oliveira S, Zêzere JL, Queirós M, Pereira JM (2017) Assessing the social context of wildfire-affected areas. The case of mainland Portugal. *Applied Geography (Sevenoaks, England)* **88**, 104–117. doi:10.1016/J.APGEOG.2017.09.004

Parisien M-A, Parks SA, Krawchuk MA, Little JM, Flannigan MD, Gowman LM, Moritz MA (2014) An analysis of controls on fire activity in boreal Canada: comparing models built with different temporal resolutions. *Ecological Applications* **24**, 1341–1356. doi:10.1890/13-1477.1

Price OF, Penman T, Bradstock R, Borah R (2016) The drivers of wildfire enlargement do not exhibit scale thresholds in south-eastern Australian forests. *Journal of Environmental Management* **181**, 208–217. doi:10.1016/J.JENVMAN.2016.06.033

Rego F, Silva JS (2014) Wildfires and landscape dynamics in Portugal: a regional assessment and global implications. In 'Forest landscapes and global change'. (Eds JC Azevedo, AH Perera, MA Pinto) pp. 227–248. (Springer: New York). doi:10.1007/978-1-4939-0953-7.

Rodrigues M (2015) Review and new methodological approaches in human-caused wildfire modeling and ecological vulnerability: risk modeling at mainland Spain. Universidad de Zaragoza. Available at https://dialnet.unirioja.es/servlet/dctes?codigo=99499.

Rodrigues M, De la Riva J (2014) An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software* **57**, 192–201. doi:10.1016/J.ENVSOFT.2014.03.003

Rodrigues M, de la Riva J, Fotheringham AS (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. doi:10.1016/J.APGEOG.2014.01.011

Rodrigues M, Jiménez A, de la Riva J (2016) Analysis of recent spatial–temporal evolution of human driving factors of wildfires in Spain. *Natural Hazards* **84**, 2049–2070. doi:10.1007/S11069-016-2533-4

San-Miguel-Ayanz J, Moreno JM, Camia A (2013) Analysis of large fires in European Mediterranean landscapes: lessons learned and perspectives. *Forest Ecology and Management* **294**, 11–22. doi:10.1016/J.FORECO.2012.10.050

Sande Silva J, Rego F, Fernandes P, Rigolot E (Eds) (2010) 'Towards integrated fire management – outcomes of the European Project Fire Paradox.' (European Forest Institute)

Sharma S (1996) 'Applied multivariate techniques.' (John Wiley & Sons: New York). doi:10.2307/1270777

Srivastava SK, Saran S, de By RA, Dadhwal VK (2014) A geo-information system approach for forest fire likelihood based on causative and anti-causative factors. *International Journal of Geographical Information Science* **28**, 427–454. doi:10.1080/13658816.2013.797984

Tedim F, Remelgado R, Borges C, Carvalho S, Martins J (2013) Exploring the occurrence of mega-fires in Portugal. *Forest Ecology and Management* **294**, 86–96. doi:10.1016/J.FORECO.2012.07.031

Turco M, Jerez S, Augusto S, Tarín-Carrasco P, Ratola N, Jiménez-Guerrero P, Trigo RM (2019) Climate drivers of the 2017 devastating fires in Portugal. *Scientific Reports* **9**, 13886. doi:10.1038/S41598-019-50281-2

Verde JC, Zêzere JL (2010) Assessment and validation of wildfire susceptibility and hazard in Portugal. *Natural Hazards and Earth System Sciences* **10**, 485–497. doi:10.5194/NHESS-10-485-2010

Viedma O, Moity N, Moreno JM (2015) Changes in landscape fire-hazard during the second half of the 20th century: agriculture abandonment and the changing role of driving factors. *Agriculture, Ecosystems & Environment* **207**, 126–140. doi:10.1016/J.AGEE.2015.04.011

Vilar L, Nieto H, Martín MP (2010) Integration of lightning- and human-caused wildfire occurrence models. *Human and Ecological Risk Assessment* **16**, 340–364. doi:10.1080/10807031003670469

Vilar L, Gómez I, Martínez-Vega J, Echavarría P, Riaño D, Martín MP (2016) Multitemporal modelling of socio-economic wildfire drivers in central Spain between the 1980s and the 2000s: comparing generalized linear models to machine learning algorithms. *PLoS One* **11**, e0161344. doi:10.1371/JOURNAL.PONE.0161344

Wooldridge JM (2012) 'Introductory econometrics: a modern approach.' (South-Western, Cengage Learning)