

1                    **Estimating the probability of wildfire**  
2 **occurrence in Mediterranean landscapes using**  
3                    **Artificial Neural Networks**

4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

39 **Abstract**

40

41 Wildfires are a major disturbance in the Mediterranean Basin and an ecological factor that constantly  
42 alters the landscape. In this context, it is crucial to understand where wildfires are more likely to occur  
43 as well as the drivers guiding them in complex landscapes such as the Mediterranean area. The  
44 objectives of this study are to estimate wildfire probability occurrence as a function of biophysical  
45 and human-related drivers, to provide an assessment of the relative impact of each driver and analyze  
46 the performance of machine learning techniques compared to traditional regression modeling. By  
47 employing an Artificial Neural Network model and fire data (2004-2012), we estimated wildfire  
48 probability across two geographical regions covering most of the Italian territory: Alpine and  
49 subalpine region and Insular and peninsular region. The high classification accuracy (0.68 for the  
50 Alpine and subalpine region and 0.76 for the Insular and peninsular region) and good performances  
51 of the technique (AUC values of 0.82 and 0.76, respectively) suggest that our model can be used in  
52 the areas studied to assess wildfire probability occurrence. We compared our model with a logistic  
53 function, which showed a weaker predictive power (AUC values of 0.78 for the Alpine and subalpine  
54 region and 0.65 for the Insular and peninsular region) compared to the Artificial Neural Network. In  
55 addition, we assessed the importance of each variable by isolating it in the model. The importance of  
56 an individual variable differed between the two regions, underscoring the high diversity of wildfire  
57 occurrence drivers in Mediterranean landscapes. Results show that in the Alpine and subalpine region,  
58 the presence of forest is the most important variable, while climate resulted as being the most  
59 important variable in the Insular and peninsular region. The majority of areas recently affected by  
60 large wildfires in both regions have been correctly classified by the ANN model as ‘high fire  
61 probability’. Hence, the use of an Artificial Neural Network is efficient and robust for understanding  
62 the probability of wildfire occurrence in Italy and other similar complex landscapes.

63

64

65

66

67

68 *Key words:* Wildfire probability, Machine Learning, Logistic function, Wildfire drivers,

69

70

71

72

73

74

75

76

77

78

79

## 80 **1. Introduction**

81

82 Fire disturbance is a key driver of many natural landscapes and for the delivery of ecosystem services  
83 (Johnstone et al., 2016; Molina et al., 2019). However, wildfires have detrimental effects on natural  
84 resources and human life when they occur in urban interfaces (Argañaraz et al., 2017; Modugno et  
85 al., 2016; San-Miguel-Ayanz et al., 2013).

86 Reports of the European Commission suggest that over the past 30 years Europe has seen an increase  
87 of extreme wildfire events generating major socio-ecological impacts (Elia et al., 2016; Lozano et al.,  
88 2017; Paveglio et al., 2018). In Italy, the magnitude of the wildfire dilemma is similar to that of other  
89 Mediterranean countries (Carlucci et al., 2019; Mancini et al., 2018b). In 2017 alone, more than 7,800  
90 wildfires occurred in the peninsula burning over 162,000 ha. Statistics also reveal that the average  
91 number of wildfire casualties is 5 per year, while the mean number of injured is 39 per year (Union,  
92 2018). Despite the continuous support of the European Commission and the efforts of national and  
93 regional government to improve fire management policies, these data depict a dramatic picture. The  
94 numbers remind us of the importance and urgency to integrate the emergency approach (e.g., fire  
95 suppression) with a more efficient preventive fire management strategy, specifically focused on  
96 favoring the development of fire-resistant and resilient landscapes (Moreira et al., 2020; Twidwell et  
97 al., 2019). To this end, landscape management agencies need to understand where wildfires are most  
98 likely to occur (i.e., wildfire probability) as well as the key drivers of current wildfires (Faivre et al.,  
99 2014; Guo et al., 2017; Pricope and Binford, 2012; Rodrigues et al., 2016).

100 The scientific community has developed many approaches to disentangle the multifactorial aspects  
101 that lead to wildfire probability in a specific landscape. For example, the first pioneering study by  
102 Chuvieco and Congalton (1989) suggests integrating geo-environmental data and logistic regression  
103 to assess wildfire probability. Preisler et al. (2004) published a study presenting a probability-based  
104 model for estimating wildfire occurrences. Subsequent to these first studies wildfire probability  
105 estimation became an increasingly popular research theme, fostering a large variety of innovative  
106 approaches worldwide (Amatulli et al., 2013; Oliveira et al., 2012; Jaafari et al., 2017; Mancini et al.  
107 2018a; Michetti and Pinar 2019).

108 In this regard, the use of machine learning techniques (MLTs) can greatly improve the understanding  
109 of wildfire probability in complex territories as the Mediterranean (Jain et al., 2020). For example,  
110 Artificial Neural Network (ANN) models, in comparison to simpler models (e.g., linear regression),

111 have the ability to explore a set of existing non-linear relationships in the data leading to improved  
112 model accuracy (Yang et al., 2006).

113 Previous studies have employed ANNs to investigate fire danger (Bisquert et al., 2012; Pai et al.,  
114 2020), wildfire vulnerability (Dimuccio et al., 2011; de Bem et al., 2019), wildfire risk assessment  
115 (Li et al., 2009; Jafari Goldarag et al., 2016; Lall and Mathibela, 2016), burned area detection (Maeda  
116 et al., 2009; Gómez and Martín, 2011), pre- and post-fire vegetation (Debouk et al., 2013; Polinova  
117 et al., 2019), causes of wildfires (Rodrigues and de la Riva, 2014), and flame and smoke detection  
118 (Chetehouna et al., 2015; Hossain et al., 2019). However, the literature currently does not provide a  
119 satisfactorily large number of studies in which ANNs have been employed to estimate the probability  
120 of wildfire occurrence in many cases the scale of analysis is restricted to small geographical areas  
121 (Vega Garcia et al., 1996; Vasilakos et al., 2009; Ruiz-Mirazo et al., 2012; Safi and Bouroumi, 2013;  
122 Satir et al., 2016).

123 To fill these gaps, we developed an ANN model to estimate the probability of wildfire occurrence in  
124 the complex Mediterranean landscape of Italy. The Italian landscape represents a suitable testbed for  
125 our study given its wide variety of vegetation types, topographical and ecological features in  
126 heterogeneous urbanization contexts, and different weather conditions. The specific objectives of the  
127 study are: (1) to estimate the probability of wildfire occurrence as a function of biophysical and  
128 human-related drivers; (2) to assess the relative magnitude of each driver; and (3) to analyze the  
129 performance of the ANN model compared to traditional regression modeling.

130 Because of varying fire regimes and the difficulty in obtaining a robust model for the entire country,  
131 the Italian peninsula was divided into two main study areas: (i) the Alpine and subalpine region (ASR)  
132 and (ii) Insular and peninsular region (IPR). The models and relative validations were applied to each  
133 region. The findings represent a further step toward a better understanding of wildfire probability  
134 occurrence, which can be useful for other related studies across the globe.

135

## 136 **2. Materials and Methods**

137

### 138 *2.1 Study areas*

139

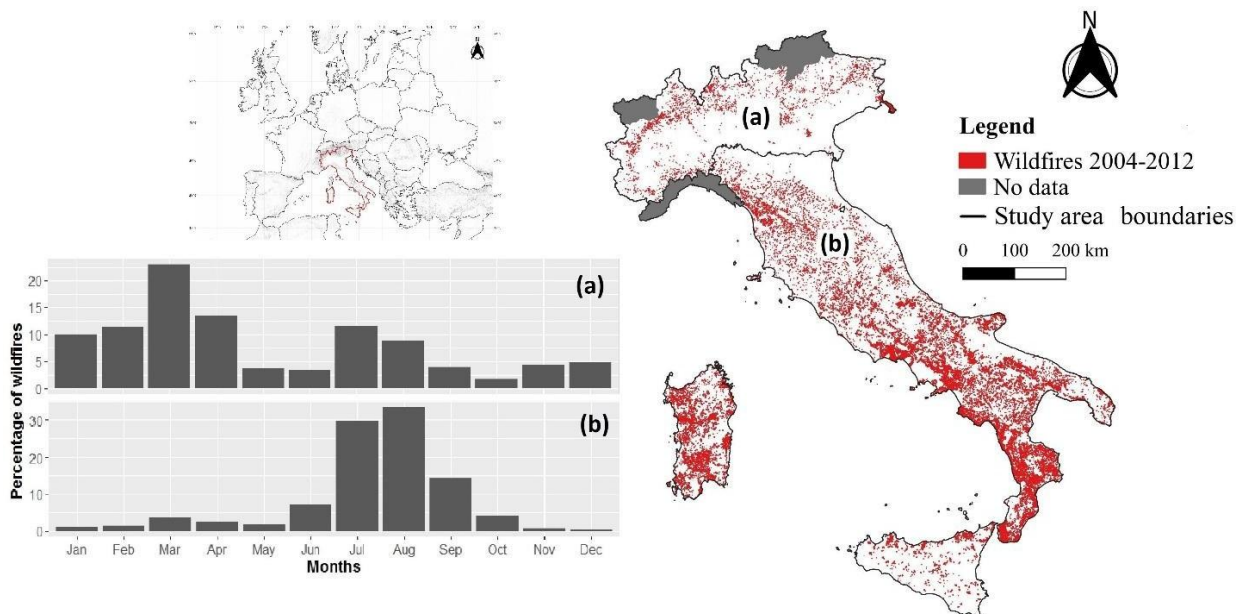
140 Italy is located at a northern latitude between 36° and 47°30' and an eastern longitude between 5°30'  
141 and 18°30', extending for the most part into the Mediterranean Sea (Fig. 1). The territory covers an  
142 area of approximately 301,330 km<sup>2</sup>, 23% of which is classified as lowland, 42% as upland, and 35%  
143 as mountainous landscape. Italy is crossed by two important mountain chains – the Alps in the North  
144 and the Apennines in the south-central region. This orography gives rise to a climate gradient from

145 North to South, ranging from Mediterranean warm to temperate cool. Most of the remaining region  
 146 is surrounded by the Mediterranean Sea creating conditions for the presence of a wide variety of flora  
 147 and fauna species. Forest resources in Italy extend over 10.9 million ha (RAF, 2019). According to  
 148 Habitat Directive no. 43/92 of the European Commission, 32% of forest lands cover the Alpine bio-  
 149 geographical region, 16% the Continental region, and 52% the Mediterranean region.

150 Italy is one of the southern European countries most affected by wildfires. Its fire regime changes  
 151 while proceeding from North to South (Conedera et al., 2018). In the ASR, the majority of wildfires  
 152 in recent decades have mainly occurred in the first three months of the year (from January to March),  
 153 reaching a peak in March of 1,000 events (Fig.1-a). This winter fire regime is due to the continental  
 154 climate characterized by cold-dry winters, fully cured vegetation and frequent episodes of strong  
 155 warm-dry winds (foehn) that further dry out vegetation and make it fire prone (Valese et al., 2014).

156 On the contrary, in the IPR wildfires mostly occur in summer (third quarter of the year), reaching a  
 157 maximum of about 14,000 events (Fig.1-b). During this season the climate is of the Mediterranean  
 158 type characterized by minimum precipitation in July, dry winds from North Africa and high  
 159 temperatures causing the loss of fuel moisture, hence increasing the probability of fire ignition and  
 160 spread (Michetti and Pinar, 2019).

161  
 162



163  
 164  
 165 **Fig. 1** - Location of the study areas in Italy (black line indicates the boundaries of the Alpine and  
 166 subalpine region and Insular and peninsular region) within the Mediterranean Basin, and map of  
 167 wildfires during the period of investigation. The histograms show the number of fire events

168 (percentage) across months for each study area: (a) Alpine and subalpine region; (b) Insular and  
169 peninsular region.

170

## 171 *2.2 Response variable*

172

173 To estimate the response variable, we used historical wildfire georeferenced polygons derived from  
174 the Comando Unità Forestali, Ambientali e Agroalimentari (CUFAA), Carabinieri Force, and forest  
175 services of Autonomous Regions. The response variable represents wildfire occurrence  
176 (presence/absence) in each 1-km<sup>2</sup> grid cell of the two study regions. If the fire polygon or a portion  
177 of it fell in a 1-km<sup>2</sup> cell we considered it as presence of wildfire, whereas if it did not it was considered  
178 as absence of wildfire occurrence. We opted to use the 2004 to 2012 time period for our study, which  
179 displays the most harmonized data throughout Italy. The Liguria region, the Valle d'Aosta region and  
180 the Autonomous Province of Bolzano were excluded from the analysis due to the difficulty in  
181 obtaining a complete spatial and temporal dataset of wildfire events for the entire territory.

## 182 *2.3 Biophysical explanatory variables*

183

184 One of the most important steps in probability analysis is building a set of explanatory variables based  
185 on their potential relation to the response variable (wildfire occurrence) and data availability (Table  
186 1). Therefore, we collected human-related and biophysical variables and transformed each variable  
187 into a continuous scale at a 1-km<sup>2</sup> resolution grid (Elia et al., 2019; Guo et al., 2016).

188 To characterize land cover, the Corine Land Cover initiative 1:100,000 (2012) was adopted, which  
189 ensured a complete mapping of the Italian landscape at the fourth hierarchical level (CLC, 2012). We  
190 aggregated the land cover classes by creating eight variables, each corresponding to a specific group  
191 of land use such as agriculture, wetland, waterland, forest, grassland, shrubland, otherland, and urban.  
192 We then calculated the presence (percentage) of each group within the 1-km<sup>2</sup> grid cells.

193 Another important variable affecting the probability of wildfire occurrence is tree cover (Satir et al.  
194 2016). In Mediterranean coniferous forests tree cover might be positively related to crown fire  
195 behavior, while dense broadleaved forests could hamper fire spread due to the high moisture content  
196 in the understory and the limited presence of flammable grasses and shrubs. We derived tree cover  
197 from the Copernicus Land Monitoring Service (<https://land.copernicus.eu/>). The product consisted of  
198 status layers (for 2012) showing the level of tree cover density in a range from 0-100%; these layers  
199 were then converted into a 1-km<sup>2</sup> resolution grid scale.

200 Topographical variables were selected for their relevance to wildfire occurrence based on previous  
201 research and data availability. The topographical features of a landscape heavily affect species  
202 composition, the microclimate and fire behavior (e.g., stack effect) (Syphard et al., 2008).  
203 Furthermore, previous studies (Gralewicz et al., 2012) have found that the higher the altitude, the  
204 fewer the occurrences of wildfires. Elevation and slope were derived from the Digital Elevation  
205 Model at European level (Reuter et al., 2007) and reclassified at a 1-km<sup>2</sup> resolution grid scale.

206 With regard to climatic variables, we used an index combining information on different weather  
207 parameters in the period of investigation (2004-2012). Data on maximum wind speed, maximum  
208 temperature and relative humidity in the two study areas were collected daily. Climate data for the  
209 months of January, February, March and April were obtained for the ASR, while the same data for  
210 June, July, August and September were collected for the IPR. Consequently, for each region we  
211 estimated a mean value of the three climatic parameters for the entire time period of investigation.  
212 Once the parameters were collected, a Fire Climate Index (FCI) was estimated for each 1-km<sup>2</sup> cell of  
213 the two study areas (Fig. 2) based on previous studies (Barbero et al., 2015; Fox et al., 2015; Hamadeh  
214 et al., 2017; Satir et al., 2016; Sharples et al., 2009). This index was calculated using the following  
215 formula:

$$216 \text{ Fire Climate Index} = \max(U)/FMI \quad [1]$$

217 where  $U$  represents the wind speed (km h<sup>-1</sup>) and  $FMI$  stands for the fuel moisture index. The  $FMI$  was  
218 developed by Sharples et al. (2009) using the following formula:

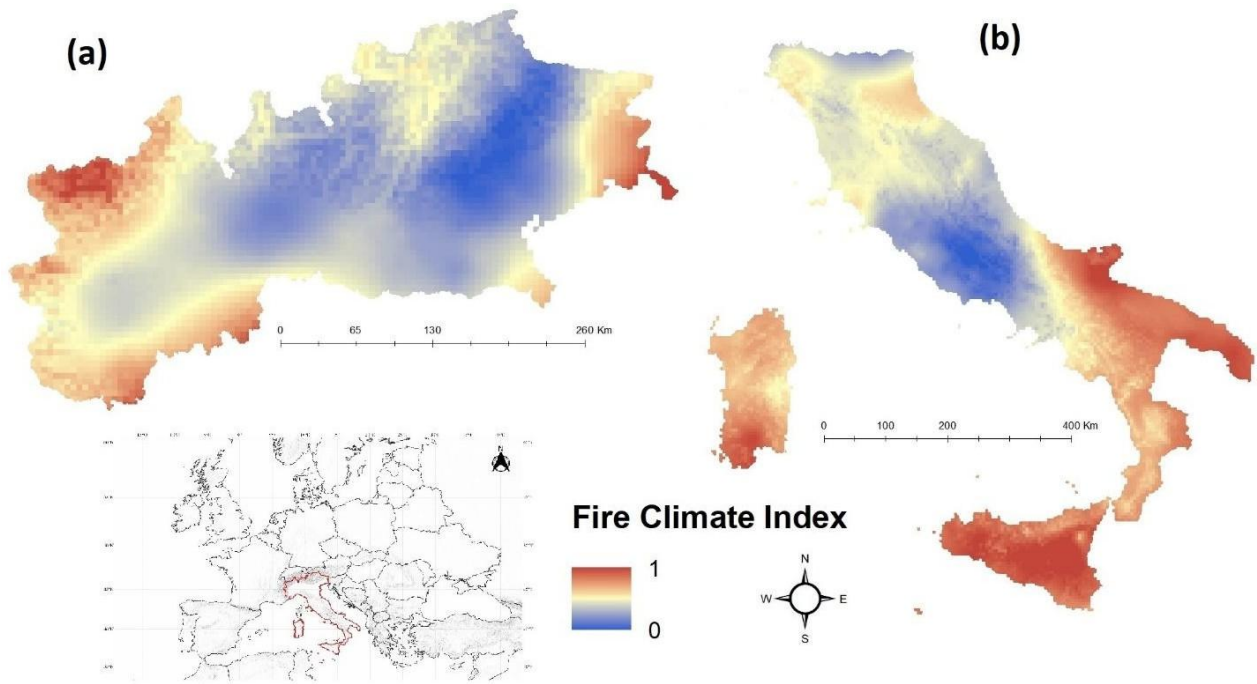
$$219 FMI = 10 - 0.25 (T - H) \quad [2]$$

220 where  $T$  is the temperature (°C) and  $H$  stands for the relative humidity (%).

221 All the data were downloaded from the SCIA (National System for the collection, processing and  
222 dissemination of climate data) website  
223 ([http://www.scia.isprambiente.it/wwwrootscia/Home\\_new\\_eng.html](http://www.scia.isprambiente.it/wwwrootscia/Home_new_eng.html)) of the Italian Institute for  
224 Environmental Protection and Research (ISPRA). The website provides main climatic parameters  
225 that can be downloaded and displayed in the form of tables, diagrams and maps.

226

227



228

229 **Fig. 2** - The Fire Climate Index estimated for each km<sup>2</sup> cell within the two study regions: (a) Alpine  
 230 and subalpine region; (b) Insular and peninsular region.

231

232

#### 233 *2.4 Human-related explanatory variables*

234

235 Most wildfires in Italy are linked to human activity. For example, fires may ignite as a result of pasture  
 236 renewal or the burning of stubble and then spread to nearby forest patches. These practices  
 237 demonstrate the need to include human-related variables in our analytical model to explain the  
 238 presence of anthropogenic activity. Based on previous studies (Lein and Stump, 2009; Maingi and  
 239 Henry, 2007; Ricotta and Di Vito, 2014) and the available data, we opted to consider three main  
 240 human-related predictors: major roads, railways, and distance from human settlements (e.g., houses,  
 241 industrial areas, airports) (Table 1). The layers were extracted from the Open Street Map website on  
 242 a scale of 1:50,000. From these layers we derived the raster distance from major roads, the distance  
 243 from railways and from settlements resampled at a 1-km<sup>2</sup> resolution grid. All the datasets were  
 244 processed for the ASR and IPR.

245

246 **Table 1.** Overview of the explanatory variables selected to perform the study.

247

Data	Input	Source	Output	Coding
<b>Climate</b>	Relative humidity (%)	SCIA	Fire Climate Index	FCI
	Absolute maximum temperature (°C)			
	Maximum wind (m/s)			
<b>Anthropogenic</b>	Road maps	Open Street maps	Distance from roads, settlements and railways Population density map	Road_dist, Urban_dist Rail_dist, Pop_density
	Rail maps			
	Settlement locations	Gallego, 2010		
	Population			
<b>Topographic</b>	DTM	National Geoportal	Digital elevation map	Elev, slope
	Slope (%)		Slope map	
<b>Landscape</b>	Corine Land Cover	Copernicus Program	Corine Classes Percentage	Urban, Agric, Shrubland, Forest, Wetland, Waterland, Grassland, Otherland, Cover_tree
	Tree cover density (%)		Tree canopy Percentage	

248 DTM = Digital Terrain Model; SCIA = National System for the collection, processing and  
249 dissemination of climate data.

250

### 251 *2.5 Pre-processing and model selection*

252

253 For each study region, a predictive model of wildfire probability occurrence was built using MLTs.  
254 Pre-processing is an essential step in machine learning consisting of normalization, data split and  
255 balancing of the database. The data were normalized in a range from 0 to 1 to homogenize the entire  
256 dataset. Before fitting the model, the original dataset was divided into training (70% of data) and  
257 testing (30% of data) sets. Because our dataset was strongly unbalanced with a major absence of  
258 wildfires (0) and a minor presence of wildfire occurrences (1), the training set of each area was  
259 subjected to an under-sampling technique, i.e. Random Over-Sampling Examples (ROSE). This  
260 technique consists in randomly under-sizing the most represented class (Liu et al., 2009; Menardi and  
261 Torelli, 2014). Subsequently, the training set was used to train the model algorithm while the test set  
262 was used for its validation.

263 Among the MLTs we chose to employ ANNs, which are mathematical models of supervised learning  
264 (Abiodun et al., 2018). ANNs take inspiration from modeling the human brain and try to replicate its  
265 structure. Nodes or neurons are the basic units of neural networks. They combine data inputs with

266 coefficients, or weights, which amplify or reduce the importance of inputs according to the algorithm.  
267 For each node, the sum of the inputs multiplied by their weights passes through an activation function  
268 which determines how this signal influences the results. The selection of the optimal values for the  
269 weights is referred to as the “training phase”, where the model is trained using a resilient back  
270 propagation algorithm with weight backtracking.

271 For ASR the “*neuralnet*” package of R was applied. We employed one hidden layer and a logistic  
272 activation function [Eq. 3] setting the value between 0 and 1.

$$273 \quad AF = \frac{1}{(1+\exp(-x))} \quad [3]$$

274 In this case the training set, consisting of 64547 observations of which 61448 were classified as  
275 absence and 3099 as presence, was under-sampled as previously mentioned to reduce the number of  
276 non-occurrences to 3099.

277 In this case the training set, consisting of 64,547 observations of which 61,448 were classified as  
278 absence of wildfires and 3,099 as presence of wildfires, was under-sampled as previously mentioned  
279 to reduce the number of wildfire non-occurrences to 3,099. For the IPR it was necessary to apply a  
280 deep neural network due to the size of the area and large amount of data. In this case, we used “*keras*”  
281 package for R (<https://cran.r-project.org/web/packages/keras/index.html>). Two hidden layers were  
282 employed, and a sequential model was created containing two activation functions: the *Relu* function  
283 for the intermediate hidden layers, and the *Sigmoid* activation function to obtain output from 0 to 1.  
284 The optimization algorithm used was *Adam* (Adaptive Momentum estimation), and binary cross-  
285 entropy was used as loss of function. In this case the training set consisting of 143,200 observations,  
286 of which 118,129 were classified as “absence” and 25,071 as “presence” of wildfires, was under-  
287 sampled with the same technique to reduce the number of wildfire non-occurrences to 25,071.

288

289

## 290 *2.6 Model performance and variable importance assessment*

291

292 The model performances were assessed using the confusion matrix and Receiver Operating  
293 Characteristic (ROC) curve test method. The area under the curve (AUC) was employed as it is one  
294 of the most common statistical methods adopted to estimate model fitting (Guo et al., 2017; Jiménez-  
295 Valverde, 2012; Vilar del Hoyo et al., 2011). For AUC values  $\geq 0.7$  the predictors indicated good  
296 performance in predicting the dependent variable (Elia et al., 2019). Accuracy was also considered in  
297 assessing the correct classification; results were then compared and validated as a function of their  
298 accuracies.

299 The ANN models were validated through cross-validation. Model calibrations were performed five  
 300 times on random subsamples of the training and test sets and the metrics estimated for each  
 301 subsample. In addition, we attempted to describe the explanatory power of the topographic variables  
 302 by isolating them in the overall models (Fig. 3). Variable importance can be estimated by observing  
 303 how much the score, AUC in our study, decreases when a feature is not used in the model calibration  
 304 process. Importance is represented by subtracting the estimated AUC value without the considered  
 305 variable from the total AUC of the all variables.

306 Lastly, we developed a logistic model with our dataset and compared the results with the ANN models  
 307 on the basis of the AUC parameter for each study area. We also developed probability maps for both  
 308 models and discussed them from a management perspective.

### 309 **3. Results**

#### 311 *3.1 Artificial neural networks and variable importance*

312 After the ANNs were constructed and trained using the test set (70% of the total dataset), the  
 313 remaining 30% of the data was used to assess ANN performance. Table 2 summarizes the  
 314 performance metrics for the ANN model used to estimate the probability of wildfire occurrence for  
 315 each study region. Overall, the performances of the ANNs across each study area were robust. Of the  
 316 two areas, the ASR recorded the highest AUC value (0.82) and lowest accuracy value (0.68) in  
 317 contrast to the IPR, respectively.

319 **Table 2** – Performance metrics of the Artificial Neural Network model for each study area

	AUC	Accuracy
Alpine and subalpine region (ASR)	0.82	0.68
Insular and peninsular region (IPR)	0.76	0.76

320 AUC=area under the curve.

321

322 The cross-validations confirmed model robustness (Table 3). The results showed similar metrics  
 323 values among the five subsamples of training and test set models for each study area, indicating that  
 324 the overall model did not show explicit overfitting.

325

326

327 **Table 3** – Performance metrics in five random subsamples of the training and test sets for each study  
 328 area

AUC		Accuracy	
ASR	IPR	ASR	IPR

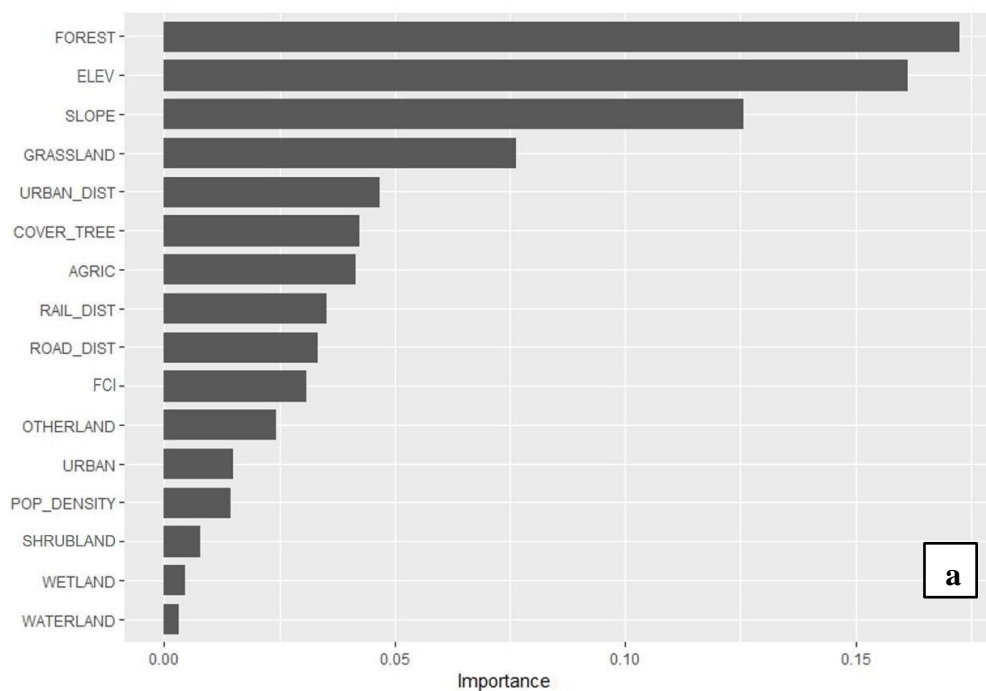
Sample 1	0.82	0.76	0.69	0.78
Sample 2	0.82	0.76	0.69	0.80
Sample 3	0.83	0.76	0.69	0.79
Sample 4	0.82	0.76	0.68	0.78
Sample 5	0.82	0.76	0.70	0.79

329 AUC = area under the curve;  
330 ASR = Alpine-subalpine region;  
331 IPR = Insular and peninsular region  
332  
333

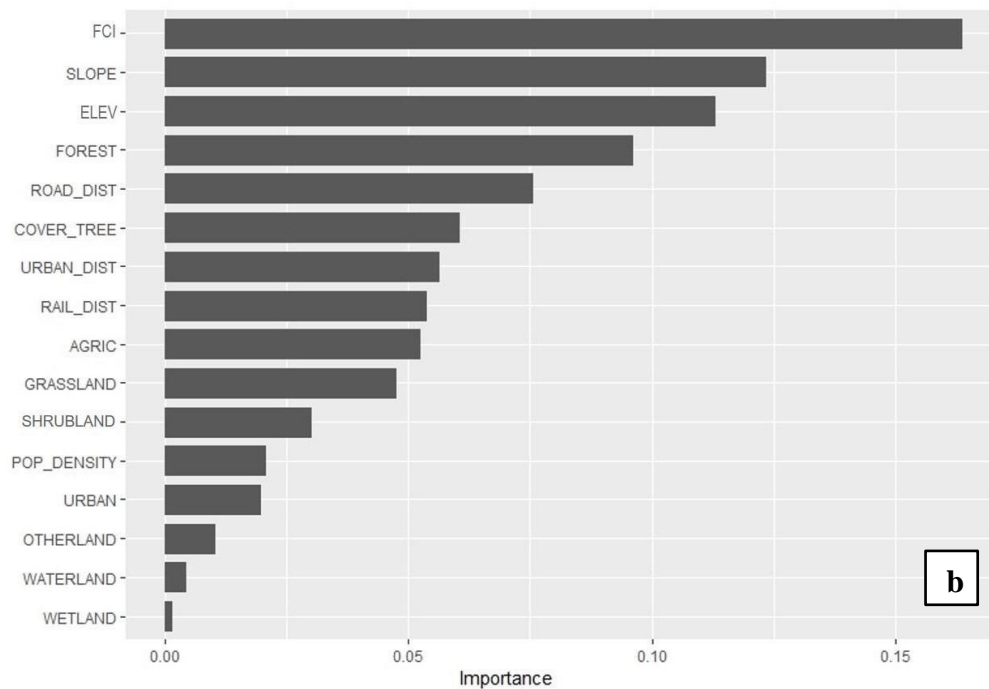
334 In the ASR the presence of forest showed the highest importance with a value of 0.17, followed by  
335 the topographic variables, Elevation and Slope with values of 0.16 and 0.13, respectively (Fig. 3-a).  
336 In addition, the human-related variables exhibited a lower value of importance in comparison to the  
337 IPR. For example, by eliminating road distance the AUC decreased by a value of 0.03, which is less  
338 than that of the IPR (0.07).

339 In the IPR, the FCI and topographic variables, Slope and Elevation, yielded high importance values,  
340 i.e., 0.16, 0.12 and 0.11, respectively, which were greater than for the remaining predictors (Fig. 3-  
341 b). The other variables related to land cover and human-related drivers exhibited a lower value of  
342 importance (~0.02) for population density and urban lands.

343



344



345

346

347 **Fig. 3** - Explanatory power of the variables through model isolation for the (a) Alpine and subalpine  
 348 region and (b) Insular and peninsular region. Variable names are also found in Table 1.

349

350

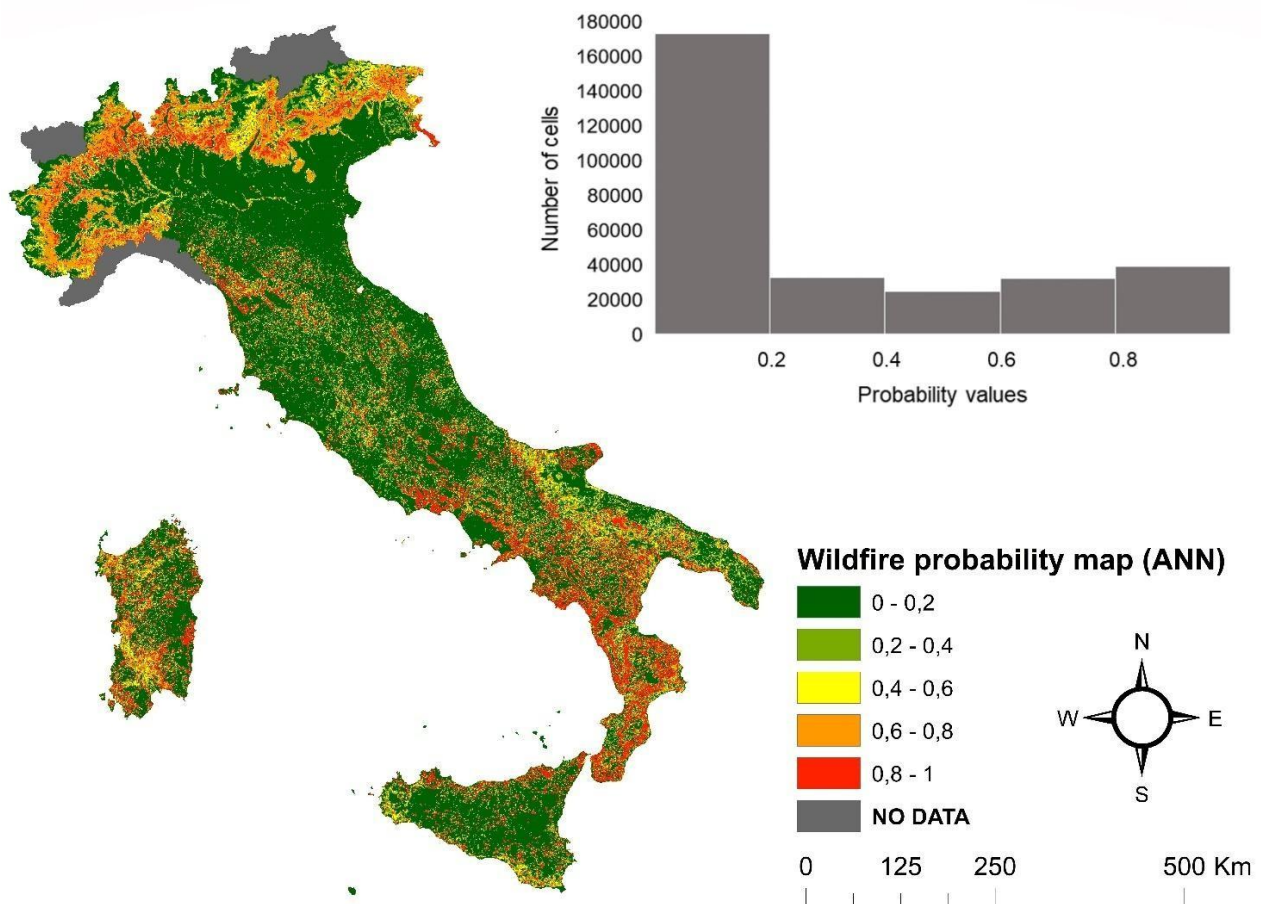
### 351 *3.2 Probability of wildfire occurrence*

352 The high classification accuracy and good performances of ANNs suggest that ANN models can be  
 353 used in the two studied regions to estimate wildfire probability (Fig. 4). In the ASR, the model  
 354 highlighted the high probability of wildfire occurrence throughout the entire esalpic mountain belt of  
 355 the southern European Alps, starting from the Maritime Alps at the western-most side to the southern  
 356 limestone Alps in the Eastern Alps. Interestingly, the model discriminated between higher wildfire  
 357 probability in the lower esalpic part of the Alpine valleys and a lower probability in the higher inland  
 358 area, a gradient that is particularly marked in the Eastern Alps. Furthermore, the model consistently  
 359 correlated recurrent extensive wildfires with several well-known hotspots such as the slopes  
 360 surrounding major Alpine lakes of the Insubric region and the coast of the Gulf of Trieste at the  
 361 eastern-most side of the Alpine region. The map also exhibited high wildfire probability occurrence  
 362 across the southern slopes of the lower Susa valley in the southwestern Alps where in the summer of  
 363 2017 the largest wildfire in Italy occurred.

364 With regard to the IPR, an increasing North-South gradient in wildfire probability occurrence was  
 365 evident with some isolated hotspots. This gradient is mostly due to southern summer climatic  
 366 conditions that trigger dramatic increases in temperature, dryness and wind speed compared to

367 northern areas. However, the ANN model effectively detected relevant isolated hotspots in central-  
 368 north Italy, such as the hills west of Florence, the Conero reserve south of Ancona, or the mountain  
 369 ridge of Monti Pisani, which features one of the most flammable forests in the Tuscany region; here,  
 370 in 2018, the largest fire event of the last 30 years took place. Notably, the model correctly classified  
 371 the area in the IPR with the highest wildfire probability, which comprises the mountain ridge of Monti  
 372 Aurunci, (the highest fire recurrence rate in Italy, i.e., three fire events during the study period), and  
 373 high fire-prone areas such as the Tyrrhenian coast from Cilento to Calabria and the Ogliastra region  
 374 of the island of Sardegna. Similarly, the model effectively identified isolated hotspots such as Vesuvio  
 375 National Park, where in 2017 a large wildfire burned 44% of the protected area (Espinosa et al., 2018),  
 376 and the coastline south of Peschici, where in 2007 one of the most dramatic and largest wildfires  
 377 occurred in Italy's history in terms of human fatalities.

378



379

380 **Fig. 4** – Map of wildfire probability occurrence generated by the Artificial Neural Network (ANN)  
 381 model across the entire peninsula of Italy merging the Alpine and subalpine region and Insular and  
 382 peninsular region. The histograms indicate cell distribution according to the probability of wildfire  
 383 occurrence.

384

385

386 *3.3 Comparison with the logistic model*

387 As stated above (section 2.6), we developed a logistic (Logit) function to compare our ANN models  
388 (Table 4). Based on the AUC parameter, the ANN models showed a higher predictive power than the  
389 Logit function in each of the two study areas. The major difference was found in IPR, where the AUC  
390 value of our ANN models was 0.76 versus 0.65 for the Logit function. More information about the  
391 Logit function (e.g., coefficient of explanatory variables and relative P-values) is available in the  
392 Supplementary materials (Tables S1 and S2).

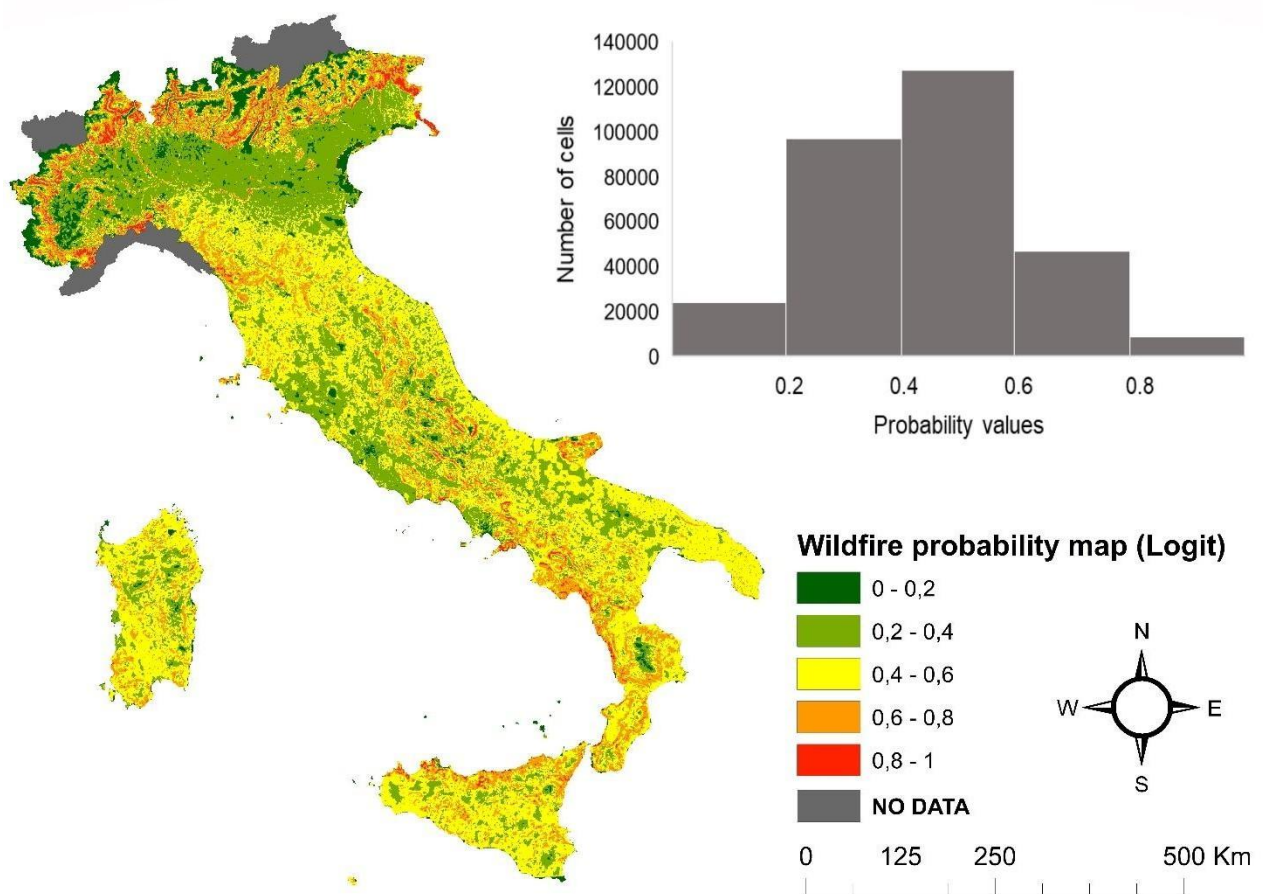
393 With regard to the ANNs, the map generated by the Logit model (Fig. 5) revealed remarkable  
394 differences in estimations of wildfire probability occurrence. The Logit model seemed to overestimate  
395 the values across the landscape. For example, the logistic map suggested an average probability value  
396 of 0.35 for the Pianura Padana. However, the area is mostly dominated by agriculture and few  
397 wildfires were recorded in the past due to lack of forest fuel prone to wildfire occurrence. On the  
398 contrary, the ANN map showed that the same area exhibited an average probability value of 0.06,  
399 which is much lower. This finding suggests a more accurate performance of the ANN in  
400 discriminating areas with high probability from areas with low probability of wildfire occurrence.

401 **Table 4** – Comparison between the logistic and ANN models using the AUC parameter.

AUC		
Model	ASR	IPR
ANNs	0.82	0.76
Logit	0.78	0.65

402 AUC = area under the curve; ANNs = Artificial Neural Networks; ASR = Alpine and subalpine region; IPR =  
403 Insular and peninsular region; Logit = logistic.

404



405

406

407 **Fig. 5** – Map of probability of wildfire occurrence generated by the logistic (Logit) model across the  
 408 entire territory of Italy by merging the Alpine and subalpine region and Insular and peninsular region.  
 409 The histograms indicate cell distribution according to the probability of wildfire occurrence.

410

411

#### 412 **4. Discussion**

413

414 Using novel approaches to understand the main predictors of wildfire occurrence is crucial in the  
 415 context of broader wildfire risk assessment (Jaafari et al., 2019; Laforteza et al., 2015; Polinova et  
 416 al., 2019). Hence, the aim of this study was to increase our knowledge of wildfire probability  
 417 occurrences in Italy using MLTs. Previous studies have also employed MLTs for estimating wildfire  
 418 probability in Mediterranean areas. However, our investigation is one of the first to apply ANN  
 419 models at a regional scale to better understand the impact of biophysical and human-related drivers  
 420 on the probability of wildfire occurrence in the complex Italian landscape.

421 Consistent with other studies (Oliveira et al., 2012; Syphard et al., 2008; Vilar et al., 2019), our model  
422 suggests that the probability of wildfire occurrence is affected by both biophysical and human-related  
423 drivers and that a non-linear trend exists between them and the variable response. For example,  
424 Olivera et al. (2012) adopted a MLT (Random Forest) to explain fire density in Europe. The authors  
425 found a non-linear relationship between predictors and fire density and suggested that a non-  
426 parametric model could be more suitable to explain the response variable.

427 The comparison between the ANN model and logistic function (see Table 4) showed a more robust  
428 predictive power for ANNs in both study areas, while logistic regression demonstrated weaker  
429 predictive power between the response variable and predictors. More specifically, our results indicate  
430 a better performance (i.e., higher AUC) for ANNs compared to the logistic function, suggesting the  
431 ability of ANNs to depict the spatial variability of wildfires in Italy based on high landscape  
432 heterogeneity (i.e., biophysical and human-driven variability) and fire regimes. Our results are  
433 consistent with those of previous studies focusing on differences between the logistic function and  
434 ANNs. For instance, Jafari Goldarag et al. (2016) found that accuracy was low (50.84%) for logistic  
435 regression compared to ANNs (92.3%) in developing fire prediction maps in Iran. In predicting fire  
436 occurrences in Brazil, de Bem et al. (2019) found AUC values of 0.77 and 0.75 using an ANN  
437 approach and logistic regression, respectively. Bisquert et al. (2012) used the ANN approach to  
438 investigate fire occurrences in Galicia (Spain) and found an average accuracy of 75%. Other authors  
439 as well have proven the soundness of ANN in estimating wildfire occurrences. For example, two  
440 studies in Lebanon (Hamadeh et al., 2015) analyzed the effects of climatological data on forest fire  
441 occurrences. Their results are consistent with ours, since they found an accuracy and an AUC value  
442 of 94% and 98%, respectively. Similarly, Sakr et al. (2011) demonstrated the robust capacity of ANN  
443 to correctly estimate the probability of fire occurrence with an accuracy of 90%.

#### 444 *4.1 Probability of wildfire occurrence and variable importance*

445 In the Italian peninsula characterized by different ecological features and patterns of urbanization,  
446 environmental and anthropogenic elements play differently across the ASR and IPR study areas.  
447 Therefore, it is crucial to collect a comprehensive set of biophysical and human-related drivers  
448 (Prasad et al., 2008). The topographic variables of the ASR and IPR displayed a common pattern in  
449 terms of importance. Slope and elevation significantly influenced the probability of wildfire  
450 occurrence ranking high as variables of importance in both study areas (Fig. 4). This is most likely  
451 correlated to the positive influence of slope (e.g., chimney effect) and altitude on fire behavior and  
452 spread (Butler et al., 2007). Furthermore, it is rare to find wildland plains with flammable vegetation  
453 in Italy; the majority of wildfires occur in mountainous areas or in areas where the slope is steep.

454 Previous studies have highlighted the strong influence of elevation on fire occurrence (Ajin et al.,  
455 2016). Wildfire occurrences in areas at a high altitude can be directly influenced by consistent sun  
456 exposure and increased lightening (de Bem et al., 2018). Our results are consistent with those of  
457 Mancini et al. (2018b), who found that elevation significantly influences both fire frequency and  
458 incidence in Italy (.

459 The study results also show that the two land cover variables proportion of forest and grassland  
460 present high importance values in the ASR. This region is characterized by dense forest cover of  
461 relatively flammable fuel complexes, such as understory of chestnut, oak, and pine plantations, which  
462 dry out during the winter because of minimal precipitation and foehn winds (Valese et al., 2014). As  
463 expected, the forest variable was important in both study areas, most likely due to the substantial  
464 amount of fuel available to burn and to the continuity and connectivity of forested landscapes. This  
465 finding has highlighted the importance of one of the main drivers of fire regimes in Italy, which in  
466 the last century has been forest expansion consequent to agriculture and grazing abandonment in  
467 mountainous and hilly areas (Bovio et al., 2017). Grassland was the more important variable in the  
468 ASR. Grasslands are mostly composed of flash fuels and are easy to ignite, thus their association to  
469 fire occurrence is somewhat predictable. In fact, during the winter season the herbaceous layers in  
470 this region are severely dried by the freezing winds blowing from northern Europe, thus becoming  
471 prone to fires.

472 The presence of shrublands was surprisingly less important than the above-mentioned land cover  
473 variables in both study areas, a finding that contrasts with previous studies. For example, Sebastián-  
474 López et al. (2008) considered shrubs as the principal predictor of fire danger in their model for  
475 southern Europe. Other studies (Moreira et al., 2011; Nunes et al., 2005) have provided evidence that  
476 shrubland is usually the land cover which is most fire prone in Mediterranean ecosystems. In our  
477 study areas, the presence of shrubland was less representative than forest and grassland and therefore  
478 exhibited lower predictive power. However, this outcome may contrast with local results suggesting  
479 that shrublands are more prone to burn if adjacent to roads and urban areas (Elia et al., 2020).

480 From a climatic perspective, FCI was recognized in the analysis as the most influential variable in  
481 the IPR. Conversely, the FCI did not attribute much importance to the ASR, ranking 10<sup>th</sup> as a variable  
482 of importance among all (Fig.3-a). As expected, in the IPR during the fire season dry and hot summers  
483 make fuel prone to ignition and create flammable conditions. In addition, strong warm winds from  
484 North Africa push wildfires and in many cases lead to extreme and dangerous outcomes

485 Some predictors did not exhibit much importance individually, and therefore it was easier to assess  
486 their impact on probability of wildfire occurrence within a group of variables. For instance, human-

487 related variables such as distance from roads, railways and human settlements had an evident impact  
488 on the maps derived by the model. Proximity to the above-mentioned variables increases the  
489 probability of wildfire occurrence, even in areas of low population density (Bar Massada et al., 2013).  
490 The gradient of higher wildfire probability in the lower esalpic part of the Alpine valleys toward a  
491 lower probability in the higher inland area is a typical feature of fire regimes in the Alps (Valese et  
492 al., 2014), which is partly related to decreasing population density and lower foehn intensity in the  
493 upper valleys. Unexpectedly, the ANN model ranked 5<sup>th</sup> as a variable of importance among the  
494 different anthropic predictors for each study area. In the ASR, the human-related variable urban  
495 distance ranked 5<sup>th</sup> mostly due to the fact that the majority of wildfires originate close to the urban  
496 interface at the bottom of valleys, but then propagate and spread up the slopes affecting forests at  
497 higher elevations than most inhabited areas. Moreover, in this region a significant portion of fires  
498 originate from pastoral burns to maintain grazing areas (Ascoli and Bovio, 2013) at the top of  
499 mountain ridges distant from urban areas.

500 In the IPR, the first human-related variable considered was road distance. Many authors have pointed  
501 out how road distance can affect wildfire occurrences in the Mediterranean region. Others still  
502 (Cardille et al., 2001; Faivre et al., 2014; Gralewicz et al., 2012; Jaafari et al., 2018) have found  
503 significant correlations between wildfire occurrences and proximity to roads. Maingi and Henry  
504 (2007) estimated that in the US (Appalachian counties of Eastern Kentucky) distance to roads  
505 explained 54% of the total variation observed in wildfire occurrences. Elia et al. (2020) found that  
506 distance to roads influenced the likelihood and frequency of wildfires in southern Europe.

#### 507 *4.2 Management implications*

508 Wildfire prevention, suppression and mitigation are critical issues for forest managers and decision  
509 makers because of the stochastic variability of the phenomenon across space. A deeper understanding  
510 of the wildfire phenomenon will inform about where wildfires are likely to occur and the drivers  
511 guiding potential new occurrences. This need has stimulated research efforts on wildfire probability  
512 studies, especially in Italy where the landscape is characterized by heterogeneous ecosystems from  
513 North to South and from coastlines to mountainous hinterlands.

514 Once areas with the highest wildfire probability are detected by ANN models, forest managers can  
515 use the resulting maps to prioritize fire management interventions (Elia et al., 2014). Additionally,  
516 by adopting these maps decision makers can develop civil protection plans, particularly in areas where  
517 natural and human systems mix (e.g., wildland urban interfaces) and with a high probability of  
518 wildfire occurrence.

519 **5. Conclusions**

520

521 In this study we applied an ANN model to estimate the probability of wildfire occurrence in Italy  
522 using a comprehensive set of biophysical and human predictors. The findings demonstrate that in a  
523 complex landscape such as the Italian peninsula, characterized by a large variety of anthropic and  
524 environmental features, the use of ANNs is efficient and statistically robust for understanding the  
525 probability of wildfire occurrence. Our model, in fact, suggests that the importance of a single  
526 variable differs along the North to South gradient, which underscores the high variability of fire  
527 drivers in a changing landscape. In addition, compared to the logistic function the ANN model  
528 produced a higher AUC value and demonstrated greater accuracy when evaluating wildfire  
529 probability.

530 Wildfire probability estimation using ANN models in the Mediterranean Basin still offers wide room  
531 for improvement. Although our method has been applied to a given landscape (Italy) and over a  
532 certain time period (2004-2012), it has the potential to be employed for longer periods of time and in  
533 cross-regional areas. This would require an adaptation of the ANN algorithm, thus involving a wide  
534 range of network architectures. Additionally, we recommend further investigation of the relationship  
535 between explanatory variables and the probability of wildfire occurrence by focusing on new  
536 predictors (e.g., socio-economic) or by further examining those used in the present work.

537 A further intent of our work was to corroborate prior studies in the field of machine learning  
538 techniques to understand wildfire probability occurrence. For this purpose, the operational use of the  
539 above-mentioned algorithms might be worth investigating in the future. A key role in this regard is  
540 played by the scientific community for both the development of new models and the transmission of  
541 knowledge to the operative world (e.g., for fire risk assessment in Regional Fire Management Plans  
542 in Italy, Art. 3 – Law 353/2000), that still favors traditional approaches rather than models with “black  
543 boxes” as ANNs (Yang et al., 2006).

544

545 **Acknowledgements**

546

547 This study has been developed within the F.OR.MA. (Fuel models and fORest roads Maps) project  
548 in collaboration with the Civil Protection Department of the Apulia Region. The author wish to thank  
549 Yole DeBellis for reviewing the manuscript.

550

551 **References**

552

- 553 Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H., 2018. State-of-the-art in  
554 artificial neural network applications: A survey. *Heliyon* 4.  
555 <https://doi.org/10.1016/j.heliyon.2018.e00938>
- 556 Ajin, A., Loghin, A.-M., Mathew K. Jacob, P.g, V., R.r, K., 2016. The Risk Assessment Study of Potential Forest  
557 Fire in Idukki Wildlife Sanctuary using RS and GIS Techniques. *Int. J. Adv. Earth Sci. Eng.* 5, 308–318.
- 558 Amatulli, G., Camia, A., San-Miguel-Ayanz, J., 2013. Estimating future burned areas under changing climate  
559 in the EU-Mediterranean countries. *Sci. Total Environ.* 450–451, 209–222.  
560 <https://doi.org/10.1016/j.scitotenv.2013.02.014>
- 561 Argañaraz, J.P., Radeloff, V.C., Bar-Massada, A., Gavier-Pizarro, G.I., Scavuzzo, C.M., Bellis, L.M., 2017.  
562 Assessing wildfire exposure in the Wildland-Urban Interface area of the mountains of central  
563 Argentina. *J. Environ. Manage.* 196, 499–510. <https://doi.org/10.1016/j.jenvman.2017.03.058>
- 564 Ascoli, D., Bovio, G., 2013. Prescribed burning in Italy: issues, advances and challenges. *IForest -*  
565 *Biogeosciences For.* 6, 79. <https://doi.org/10.3832/ifor0803-006>
- 566 Bar Massada, A.B., Syphard, A.D., Stewart, S.I., Radeloff, V.C., 2013. Wildfire ignition-distribution modelling:  
567 a comparative study in the Huron–Manistee National Forest, Michigan, USA. *Int. J. Wildland Fire* 22,  
568 174–183. <https://doi.org/10.1071/WF11178>
- 569 Barbero, R., Abatzoglou, J.T., Larkin, N.K., Kolden, C.A., Stocks, B., 2015. Climate change presents increased  
570 potential for very large fires in the contiguous United States. *Int. J. Wildland Fire* 24, 892–899.  
571 <https://doi.org/10.1071/WF15083>
- 572 Bisquert, M., Caselles, E., Sánchez, J.M., Caselles, V., 2012. Application of artificial neural networks and  
573 logistic regression to the prediction of forest fire danger in Galicia using MODIS data. *Int. J.*  
574 *Wildland Fire* 21, 1025. <https://doi.org/10.1071/WF11105>
- 575 Bovio, M, M., L, T., M, S., G, V., R, L., M, E., P, F., D, A., 2017. Gli incendi boschivi stanno cambiando:  
576 cambiamo le strategie per governarli. *For. - J. Silvic. For. Ecol.* 14, 202.  
577 <https://doi.org/10.3832/efor2537-014>
- 578 Butler, B.W., Anderson, W.R., Catchpole, E.A., 2007. Influence of slope on fire spread rate. *Butl. Bret W*  
579 *Cook Wayne Comps Fire Environ.-Innov. Manag. Policy Conf. Proc.* 26-30 March 2007 Destin FL  
580 *Proc. RMRS-P-46CD Fort Collins CO US Dep. Agric. For. Serv. Rocky Mt. Res. Stn. CD-ROM P 75-82*  
581 046.
- 582 Cardille, J.A., Ventura, S.J., Turner, M.G., 2001. Environmental and Social Factors Influencing Wildfires in the  
583 Upper Midwest, United States. *Ecol. Appl.* 11, 111–127. [https://doi.org/10.1890/1051-0761\(2001\)011\[0111:EASFIW\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2001)011[0111:EASFIW]2.0.CO;2)
- 585 Carlucci, M., Zambon, I., Colantoni, A., Salvati, L., 2019. Socioeconomic Development, Demographic  
586 Dynamics and Forest Fires in Italy, 1961–2017: A Time-Series Analysis. *Sustainability* 11, 1305.  
587 <https://doi.org/10.3390/su11051305>
- 588 Chetehouna, K., Tabach, E.E., Bouazaoui, L., Gascoin, N., 2015. Predicting the flame characteristics and rate  
589 of spread in fires propagating in a bed of *Pinus pinaster* using Artificial Neural Networks. *Process*  
590 *Saf. Environ. Prot.* 98, 50–56. <https://doi.org/10.1016/j.psep.2015.06.010>
- 591 Chuvieco, E., Aguado, I., Jurdao, S., Pettinari, M.L., Yebra, M., Salas, J., Hantson, S., Riva, J. de la, Ibarra, P.,  
592 Rodrigues, M., Echeverría, M., Azqueta, D., Román, M.V., Bastarrika, A., Martínez, S., Recondo, C.,  
593 Zapico, E., Martínez-Vega, F.J., 2014. Integrating geospatial information into fire risk assessment.  
594 *Int. J. Wildland Fire* 23, 606–619. <https://doi.org/10.1071/WF12052>
- 595 Chuvieco, E., Congalton, R.G., 1989. Application of remote sensing and geographic information systems to  
596 forest fire hazard mapping. *Remote Sens. Environ.* 29, 147–159. [https://doi.org/10.1016/0034-4257\(89\)90023-0](https://doi.org/10.1016/0034-4257(89)90023-0)
- 598 CLC, 2012. Corine Land Cover 2012 IV livello — Italiano [WWW Document]. URL  
599 [http://www.sinanet.isprambiente.it/it/sia-ispra/download-mais/corine-land-cover/corine-land-](http://www.sinanet.isprambiente.it/it/sia-ispra/download-mais/corine-land-cover/corine-land-cover-2012-iv-livello/view)  
600 [cover-2012-iv-livello/view](http://www.sinanet.isprambiente.it/it/sia-ispra/download-mais/corine-land-cover/corine-land-cover-2012-iv-livello/view) (accessed 1.29.20).

601 Conedera, M., Krebs, P., Valesse, E., Cocca, G., Schunk, C., Menzel, A., Vacik, H., Cane, D., Japelj, A., Muri, B.,  
602 Ricotta, C., Oliveri, S., Pezzatti, G.B., 2018. Characterizing Alpine pyrogeography from fire statistics.  
603 *Appl. Geogr.* 98, 87–99. <https://doi.org/10.1016/j.apgeog.2018.07.011>  
604 de Bem, P.P., de Carvalho Júnior, O.A., Matricardi, E.A.T., Guimarães, R.F., Gomes, R.A.T., 2019. Predicting  
605 wildfire vulnerability using logistic regression and artificial neural networks: a case study in Brazil's  
606 Federal District. *Int. J. Wildland Fire* 28, 35. <https://doi.org/10.1071/WF18018>  
607 Debouk, H., Riera-Tatché, R., Vega-García, C., 2013. Assessing Post-Fire Regeneration in a Mediterranean  
608 Mixed Forest Using Lidar Data and Artificial Neural Networks. *Photogramm. Eng. Remote Sens.* 79,  
609 1121–1130. <https://doi.org/10.14358/PERS.79.12.1121>  
610 Deeming, J.E., 1972. National Fire-Danger Rating System. Rocky Mountain Forest and Range Experiment  
611 Station, Forest Service, U.S. Department of Agriculture.  
612 Dimuccio, L.A., Ferreira, R., Cunha, L., Almeida, A.C. de, 2011. Regional forest-fire susceptibility analysis in  
613 central Portugal using a probabilistic ratings procedure and artificial neural network weights  
614 assignment. *Int. J. Wildland Fire* 20, 776–791. <https://doi.org/10.1071/WF09083>  
615 Elia, M., Giannico, V., Laforteza, R., Sanesi, G., 2019. Modeling fire ignition patterns in Mediterranean  
616 urban interfaces. *Stoch. Environ. Res. Risk Assess.* 33, 169–181. [https://doi.org/10.1007/s00477-](https://doi.org/10.1007/s00477-018-1558-5)  
617 [018-1558-5](https://doi.org/10.1007/s00477-018-1558-5)  
618 Elia, M., Giannico, V., Spano, G., Laforteza, R., Sanesi, G., 2020. Likelihood and frequency of recurrent fire  
619 ignitions in highly urbanised Mediterranean landscapes. *Int. J. Wildland Fire.*  
620 <https://doi.org/10.1071/WF19070>  
621 Elia, M., Laforteza, R., Colangelo, G., Sanesi, G., 2014. A streamlined approach for the spatial allocation of  
622 fuel removals in wildland–urban interfaces. *Landsc. Ecol.* 29, 1771–1784.  
623 <https://doi.org/10.1007/s10980-014-0070-7>  
624 Elia, M., Lovreglio, R., Ranieri, N., Sanesi, G., Laforteza, R., 2016. Cost-Effectiveness of Fuel Removals in  
625 Mediterranean Wildland-Urban Interfaces Threatened by Wildfires. *Forests* 7, 149.  
626 <https://doi.org/10.3390/f7070149>  
627 Espinosa, J., Palheiro, P., Loureiro, C., Ascoli, D., Esposito, A., Fernandes, P.M., 2018. Fire-severity mitigation  
628 by prescribed burning assessed from fire-treatment encounters in maritime pine stands. *Can. J. For.*  
629 *Res.* 49, 205–211. <https://doi.org/10.1139/cjfr-2018-0263>  
630 Faivre, N., Jin, Y., Goulden, M.L., Randerson, J.T., 2014. Controls on the spatial pattern of wildfire ignitions  
631 in Southern California. *Int. J. Wildland Fire* 23, 799. <https://doi.org/10.1071/WF13136>  
632 Fox, D.M., Martin, N., Carrega, P., Andrieu, J., Adnès, C., Emsellem, K., Ganga, O., Moebius, F., Tortorollo,  
633 N., Fox, E.A., 2015. Increases in fire risk due to warmer summer temperatures and wildland urban  
634 interface changes do not necessarily lead to more fires. *Appl. Geogr.* 56, 1–12.  
635 <https://doi.org/10.1016/j.apgeog.2014.10.001>  
636 Gallego, F.J., 2010. A population density grid of the European Union. *Popul. Environ.* 31, 460–473.  
637 <https://doi.org/10.1007/s11111-010-0108-y>  
638 Gómez, I., Martín, M.P., 2011. Prototyping an artificial neural network for burned area mapping on a  
639 regional scale in Mediterranean areas using MODIS images. *Int. J. Appl. Earth Obs. Geoinformation*  
640 13, 741–752. <https://doi.org/10.1016/j.jag.2011.05.002>  
641 Gralewicz, N.J., Nelson, T.A., Wulder, M.A., 2012. Factors influencing national scale wildfire susceptibility in  
642 Canada. *For. Ecol. Manag.* 265, 20–29. <https://doi.org/10.1016/j.foreco.2011.10.031>  
643 Guo, F., Su, Z., Wang, G., Sun, L., Lin, F., Liu, A., 2016. Wildfire ignition in the forests of southeast China:  
644 Identifying drivers and spatial distribution to predict wildfire likelihood. *Appl. Geogr.* 66, 12–21.  
645 <https://doi.org/10.1016/j.apgeog.2015.11.014>  
646 Guo, F., Su, Z., Wang, G., Sun, L., Tigabu, M., Yang, X., Hu, H., 2017. Understanding fire drivers and relative  
647 impacts in different Chinese forest ecosystems. *Sci. Total Environ.* 605–606, 411–425.  
648 <https://doi.org/10.1016/j.scitotenv.2017.06.219>  
649 Hamadeh, N., Hilal, A., Daya, B., Chauvet, P., 2015. Studying the factors affecting the risk of forest fire  
650 occurrence and applying neural networks for prediction, in: 2015 SAI Intelligent Systems  
651 Conference (IntelliSys). Presented at the 2015 SAI Intelligent Systems Conference (IntelliSys), pp.  
652 522–526. <https://doi.org/10.1109/IntelliSys.2015.7361189>

653 Hamadeh, N., Karouni, A., Daya, B., Chauvet, P., 2017. Using correlative data analysis to develop weather  
654 index that estimates the risk of forest fires in Lebanon & Mediterranean: Assessment versus  
655 prevalent meteorological indices. *Case Stud. Fire Saf.* 7, 8–22.  
656 <https://doi.org/10.1016/j.csfs.2016.12.001>

657 Hardy, C.C., 2005. Wildland fire hazard and risk: Problems, definitions, and context. *For. Ecol. Manag.*,  
658 Relative Risk Assessments for Decision –Making Related To Uncharacteristic Wildfire 211, 73–82.  
659 <https://doi.org/10.1016/j.foreco.2005.01.029>

660 Hossain, F.M.A., Zhang, Y., Yuan, C., Su, C.-Y., 2019. Wildfire Flame and Smoke Detection Using Static Image  
661 Features and Artificial Neural Network, in: 2019 1st International Conference on Industrial Artificial  
662 Intelligence (IAI). Presented at the 2019 1st International Conference on Industrial Artificial  
663 Intelligence (IAI), pp. 1–6. <https://doi.org/10.1109/ICIAI.2019.8850811>

664 Jaafari, A., Gholami, D.M., Zenner, E.K., 2017. A Bayesian modeling of wildfire probability in the Zagros  
665 Mountains, Iran. *Ecol. Inform.* 39, 32–44. <https://doi.org/10.1016/j.ecoinf.2017.03.003>

666 Jaafari, A., Razavi Termeh, S.V., Bui, D.T., 2019. Genetic and firefly metaheuristic algorithms for an  
667 optimized neuro-fuzzy prediction modeling of wildfire probability. *J. Environ. Manage.* 243, 358–  
668 369. <https://doi.org/10.1016/j.jenvman.2019.04.117>

669 Jaafari, A., Zenner, E.K., Pham, B.T., 2018. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A  
670 comparative study of decision tree based classifiers. *Ecol. Inform.* 43, 200–211.  
671 <https://doi.org/10.1016/j.ecoinf.2017.12.006>

672 Jafari Goldarag, Y., Mohammadzadeh, A., Ardakani, A.S., 2016. Fire Risk Assessment Using Neural Network  
673 and Logistic Regression. *J. Indian Soc. Remote Sens.* 44, 885–894. [https://doi.org/10.1007/s12524-](https://doi.org/10.1007/s12524-016-0557-6)  
674 [016-0557-6](https://doi.org/10.1007/s12524-016-0557-6)

675 Jain, P., Coogan, S.C.P., Subramanian, S.G., Crowley, M., Taylor, S., Flannigan, M.D., 2020. A review of  
676 machine learning applications in wildfire science and management. *ArXiv200300646 Cs Stat.*

677 Jiménez-Valverde, A., 2012. Insights into the area under the receiver operating characteristic curve (AUC)  
678 as a discrimination measure in species distribution modelling. *Glob. Ecol. Biogeogr.* 21, 498–507.  
679 <https://doi.org/10.1111/j.1466-8238.2011.00683.x>

680 Johnstone, J.F., Allen, C.D., Franklin, J.F., Frelich, L.E., Harvey, B.J., Higuera, P.E., Mack, M.C., Meentemeyer,  
681 R.K., Metz, M.R., Perry, G.L., Schoennagel, T., Turner, M.G., 2016. Changing disturbance regimes,  
682 ecological memory, and forest resilience. *Front. Ecol. Environ.* 14, 369–378.  
683 <https://doi.org/10.1002/fee.1311>

684 Laforteza, R., Tanentzap, A.J., Elia, M., John, R., Sanesi, G., Chen, J., 2015. Prioritizing fuel management in  
685 urban interfaces threatened by wildfires. *Ecol. Indic.* 48, 342–347.  
686 <https://doi.org/10.1016/j.ecolind.2014.08.034>

687 Lall, S., Mathibela, B., 2016. The application of artificial neural networks for wildfire risk prediction, in: 2016  
688 International Conference on Robotics and Automation for Humanitarian Applications (RAHA).  
689 Presented at the 2016 International Conference on Robotics and Automation for Humanitarian  
690 Applications (RAHA), pp. 1–6. <https://doi.org/10.1109/RAHA.2016.7931880>

691 Lein, J.K., Stump, N.I., 2009. Assessing wildfire potential within the wildland–urban interface: A  
692 southeastern Ohio example. *Appl. Geogr.* 29, 21–34. <https://doi.org/10.1016/j.apgeog.2008.06.002>

693 Li, L.-M., Song, W.-G., Ma, J., Satoh, K., 2009. Artificial neural network approach for modeling the impact of  
694 population density and weather parameters on forest fire risk. *Int. J. Wildland Fire* 18, 640–647.  
695 <https://doi.org/10.1071/WF07136>

696 Liu, X.-Y., Wu, J., Zhou, Z.-H., 2009. Exploratory Undersampling for Class-Imbalance Learning. *IEEE Trans.*  
697 *Syst. Man Cybern. Part B Cybern.* 39, 539–550. <https://doi.org/10.1109/TSMCB.2008.2007853>

698 Lozano, O.M., Salis, M., Ager, A.A., Arca, B., Alcasena, F.J., Monteiro, A.T., Finney, M.A., Del Giudice, L.,  
699 Scoccimarro, E., Spano, D., 2017. Assessing Climate Change Impacts on Wildfire Exposure in  
700 Mediterranean Areas: Climate Change Impacts on Mediterranean Wildfires. *Risk Anal.* 37, 1898–  
701 1916. <https://doi.org/10.1111/risa.12739>

702 Maeda, E.E., Formaggio, A.R., Shimabukuro, Y.E., Arcoverde, G.F.B., Hansen, M.C., 2009. Predicting forest  
703 fire in the Brazilian Amazon using MODIS imagery and artificial neural networks. *Int. J. Appl. Earth*  
704 *Obs. Geoinformation* 11, 265–272. <https://doi.org/10.1016/j.jag.2009.03.003>

705 Maingi, J.K., Henry, M.C., 2007. Factors influencing wildfire occurrence and distribution in eastern  
706 Kentucky, USA. *Int. J. Wildland Fire* 16, 23. <https://doi.org/10.1071/WF06007>

707 Mancini, L.D., Corona, P., Salvati, L., 2018a. Ranking the importance of Wildfires' human drivers through a  
708 multi-model regression approach. *Environ. Impact Assess. Rev.* 72, 177–186.  
709 <https://doi.org/10.1016/j.eiar.2018.06.003>

710 Mancini, L.D., Elia, M., Barbati, A., Salvati, L., Corona, P., Laforzezza, R., Sanesi, G., 2018b. Are Wildfires  
711 Knocking on the Built-Up Areas Door? *Forests* 9, 234. <https://doi.org/10.3390/f9050234>

712 Menardi, G., Torelli, N., 2014. Training and assessing classification rules with imbalanced data. *Data Min.*  
713 *Knowl. Discov.* 28, 92–122. <https://doi.org/10.1007/s10618-012-0295-5>

714 Michetti, M., Pinar, M., 2019. Forest Fires Across Italian Regions and Implications for Climate Change: A  
715 Panel Data Analysis. *Environ. Resour. Econ.* 72, 207–246. [https://doi.org/10.1007/s10640-018-](https://doi.org/10.1007/s10640-018-0279-z)  
716 [0279-z](https://doi.org/10.1007/s10640-018-0279-z)

717 Modugno, S., Balzter, H., Cole, B., Borrelli, P., 2016. Mapping regional patterns of large forest fires in  
718 Wildland–Urban Interface areas in Europe. *J. Environ. Manage.* 172, 112–126.  
719 <https://doi.org/10.1016/j.jenvman.2016.02.013>

720 Molina, J.R., Herrera, M.A., Rodríguez y Silva, F., 2019. Wildfire-induced reduction in the carbon storage of  
721 Mediterranean ecosystems: An application to brush and forest fires impacts assessment. *Environ.*  
722 *Impact Assess. Rev.* 76, 88–97. <https://doi.org/10.1016/j.eiar.2019.02.008>

723 Moreira, F., Ascoli, D., Safford, H., Adams, M.A., Moreno, J.M., Pereira, J.M.C., Catry, F.X., Armesto, J., Bond,  
724 W., González, M.E., Curt, T., Koutsias, N., McCaw, L., Price, O., Pausas, J.G., Rigolot, E., Stephens, S.,  
725 Tavsanoğlu, C., Vallejo, V.R., Wilgen, B.W.V., Xanthopoulos, G., Fernandes, P.M., 2020. Wildfire  
726 management in Mediterranean-type regions: paradigm change needed. *Environ. Res. Lett.* 15,  
727 011001. <https://doi.org/10.1088/1748-9326/ab541e>

728 Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., Barbati, A., Corona, P., Vaz, P.,  
729 Xanthopoulos, G., Mouillot, F., Bilgili, E., 2011. Landscape – wildfire interactions in southern  
730 Europe: Implications for landscape management. *J. Environ. Manage.* 92, 2389–2402.  
731 <https://doi.org/10.1016/j.jenvman.2011.06.028>

732 Nunes, M.C.S., Vasconcelos, M.J., Pereira, J.M.C., Dasgupta, N., Alldredge, R.J., Rego, F.C., 2005. Land Cover  
733 Type and Fire in Portugal: Do Fires Burn Land Cover Selectively? *Landsc. Ecol.* 20, 661–673.  
734 <https://doi.org/10.1007/s10980-005-0070-8>

735 Oliveira, S., Oehler, F., San-Miguel-Ayán, J., Camia, A., Pereira, J.M.C., 2012. Modeling spatial patterns of  
736 fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For. Ecol.*  
737 *Manag.* 275, 117–129. <https://doi.org/10.1016/j.foreco.2012.03.003>

738 Pai, M.L., Varsha, K.S., Arya, R., 2020. Application of Artificial Neural Networks and Genetic Algorithm for  
739 the Prediction of Forest Fire Danger in Kerala, in: Abraham, A., Cherukuri, A.K., Melin, P., Gandhi, N.  
740 (Eds.), *Intelligent Systems Design and Applications, Advances in Intelligent Systems and Computing.*  
741 Springer International Publishing, Cham, pp. 935–942. [https://doi.org/10.1007/978-3-030-16660-](https://doi.org/10.1007/978-3-030-16660-1_91)  
742 [1\\_91](https://doi.org/10.1007/978-3-030-16660-1_91)

743 Paveglio, T.B., Edgeley, C.M., Stasiewicz, A.M., 2018. Assessing influences on social vulnerability to wildfire  
744 using surveys, spatial data and wildfire simulations. *J. Environ. Manage.* 213, 425–439.  
745 <https://doi.org/10.1016/j.jenvman.2018.02.068>

746 Polinova, M., Wittenberg, L., Kutiel, H., Brook, A., 2019. Reconstructing pre-fire vegetation condition in the  
747 wildland urban interface (WUI) using artificial neural network. *J. Environ. Manage.* 238, 224–234.  
748 <https://doi.org/10.1016/j.jenvman.2019.02.091>

749 Prasad, V.K., Badarinath, K.V.S., Eaturu, A., 2008. Biophysical and anthropogenic controls of forest fires in  
750 the Deccan Plateau, India. *J. Environ. Manage.* 86, 1–13.  
751 <https://doi.org/10.1016/j.jenvman.2006.11.017>

752 Preisler, H., Brillinger, D.R., Burgan, R.E., Benoit, J., 2004. Probability based models for estimation of wildfire  
753 risk. *Int. J. Wildland Fire* 132 133-142 13, 133–142. <https://doi.org/10.1071/WF02061>

754 Preisler, H.K., Brillinger, D.R., Burgan, R.E., Benoit, J.W., 2004. Probability based models for estimation of  
755 wildfire risk\*. *Int. J. Wildland Fire* 13, 133–142. <https://doi.org/10.1071/WF02061>

756 Pricope, N.G., Binford, M.W., 2012. A spatio-temporal analysis of fire recurrence and extent for semi-arid  
757 savanna ecosystems in southern Africa using moderate-resolution satellite imagery. *J. Environ.*  
758 *Manage.* 100, 72–85. <https://doi.org/10.1016/j.jenvman.2012.01.024>  
759 RAF, R.R., 2019. Rapporto sullo stato delle foreste e del settore forestale in Italia - RaF [WWW Document].  
760 <https://www.reterurale.it>. URL  
761 <https://www.reterurale.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/19231> (accessed 7.29.20).  
762 Reuter, H.I., Nelson, A., Jarvis, A., 2007. An evaluation of void-filling interpolation methods for SRTM data.  
763 *Int. J. Geogr. Inf. Sci.* 21, 983–1008. <https://doi.org/10.1080/13658810601169899>  
764 Ricotta, C., Di Vito, S., 2014. Modeling the Landscape Drivers of Fire Recurrence in Sardinia (Italy). *Environ.*  
765 *Manage.* 53, 1077–1084. <https://doi.org/10.1007/s00267-014-0269-z>  
766 Rodrigues, M., de la Riva, J., 2014. An insight into machine-learning algorithms to model human-caused  
767 wildfire occurrence. *Environ. Model. Softw.* 57, 192–201.  
768 <https://doi.org/10.1016/j.envsoft.2014.03.003>  
769 Rodrigues, M., Jiménez, A., de la Riva, J., 2016. Analysis of recent spatial–temporal evolution of human  
770 driving factors of wildfires in Spain. *Nat. Hazards* 84, 2049–2070. [https://doi.org/10.1007/s11069-](https://doi.org/10.1007/s11069-016-2533-4)  
771 [016-2533-4](https://doi.org/10.1007/s11069-016-2533-4)  
772 Ruiz-Mirazo, J., Martínez-Fernández, J., Vega-García, C., 2012. Pastoral wildfires in the Mediterranean:  
773 Understanding their linkages to land cover patterns in managed landscapes. *J. Environ. Manage.* 98,  
774 43–50. <https://doi.org/10.1016/j.jenvman.2011.12.017>  
775 Safi, Y., Bouroumi, A., 2013. Prediction of forest fires using Artificial neural networks. *Appl. Math. Sci.* 7,  
776 271–286. <https://doi.org/10.12988/ams.2013.13025>  
777 Sakr, G.E., Elhajj, I.H., Mitri, G., 2011. Efficient forest fire occurrence prediction for developing countries  
778 using two weather parameters. *Eng. Appl. Artif. Intell.* 24, 888–894.  
779 <https://doi.org/10.1016/j.engappai.2011.02.017>  
780 San-Miguel-Ayanz, J., Moreno, J.M., Camia, A., 2013. Analysis of large fires in European Mediterranean  
781 landscapes: Lessons learned and perspectives. *For. Ecol. Manag.* 294, 11–22.  
782 <https://doi.org/10.1016/j.foreco.2012.10.050>  
783 Satir, O., Berberoglu, S., Donmez, C., 2016. Mapping regional forest fire probability using artificial neural  
784 network model in a Mediterranean forest ecosystem. *Geomat. Nat. Hazards Risk* 7, 1645–1658.  
785 <https://doi.org/10.1080/19475705.2015.1084541>  
786 Sebastián-López, A., Salvador-Civil, R., Gonzalo-Jiménez, J., SanMiguel-Ayanz, J., 2008. Integration of socio-  
787 economic and environmental variables for modelling long-term fire danger in Southern Europe.  
788 *Eur. J. For. Res.* 127, 149–163. <https://doi.org/10.1007/s10342-007-0191-5>  
789 Sharples, J.J., McRae, R.H.D., Weber, R.O., Gill, A.M., 2009. A simple index for assessing fire danger rating.  
790 *Environ. Model. Softw.* 24, 764–774. <https://doi.org/10.1016/j.envsoft.2008.11.004>  
791 Syphard, A.D., Radeloff, V.C., Keuler, N.S., Taylor, R.S., Hawbaker, T.J., Stewart, S.I., Clayton, M.K., 2008.  
792 Predicting spatial patterns of fire on a southern California landscape.  
793 <https://doi.org/10.1071/WF07087>  
794 Twidwell, D., Wonkka, C.L., Wang, H.-H., Grant, W.E., Allen, C.R., Fuhlendorf, S.D., Garmestani, A.S.,  
795 Angeler, D.G., Taylor, C.A., Kreuter, U.P., Rogers, W.E., 2019. Coerced resilience in fire  
796 management. *J. Environ. Manage.* 240, 368–373. <https://doi.org/10.1016/j.jenvman.2019.02.073>  
797 Union, P.O. of the E., 2018. Forest fires in Europe, Middle East and North Africa 2017. [WWW Document].  
798 URL [https://op.europa.eu/443/en/publication-detail/-/publication/77307262-ba25-11e8-99ee-](https://op.europa.eu/443/en/publication-detail/-/publication/77307262-ba25-11e8-99ee-01aa75ed71a1/language-en)  
799 [01aa75ed71a1/language-en](https://op.europa.eu/443/en/publication-detail/-/publication/77307262-ba25-11e8-99ee-01aa75ed71a1/language-en) (accessed 2.3.20).  
800 Valesse, E., Conedera, M., Held, A.C., Ascoli, D., 2014. Fire, humans and landscape in the European Alpine  
801 region during the Holocene. *Anthropocene, Landscapes in the Anthropocene* 6, 63–74.  
802 <https://doi.org/10.1016/j.ancene.2014.06.006>  
803 Vasilakos, C., Kalabokidis, K., Hatzopoulos, J., Matsinos, I., 2009. Identifying wildland fire ignition factors  
804 through sensitivity analysis of a neural network. *Nat. Hazards* 50, 125–143.  
805 <https://doi.org/10.1007/s11069-008-9326-3>  
806 Vega Garcia, C., Lee, B.S., Woodard, P.M., Titus, S.J., 1996. Applying neural network technology to human-  
807 caused wildfire occurrence prediction. *AI Appl.* 10, 9–18.

808 Vilar del Hoyo, L., Martín Isabel, M.P., Martínez Vega, F.J., 2011. Logistic regression models for human-  
809 caused wildfire risk estimation: analysing the effect of the spatial accuracy in fire occurrence data.  
810 Eur. J. For. Res. 130, 983–996. <https://doi.org/10.1007/s10342-011-0488-2>  
811 Vilar, L., Garrido, J., Echavarría, P., Martínez-Vega, J., Martín, M.P., 2019. Comparative analysis of CORINE  
812 and climate change initiative land cover maps in Europe: Implications for wildfire occurrence  
813 estimation at regional and local scales. Int. J. Appl. Earth Obs. Geoinformation 78, 102–117.  
814 <https://doi.org/10.1016/j.jag.2019.01.019>  
815 Yang, L., Dawson, C.W., Brown, M.R., Gell, M., 2006. Neural network and GA approaches for dwelling fire  
816 occurrence prediction. Knowl.-Based Syst. 19, 213–219.  
817 <https://doi.org/10.1016/j.knosys.2005.11.021>  
818

819

## 820 Supplementary material

821

822 **Table S1** - Performance metrics of the logistic function in five random subsamples of training and  
823 test sets for each study areas

	AUC		Accuracy <sup>824</sup>	
	ASR	IPR	ASR	IPR <sup>825</sup>
1 sample	0.78	0.65	0.71	0.62
2 sample	0.78	0.66	0.70	0.63 <sup>826</sup>
3 sample	0.79	0.66	0.70	0.63 <sup>827</sup>
4 sample	0.77	0.65	0.71	0.62
5 sample	0.79	0.65	0.71	0.63 <sup>828</sup>

829

830 **Table S2** – Logistic model results explaining the probability of wildfire occurrence in Alpine and  
831 Subalpine Region (ASR) and Insular and Peninsular Region (IPR).

Explanatory variable	ASR			IPR		
	Coefficients	Std. Error	P-value	Coefficients	Std. Error	P-value
(Intercept)	-0.602	0.038	***	-0.101	0.009	***
DTM	-1.135	0.089	***	0.33	0.015	***
Slope	0.288	0.050	***	0.446	0.013	***
Tree cover	0.437	0.043	***	0.154	0.012	***
FCI	0.189	0.028	***	0.218	0.010	***
Agric	0.004	0.328		0.346	0.055	***
Forest	0.714	0.296	*	0.396	0.049	***
Grassland	0.713	0.184	***	0.373	0.027	***
Otherland	0.172	0.206		0.067	0.013	***
Shrubland	0.160	0.072	*	0.147	0.025	***
Urban	0.007	0.144		0.042	0.024	
Water	0.007	0.062		-0.001	0.014	
Wet	0.091	0.035	*	0.026	0.012	*
Pop density	-0.044	0.045		-0.072	0.017	***
Rail distance	0.056	0.040		-0.125	0.011	***
Road distance	-0.029	0.044		-0.041	0.01	***

**Settlements distance** -0.533      0.057      \*\*\*      -0.249      0.011      \*\*\*

832 Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

833

834 **Table S3** – Intended definition of the fire terminology adopted in the study

<b>Term</b>	<b>Definition</b>	<b>Reference</b>
Fire probability	The probability of a wildfire occurring at a given point or area	Preisler et al., 2004
Fire risk	The product of the probability of an event and the expected outcome - typically expressed as damage - of the event	Hardy, 2005
Vulnerability	Potential effects of fire on social and ecological values and their sensitivity and resilience fire disturbance	Chuvieco et al., 2014
Fire Danger	The resultant descriptor of the combination of both constant and variable factors which affect the initiation, spread and difficulty of control of wildfires on an area	Deeming, 1972

835

836

837