







# Wildfire hazard mapping in the eastern Mediterranean landscape

Andrea Trucchia<sup>A,\*</sup> , Giorgio Meschi<sup>A</sup> , Paolo Fiorucci<sup>A</sup> , Antonello Provenzale<sup>B</sup> , Marj Tonini<sup>C</sup>  and Umberto Pernice<sup>A,D</sup> 

For full list of author affiliations and declarations see end of paper

**\*Correspondence to:**

Andrea Trucchia  
CIMA Research Foundation, I-17100  
Savona, Italy  
Email: [andrea.trucchia@cimafoundation.org](mailto:andrea.trucchia@cimafoundation.org)

## ABSTRACT

**Background.** Wildfires are a growing threat to many ecosystems, bringing devastation to human safety and health, infrastructure, the environment and wildlife. **Aims.** A thorough understanding of the characteristics determining the susceptibility of an area to wildfires is crucial to prevention and management activities. The work focused on a case study of 13 countries in the eastern Mediterranean and southern Black Sea basins. **Methods.** A data-driven approach was implemented where a decade of past wildfires was linked to geoclimatic and anthropic descriptors via a machine learning classification technique (Random Forest). Empirical classification of fuel allowed linking of fire intensity and hazard to environmental drivers. **Key results.** Wildfire susceptibility, intensity and hazard were obtained for the study area. For the first time, the methodology is applied at a supranational scale characterised by a diverse climate and vegetation landscape, relying on open data. **Conclusions.** This approach successfully allowed identification of the main wildfire drivers and led to identification of areas that are more susceptible to impactful wildfire events. **Implications.** This work demonstrated the feasibility of the proposed framework and settled the basis for its scalability at a supranational level.

**Keywords:** crossboundary wildfire management, eastern Mediterranean, hazard mapping, machine learning, Random Forest, risk management, susceptibility mapping, wildfire drivers.

## Introduction

Wildfires are growing in intensity and spreading in range across many ecosystems, causing devastation to the environment, wildlife, human health and infrastructure (United Nations Environment Programme 2022). However, the complexity of the processes controlling wildfire occurrence is high, and in other regions and ecosystems, the data indicate a decrease of the area burned by fires. For example, after the large impacts of the 1980s wildfire seasons, in most southern European countries, the average trend in wildfire distribution showed a decrease in the total annual burned area and number of events (Turco *et al.* 2016). The latter situation is certainly related to increasing firefighting capacities and awareness thanks also to improved wildfire forecasting. However, in this same area, the impacts of climate change, coupled with the drastic modifications in land use and socio-economic conditions that have occurred in the last decade are expected to trigger a significant increase in the frequency, extent and severity of wildfires if further improved prevention and control measures are not put in place (Turco *et al.* 2018).

Severe wildfires characterised the 2021 wildfire summer season, during which Greece, Italy, Algeria and Turkey experienced a large number of severe wildfire events burning more than 630 000 ha (San-Miguel-Ayanz *et al.* 2022). The 2021 Algerian wildfires killed at least 90 people (ReliefWeb 2021), resulting in the deadliest fires of recent times after the events in central Portugal (2017) and in Mathi, Greece (2018). Thus, the impacts of the recent and recurrent wildfires in the Mediterranean area indicate that human societies are facing an increasing fire risk, owing to the combination of climate conditions and landscape-scale fuel accumulation caused by the abandonment of rural activities (Ascoli *et al.* 2021).

**Received:** 7 July 2022

**Accepted:** 10 February 2023

**Published:** 16 March 2023

**Cite this:**

Trucchia A *et al.* (2023)  
*International Journal of Wildland Fire*  
32(3), 417–434. doi:10.1071/WF22138

© 2023 The Author(s) (or their employer(s)). Published by CSIRO Publishing on behalf of IAWF. This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND)

OPEN ACCESS

In order to reduce such impacts, there is an urgent need to shift from fire-management systems mainly focused on wildfire suppression to early warning systems that include structured prevention and risk mitigation strategies. In this framework, a comprehensive high-resolution geographical mapping of at least the static wildfire hazard and risk level is a first priority towards developing large-scale, integrated approaches to such strategies.

An obstacle to this shift is the absence of a collective and widespread understanding of the conditions related to Extreme Wildfire Events (EWEs) (Tedim et al. 2013, 2018) and other classes of impactful wildfires beyond the sole cause of ignition or the effect of weather, which are the main uncontrollable aspects of the phenomenon. Blaming only ignition causes and weather effects neglects the fact that there is still much room for knowledge improvements leading to the identification of priorities for effective wildfire prevention. EWEs in particular are a complex phenomenon that springs from the interplay between natural and social conditions acting in all phases of wildfire activity (Tedim et al. 2018). The first step to tackle this problem requires reaching a comprehensive understanding of the general wildfire regime, leading to a definition of priorities, a rise in prevention and preparedness level and more effective land use planning, with appropriate restoration and risk mitigation activities.

A powerful tool to deepen the understanding of these phenomena is provided by wildfire susceptibility assessments, which also allow hazard and risk maps to be obtained (Pradhan et al. 2007; Ghorbanzadeh et al. 2019a). Hazard maps provide information on the likelihood of wildfire occurrence, taking into account knowledge about fire behaviour. Risk maps include the characterisation of the exposed elements in terms of their value and vulnerability, and ultimately lead to identification of the areas where wildfires can have the most severe impacts.

To foster prevention activities, a thorough understanding of the predisposing environmental factors for severe wildfires is crucial in civil protection and fire management activities. From a technological point of view, there are plenty of data, tools and models that can be applied to this issue. Therefore, advanced hazard and risk mapping methodologies harmonised at different scales are necessary.

Synoptic time series of mapped burned areas provide a significant input to learn from the past, allowing us to better understand wildfire regimes and to identify the main drivers of catastrophic wildfires. However, often time series are not long enough and locally accurate to reach these goals. To circumvent this problem, a larger and biogeographically diverse area can be considered to infer wildfire drivers in different climates, topographic and anthropogenic conditions. In this context, wildfire susceptibility, hazard and risk maps can at least partially compensate for the lack of accurate synoptic time series of burned areas at local scale, helping decision makers and practitioners in wildfire management and long-term landscape management, and strengthening efficient prevention

activities adapted to local environmental and socio-economic contexts. The objective of such maps may range from static assessments (hazard) to dynamic ones (danger).

In this work, we describe in detail a possible methodology for wildfire hazard assessment at supranational level, along the lines of a research framework started at local level for the Liguria Region in Italy (Tonini et al. 2020) and recently expanded at national scale (Trucchia et al. 2022). Such methodology is based on the drafting of a contingency hazard matrix aiming at coupling the information related to wildfire susceptibility and potential fireline intensity. The susceptibility, describing the static spatial probability of wildfire occurrence, was estimated in a machine learning framework to determine the most fire-prone areas based on the geo-topographical characteristics of the territory and average climatic data. The potential intensity and fire behaviour, a proxy for the severity of wildfire in case of its occurrence, were modelled as a function of land cover.

Hazard mapping, as well as its main components, is provided with spatial resolution of 500 m over the whole study area, which includes different countries in the eastern European Mediterranean landscape. The availability of different static maps makes it possible to assess different types of information related to wildfires: to begin with, susceptibility mapping addresses in a static way the areas having high probability of being affected by a wildfire event. The intensity mapping in turn estimates the areas where wildfires can be characterised by extreme fire behaviour. Finally, hazard mapping highlights the areas characterised by high probability of intense wildfires. The information provided is a valid tool for supporting national civil protection agencies in the early phases of the disaster risk management cycle and makes it possible to better address cross-border wildfire management issues. An important objective of this research paper is to assess the performance of the susceptibility methodology in a full-scale case study and gather relevant information on wildfire drivers in the eastern Mediterranean area. Another objective is to test the feasibility of an expeditive, fast-track approach to hazard mapping that relies on open data, easy to gather and interpret. To the best of our knowledge, this study represents one of the first attempts to model wildfire hazard at the supranational scale in the eastern Mediterranean, where useful information can be drawn with a view to cross-border wildfire management, supporting the European Civil Protection Mechanism. This analysis also provides the starting point for wildfire risk assessment, when the exposure elements in a given area are included.

## A framework for wildfire hazard assessment and mapping

The proposed approach combines multi-source data gathering, model/expert-based processes and machine learning (ML) analyses.

This preliminary work, undertaken for a large area including several countries in the Eastern Mediterranean and southern Black Sea region, uses available global data sets, considering only open data. The main steps, each one producing maps as an output, are summarised below:

1. Wildfire susceptibility is defined as the static probability of wildfires in a certain area, depending on the intrinsic characteristics of the terrain. This can be achieved by adopting several approaches, ranging from statistical hierarchical ones to ML-based algorithms.
2. Wildfire intensity is defined as the rate of heat energy released by the fire. It is linked to flame length and more in general to fire behaviour during a specific wildfire event. At this stage, the areas where severe wildfires can develop owing to fuel cover and other environmental features are detected. This can be done by expert-based classification of fuel cover or by empirical models.
3. Wildfire hazard is indicated by the spatial distribution of the areas where a severe wildfire is likely to occur. This can be done by merging the outputs of the two previous steps by means of empirical functions or via contingency matrices.

The hazard maps obtained can then be paired with data on available highly valued resources and assets and their vulnerability to potential wildfire events in order to obtain risk maps. Such maps are usually tailored around the choice of assets to be safeguarded (infrastructural, social, ecosystemic, ...) and usually employ multi-criteria decision-making techniques and multi-objective optimisation procedures.

In the following, the terms regarding wildfire hazard and risk will be disambiguated. Extensive work has been done in the last two decades on this subject, and despite the many efforts that the community put in place, there are still discrepancies between the different approaches. In particular, a considerable range of definitions for wildfire hazard still exists, and the metrics used to express the terms are varied. According to Bachmann and Allgower, 'inconsiderate use of the terms 'danger', 'hazard' and 'risk' may ultimately result in misunderstandings that can have fatal consequences' (Bachmann and Allgower 2000, p. 67)

The approach proposed here has the objective of obtaining a clear operational characterisation of hazard, keeping at minimum the terminological differences with other established frameworks. Similarities and divergences between the proposed approach and some of those present in the literature are discussed below.

Hardy (2005) proposes the same division between hazard and risk, with the latter encompassing the exposed assets (highly valued resources and assets, HVRAs). His framework also states that static hazard mapping should be independent of weather assessment, which is instead considered in danger profiling and early warning systems. In this framework, wildfire hazard is 'a fuel complex, defined by volume, type, condition, arrangement, and location that determines

the degree of ease of ignition and the resistance to control. Fire hazard expresses the potential fire behaviour for a fuel type, regardless of the fuel type's weather-influenced fuel moisture content.' This characterisation was the synthesis of the NWCG (2003) and Canadian (Ministry of Forests (MOF) 1997, Ministry of Forests and Range (MOF) 2008) definitions of fire hazard. This definition clearly associates the hazard with the fuel that paves the way for specific fire behaviour, which in the framework proposed in the present paper corresponds only to the severity/fuel behaviour layer. However, the terms 'arrangement and location' can fit into the proposed framework as they are dealt with in the computation of the susceptibility layer.

Scott, Thompson and Calkin (USDA), in their exhaustive general technical report of 2013 (Scott *et al.* 2013) expressed a framework that is widely compatible with the one proposed here. They adopted the words 'likelihood' and 'susceptibility' for wildfire susceptibility and the exposed elements' vulnerability to fire, respectively. In particular, they expressed wildfire hazard as the contribution of both wildfire likelihood and potential intensity / fire behaviour, expressly stating that fire intensity level is independent of likelihood. In the framework of Scott and co-authors, likelihood and intensity are to be considered as integrated measures of hazard that fit with the concept of hazard as a situation with potential for damage.

Parente and Pereira (2016) expressed a risk framework for a case study of Portugal, expanding the seminal ideas of Verde and Zêzere (2010). Also in their case, risk is obtained by integrating the static hazard map with HRVAs and related vulnerability. Fire hazard is considered by blending the pixel-by-pixel fire probability, which accounts e.g. for human pressure in determining ignition, with a spatially distributed 'wildfire favourability score' that depends on land use – vegetation cover and topography. Such frameworks share with ours the fact that susceptibility is not only linked by fire ignition probability but also by the favourable conditions for its spread, which enhance the possibility of fire occurring in a given location. However, the framework of Parente and Pereira is intrinsically dependent on the probabilistic rating of different scenarios, assigning a probability to different scenarios of total annual burned area.

Finally, recent works by Ghorbanzadeh and co-authors (Ghorbanzadeh *et al.* 2019a; Gholamnia *et al.* 2020; Tavakkoli Piralilou *et al.* 2022) considered susceptibility obtained via ML techniques applied to topographic, anthropic, land cover/land use and climatic inputs as a stand-alone proxy for wildfire hazard, to be expanded to a full-fledged risk analysis using social and infrastructural vulnerability indicators.

## Materials and methods

In this section, the data and the algorithms used in the analysis are described in detail.

## Study area

The study area includes the following countries: Italy, Slovenia, Croatia, Western Balkans (Albania, Bosnia and Herzegovina, North Macedonia, Montenegro, Kosovo<sup>1</sup> and Serbia), Greece, Cyprus, Bulgaria and Turkey. Together, they embrace a vast area (~1 612 500 km<sup>2</sup>) characterised by a large number of different biogeographical regions: Mediterranean, continental, Alpine, Anatolian, Pannonian and Black Sea biogeographical regions (European Environment Agency 2002). Following the Köppen–Geiger climate classification (Beck et al. 2018), most of the area is characterised by either ‘hot-summer Mediterranean climate (Csa)’, especially along the coast, ‘semi-arid climate (Bsk)’, especially in internal continental areas, and ‘Mediterranean-influenced warm-summer humid continental climate (Dsa, Dsb, Dfb)’, especially in mountain areas and in the interior regions of eastern Europe and Turkey. Given the high diversity of environments, vegetation typically ranges from Mediterranean maquis to broadleaf woods to coniferous trees especially in the high-elevation areas, but also large portions of grasslands. Most of the area is characterised by heavy anthropic impact on the distribution of vegetation species, mainly for agro-forestry purposes. In the last decades, the abandonment of traditional agricultural and forestry activities has led to reforestation and to possible impacts on biodiversity (Plieninger et al. 2014 and reference therein) but also to potentially positive effects on carbon sequestration (Novara et al. 2017).

## Burned area database

Proper wildfire susceptibility mapping relies on a dataset of past burned area polygons. In this analysis, the European Forest Fire Information System (EFFIS) burned area database was adopted for all countries in the study area. The main dataset provides the burned area recorded from 2008 to 2019. The wildfire polygons were then merged and projected to a binary working raster (which accounts for the presence–absence of wildfire). In this way, information about the burned frequency per pixel is not present in the dataset. This procedure can decrease the impact of pasture agricultural fires in the final analysis (Trucchia et al. 2022).

This database provides the label for the classification procedure to the training set and the test set through random sampling of the pixels (see *Methodology* subsection). The wildfire database encompasses a total of 10 118 burned polygons corresponding to a grand total of 2 922 532 ha burned. As a further validation procedure, another dataset from EFFIS was retrieved, corresponding to the 2020–2021 burned areas, which includes 7204 polygons summing up to 1 100 217 ha burned.

## Climatic database

Several climatic layers were considered. The two main data sources were provided by ERA5 data, i.e. the fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis for global climate and weather. These are: the Climate Change Knowledge Portal (CCKP), that is, the hub for climate-related information, data and tools for the World Bank Group (World Bank Group 2021), and the Copernicus Climate Data Store – CCDS (Hersbach et al. 2018). The climatic layers retrieved from CCKP are the ERA5 layers (retrieved for the time window 1991–2020) corresponding to mean precipitation, maximum number of consecutive dry days, maximum temperature, mean temperature, number of summer days and number of tropical nights. From CCDS, the layer of soil moisture averaged for the years 1991–2020 was retrieved.

The map of the Köppen–Geiger climate classification, available at 1-km resolution for the present-day (1980–2016) was also retrieved (Beck et al. 2018).

The climatic layers are summarised in Table 1.

## Land use/land cover layers

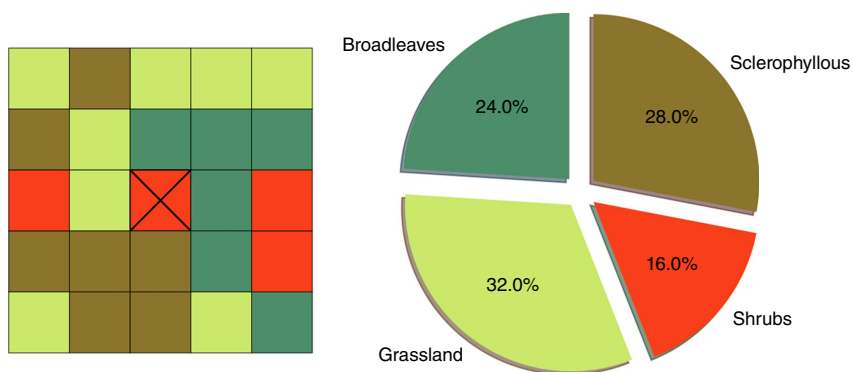
The main data source used for defining vegetation cover variables is CORINE Land Cover 2018 (CLC2018). The raster obtained, containing the CORINE code for the pixels, was processed to obtain the neighbouring vegetation variables, which express vegetation continuity over the analysed landscape. Vegetation continuity was the most important aspect in previous studies conducted in Italy (Trucchia et al. 2022) where local and very detailed fire perimeters were used in the analysis. The results from that study show that the spatial continuity of very flammable species is the main factor responsible for large and uncontrollable fires. However, the continuity of mature and undisturbed broadleaf forests limits the expansion of large wildfires. For this reason, neighbouring percentages of vegetation variables were included in the large-scale analysis. For each pixel, a Moore neighbourhood of order 2 (resulting in the 24 surrounding pixels having a total area of  $2.5 \times 2.5$  km<sup>2</sup>) was estimated (see Fig. 1). The frequency of appearance of the vegetation types was then computed. For instance, in the case where a pixel is completely surrounded by CLC2018 code ‘311’, that is, broadleaf vegetation, the variable ‘perc\_311’ is set to 1 while all the other ‘perc\_XXX’ variables are set to 0.

Aside from CORINE data, the Tree-Cover Density (TCD) layer provided by Copernicus and available from the Copernicus Land Monitoring Service (Langanke et al. 2013) was used to characterise each pixel.

<sup>1</sup>This designation is without prejudice to position on status, and is in line with United Nations Security Council resolution (UNSCR) 1244 and the International Court of Justice (ICJ) Opinion on the Kosovo declaration of independence.

**Table 1.** Description of the climatic layers adopted.

Climatic layer	Resolution (km)	Description	Source
Mean precipitation	~55	Average of yearly accumulated precipitation (mm)	Climate Change Knowledge Portal 1991–2020
Maximum no. of consecutive dry days	~55	Number of days in the longest period without significant precipitation of at least 1 mm (days)	Climate Change Knowledge Portal 1991–2020
Max temperature	~55	Average maximum temperature (°C)	Climate Change Knowledge Portal 1991–2020
Mean temperature	~55	Average mean temperature (°C)	Climate Change Knowledge Portal 1991–2020
Number of summer days	~55	Average count of days where the daily maximum temperature surpassed 25°C (days)	Climate Change Knowledge Portal 1991–2020
Number of tropical nights	~55	Average count of days where the daily minimum temperature remained above 20°C (days)	Climate Change Knowledge Portal 1991–2020
Soil moisture	~25	Volumetric soil water (layer 1), 0–7 cm	ERA5 monthly averaged data on single levels from 1979 to present, from Copernicus Climate Data Store
Köppen–Geiger climate classification	~7.4	Eight climate classes of Köppen–Geiger (–)	Beck <i>et al.</i> (2018)



**Fig. 1.** A representation of the Moore Neighbourhood and the 'neighbourhood percentages' of vegetation classes obtained (vegetation continuity). Any other neighbour vegetation type not represented in the image is set to 0%.

### Anthropic layers

A dataset accounting for anthropic effects in the fire regime was incorporated into the analysis with the following layers:

- (1) 2015 Population Density retrieved from the NASA Socioeconomic Data and Applications Center (SEDAC) product *Gridded Population of the World, Version 4 (GPWv4)*
- (2) Distance from cultivated land, elaborating Corine Land Cover data
- (3) Distance from urban areas, again elaborating Corine Land Cover data.

### Topographic layers

The following layers related to topographic information were incorporated into the analysis:

- (1) The Digital Elevation Model (DEM) from the Multi-Error-Removed Improved-Terrain (MERIT) DEM (Yamazaki *et al.* 2017)

- (2) Slope, obtained from the DEM by the means of Geospatial Data Abstraction Library (GDAL) libraries
- (3) Aspect components, northing and easting (Olaya 2009).

All climatic variables are summarised in Table 1, and all the other variables in Table 2.

## Methodology

### Susceptibility map

The proposed methodology for computation of the wild-fire susceptibility map is based on an ML algorithm (Tonini *et al.* 2020; Trucchia *et al.* 2022) structured as a classification task. It is based on the Random Forest Classifier (RF) (Breiman 2001) used to find a functional relationship between the dependent variable (the label, in this case wild-fire occurrences) and the independent variables (the predisposing environmental factors). RF is a widespread technique when it comes to studying a wide range of natural hazards such as floods (Nachappa *et al.* 2020a), landslides

**Table 2.** Input data for the susceptibility mapping: non-climatic layers for predisposing factors and observed variable (burned areas).

Input data	Source	Description
CORINE land cover	Copernicus <sup>A</sup>	Land cover raster file of Corine 2018 at 100-m resolution
Copernicus tree cover density (TCD)	Copernicus <sup>B</sup>	Density of the forestry areas at European level at 100-m resolution
Digital Elevation Model (DEM)	MERIT DEM (Yamazaki et al. 2017)	Raster file related to the elevation in metres for the study area
Burned areas	EFFIS <sup>C</sup>	Historical burned areas retrieved from EFFIS (data ranging from 2008 to 2019)

<sup>A</sup><https://land.copernicus.eu/pan-european/corine-land-cover>.

<sup>B</sup><https://land.copernicus.eu/pan-european/high-resolution-layers/forests/tree-cover-density>.

<sup>C</sup><https://effis.jrc.ec.europa.eu/>.

**Table 3.** The wildfire susceptibility classes subdivisions.

Susceptibility classes	Percentile ranges	Description
Three-class thresholds		
1	0–25	Low
2	25–75	Medium
3	75–100	High
Five-class thresholds		
1	0–30	Very low
2	30–50	Low
3	50–80	Medium
4	80–95	High
5	95–100	Very high

The three-classes general levels are used to classify the susceptibility map when used as one of the two inputs of the expert judgement-based hazard contingency matrix. The five-classes subdivision is used to assess the performance of the susceptibility map.

(Ghorbanzadeh et al. 2019b) or even multi-hazard analyses (Nachappa et al. 2020b).

The predisposing factors considered are geographic data (elevation, slope, aspect, land cover/fuel cover), climatic data (ERA5 layers, Köppen–Geiger climate classes) and anthropogenic data (population density, distance from settlements and crops), as shown in detail in this section.

All the input layers involved in the creation of the predisposing factor dataset were rasterised and rescaled to a spatial resolution of 500 m, adopting the EPSG:3035 ETRS89-extended/LAEA Europe coordinate reference system. The dataset covering the entire study area was randomly sampled in order to split the data into a training and a testing set. First, a sampled set is generated by merging all national sampled subsets. These latter are built by taking into account 30% of the total number of burned pixels and the same number of pseudo-absence pixels (that is, pixels that have not experienced a wildfire according to the wildfire historical dataset) associated with each country. Finally, the sample dataset obtained in this way is split into training and testing sets by randomly selecting 75 and 25% of the pixels respectively. The resulting dataset summarises the geo-topographical and

**Table 4.** Brief description of the different model configurations with their identification code.

Experiment code	Experiment description
E1	Model configuration with the whole set of predisposing factors
E2	Neighbouring vegetation variables are removed
E3	Climate variables are removed
E4	The unique climate variables kept in the feature dataset are the mean precipitation and mean temperature

climatic conditions of all the different countries and has the property of being balanced, that is, the number of burned pixels is equal to the number of non-burned pixels (pseudo-absences). This allows better training of the ML model, which otherwise would overestimate the most frequent label. Consequently, the RF model was trained on the training dataset and evaluated over the test dataset to compute performance indicators. In this work, we use as performance indicators the area under the receiving operator characteristic (ROC) curve (AUC) and the mean square error (MSE) (Trucchia et al. 2022), which are cited below. In addition, we also tested the models on a validation dataset that referred to a temporal period not included in the data used for training, namely, the burned area information for 2020 and 2021.

The output susceptibility can be visualised in several ways, retaining the raw continuous values or classifying them through a given choice of percentile values (see Table 3). In the first percentile choice, the susceptibility was divided into three classes (from Low to High).

This classification helped to build the susceptibility/wildfire intensity contingency matrix for assessing wildfire hazard.

This choice aims at reducing the complexity of the contingency matrix: the intensity layer is crossed with low, medium or high likelihood of experiencing a wildfire. This subdivision, which uses a low number of classes, is more suitable to identify the wildfire hazard levels than a five-class configuration for the susceptibility, introducing less subjectivity to the computation.

In the second configuration, five classes were defined (from Very Low to Very High susceptibility) once again through specific percentiles.

This configuration is more aimed at an operational use of the map, as it highlights the percentage of the territory most prone to wildfires (e.g. 5% of territory extremely prone to wildfires corresponding to the fifth class, 20% of highly

**Table 5.** Intensity classes are defined based on expert judgement as a function of the vegetation cover codified by the third level of CLC18.

Intensity classes	Description	Typical vegetation coverage
1	Low-intensity surface wildfire	Crops, grasslands
2	Medium-intensity surface wildfire	Broadleaves, tree crops
3	High-intensity surface wildfire	Shrubs, sclerophyllous vegetation
4	Very-high-intensity crown fires	Conifers, mixed forest

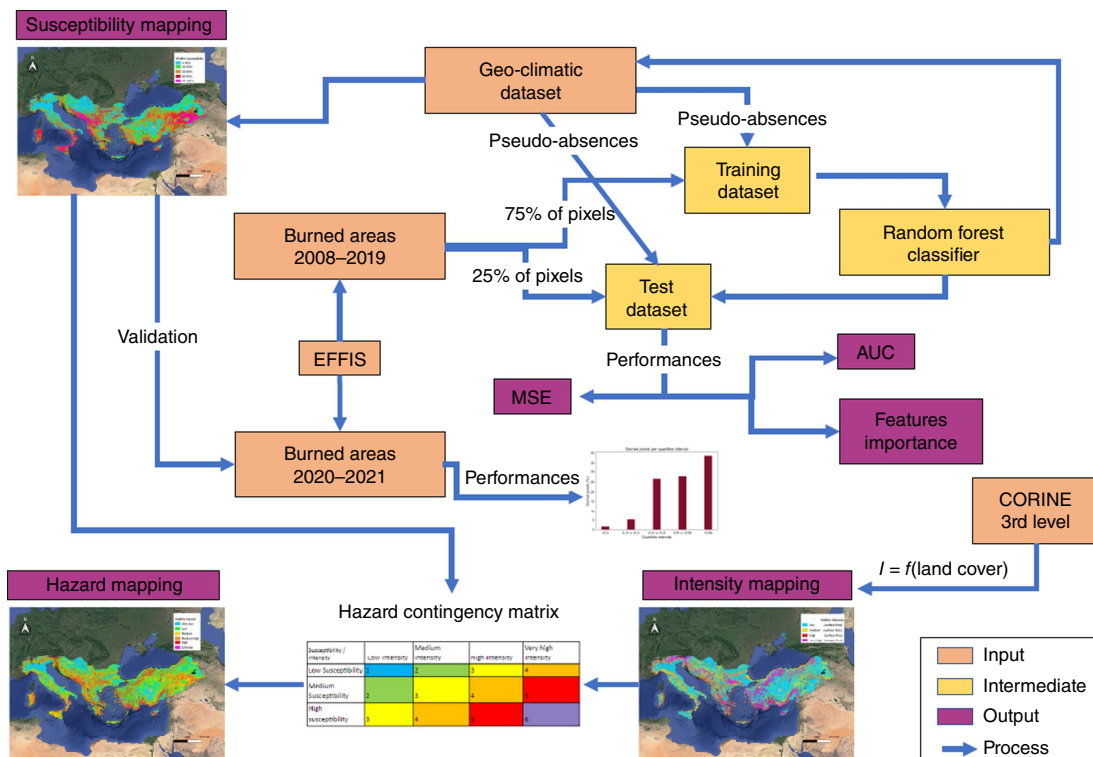
In particular, three types of surface wildfires ranging from low to high intensity were defined, while the fourth class represents crown fires, affecting mainly the coniferous vegetation typically associated with very high intensity.

susceptibility corresponding to pixels associated with Class 4 or 5, and so on), discriminating those areas from the part of the territory characterised by lower susceptibility levels.

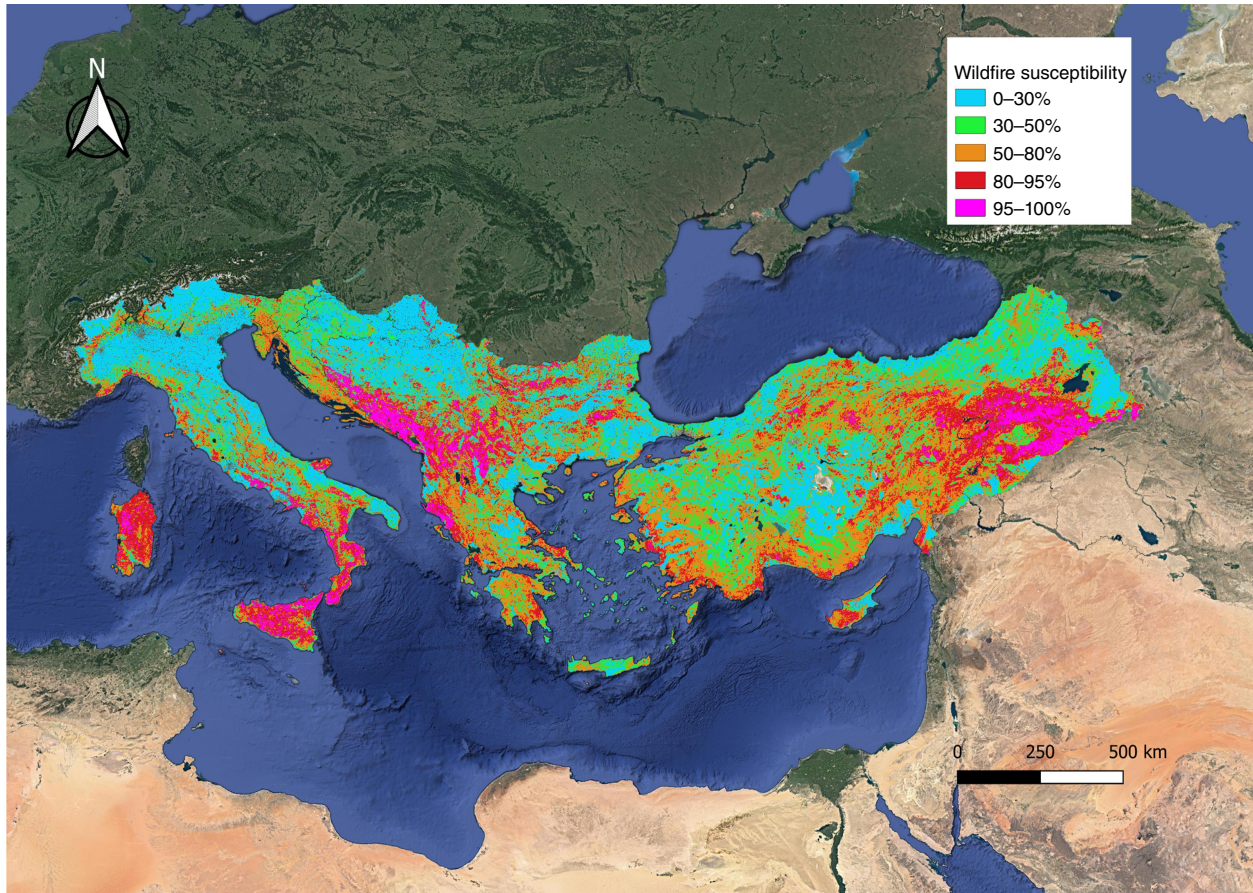
The latter subdivision was used to assess the performance indicators on the validation dataset containing the burned area for the years 2020 and 2021. Test pixels associated with past wildfires are assigned to each of the above classes and their susceptibility distribution can be visualised by a histogram. If the susceptibility map is well built, most of the testing wildfire pixels will then fall on the highest susceptibility classes.

When the ML model was trained, its input features (i.e. the predisposing factors) were ranked according to their relevance assessed by computing the Gini impurity (Breiman 2001) to identify the main drivers of wildfire occurrence in the study area.

Four different ML model configurations were selected in order to analyse how the contribution of vegetation and climate variables have an impact on model outcome. This allowed us to infer specific information on the behaviour of the variables that are directly involved in the wildfire scenario (Table 4). The tested configurations include a standard model with the whole ensemble of predisposing factors, a model that does not use the neighbouring vegetation



**Fig. 2.** Schematic representation of the adopted methodology for wildfire hazard assessment and mapping. An RF classifier is trained using information on wildfire occurrences (to get labels of training data) and the related geo-climatic predisposing factors on selected burned and unburned pixels. The model is then used to classify each pixel of the geo-climatic dataset, producing a susceptibility map. A wildfire intensity map is produced inferring categories from land use/land cover data. Susceptibility and Intensity classes combine to produce the output hazard map.



**Fig. 3.** Susceptibility map divided into five classes based on the defined percentile values of Table 3.

variable (vegetation continuity), a model that excludes the climate variables, and finally an experiment with a simplified set of climate variables, where only the mean annual cumulative precipitation and the mean annual temperature are kept, similarly to the approach to the Italian susceptibility mapping of Trucchia *et al.* (2022). The performance of all such models in terms of ROC AUC and MSE is tabulated. The importance of the predisposing factors in each of the tested configurations is then discussed.

### Intensity map

Different approaches are available to compute wildfire intensity and/or behaviour. Some of them rely on empirical relationships between fuel cover, fuel structure, and a worst-case scenario for fuel moisture and wind conditions. The one adopted in this work is simpler and relies on expert-based classification.

Here, the identification of different intensity classes is based solely on vegetation cover. In particular, four classes of wildfire severity were defined on the basis of the fuel type codified following the third level of CLC18 land use (see Table 5).

### Hazard map

In the proposed framework, the hazard map is obtained as a function of both the susceptibility and intensity layers (Fig. 2). As the intensity has been aggregated into categories, the susceptibility is also aggregated via a quantile-based analysis of the continuous distribution of susceptibility in the study area. As mentioned before, the input susceptibility is divided into three main classes: Low, Medium and High susceptibility (see Table 3). For the hazard assessment, a contingency matrix is adopted. The matrix takes as inputs the susceptibility classes (rows) and the intensity classes (columns), then six hazard levels are defined according to expert judgement.

The hazard classes numerical values range from 1 to 6. From lowest to highest, they correspond respectively to Very Low, Low, Medium, Medium-High, High and Extreme wildfire hazard. The matrix is designed in a way such that high levels of hazard are present in relation to medium/high probability of wildfire occurrence coupled with high/very high potential severity. In the case of very high levels of wildfire intensity and high susceptibility, the hazard is assigned an extreme value. Low/medium intensity values are always associated with low to medium-high hazard levels.



**Table 6.** Association between land cover codes from CLC2018 and the intensity classes described in Table 5.

CLC 18 class	Description	Intensity class
211	Non-irrigated arable land	1
212	Permanently irrigated land	1
213	Rice fields	1
221	Vineyards	1
222	Fruit trees, berry plantations	2
223	Olive groves	2
231	Pastures	1
242	Complex cultivation patterns	1
243	Land principally occupied by agriculture, with significant areas of natural vegetation	1
244	Agro-forestry areas	2
311	Broad-leaved forest	2
312	Coniferous forest	4
313	Mixed forest	4
321	Shrub and/or herbaceous vegetation associations	1
322	Moors and heathland	3
323	Sclerophyllous vegetation	3
324	Transitional woodland/shrub	3
331	Beaches, dunes, sands	1
332	Bare rock	1
333	Sparsely vegetated areas	2
334	Burnt areas	3

All codes corresponding to non-vegetated areas are aggregated into a code '0 – not burnable'.

## Results

Wildfire susceptibility, intensity and hazard obtained through the methodology above were estimated for the study area. Evaluating the trained ML model over the whole geotopographical and climatic dataset allows retrieval of the estimated probability value associated with the pixels of the study area. In this way, the susceptibility output is characterised by continuous values between 0 and 1, which are then aggregated into the classes shown in Table 3. The resulting susceptibility map refers to the base configuration (with the full ensemble of predisposing factors) and is shown in Fig. 3.

Wildfire potential intensity maps are defined over the whole study area, based on the land cover information retrieved from CLC18 at the third level of detail. Aggregating such classes as discussed in Table 5 and detailed in Table 6 allows definition of the areas in which wildfires are expected to be more severe solely as a function of the vegetation type.

As a result of such spatial fuel aggregation, the intensity map is shown in Fig. 4.

The hazard map in the eastern European Mediterranean countries selected for this study is provided through a contingency matrix (Fig. 5) that takes as inputs susceptibility and intensity. It is shown in Fig. 6.

## Performance of ML model for susceptibility mapping

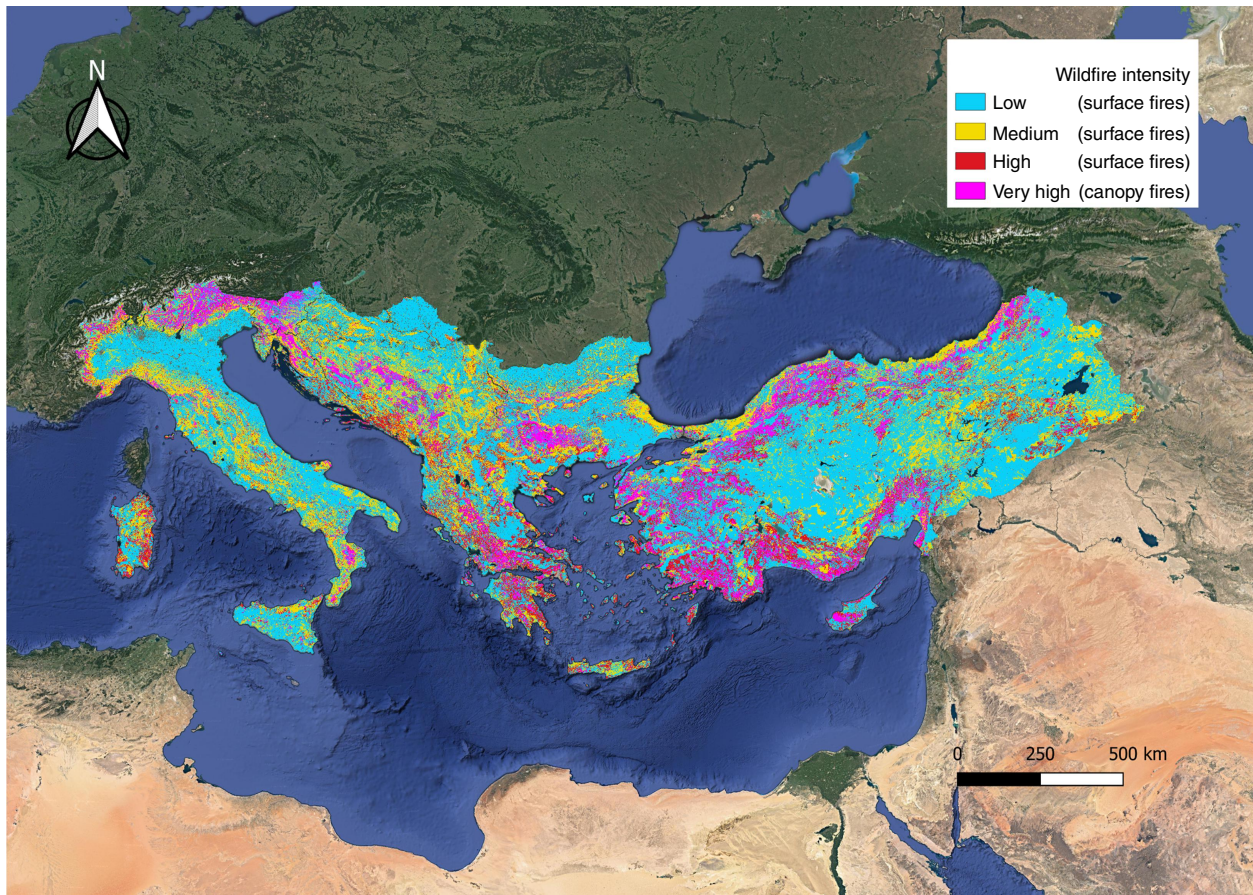
In the following, the RF model performance is assessed, considering the different model configurations adopted. Indicators like the ROC AUC and MSE are evaluated for different testing datasets. Those indicators use the raw susceptibility values, without any percentile-based aggregation. Particular attention is given to the importance of the variables assessed with different methods such as the Gini impurity or the drawing of partial dependence plots. The objective is to get a better understanding of the main predisposing factors' role associated with wildfires. The latter factors can be included in four main variable classes: climatic, vegetational, topographical and anthropic. Particular focus on the role of the neighbouring vegetation variable, especially compared with the local land cover information, is given. The first testing dataset is provided by random sampling of the 25% of pixels from the ML predisposing factors dataset, as mentioned in the Methodology section. The second set includes the ensemble of all the wildfire burned areas retrieved by EFFIS in the years 2020–2021 that were never used for the RF classifier's training phase.

Table 7 shows the ROC AUC values and the MSE as well for the two datasets.

The model configuration including all the predisposing factors performs better than the others, as expected. Good performance in terms of AUC and MSE is also achieved removing the neighbouring vegetation; however, in the following analyses, we illustrate how this affects the susceptibility mapping. Excluding the eight climate variables results instead in a decrease of the AUC values as well as an increase of the MSE for both the test pixels and the fire validation set. Introducing just two climate variables (experiment E4), one observes a definite improvement in the ROC AUC indicator. These results confirm that the climatic variables have an important role in assessing wildfire susceptibility, as will be described in the following paragraphs, but also that including more variables, up to eight, (see Table 1) does not improve model performance in a linear way, mainly owing to the physical correlation between many climatic variables. Fig. 7 shows the Pearson correlation coefficient matrix of the predisposing factors, highlighting the high level of correlation between the climatic variables.

The ROC curves evaluated for the validation dataset (2020–2021 burned areas) are plotted in Fig. 8.

As an additional computation, the distribution of susceptibility over the burned areas of 2020–2021 is assessed. The



**Fig. 4.** Wildfire potential intensity map in the eastern European Mediterranean countries. Blue, yellow and red represent the different intensity levels associated with the occurrence of surface fires. Purple identifies the possibility of having severe canopy wildfires, mainly owing to the presence of conifers.

Susceptibility/intensity	Low intensity	Medium intensity	High intensity	Very high intensity
Low susceptibility	1	2	3	4
Medium susceptibility	2	3	4	5
High susceptibility	3	4	5	6

**Fig. 5.** Hazard contingency matrix (for increasing hazard levels from 1 to 6). It combines the input classes of intensity (columns) and susceptibility (rows). Every entry of the matrix is the hazard level relating to a specific susceptibility and intensity class.

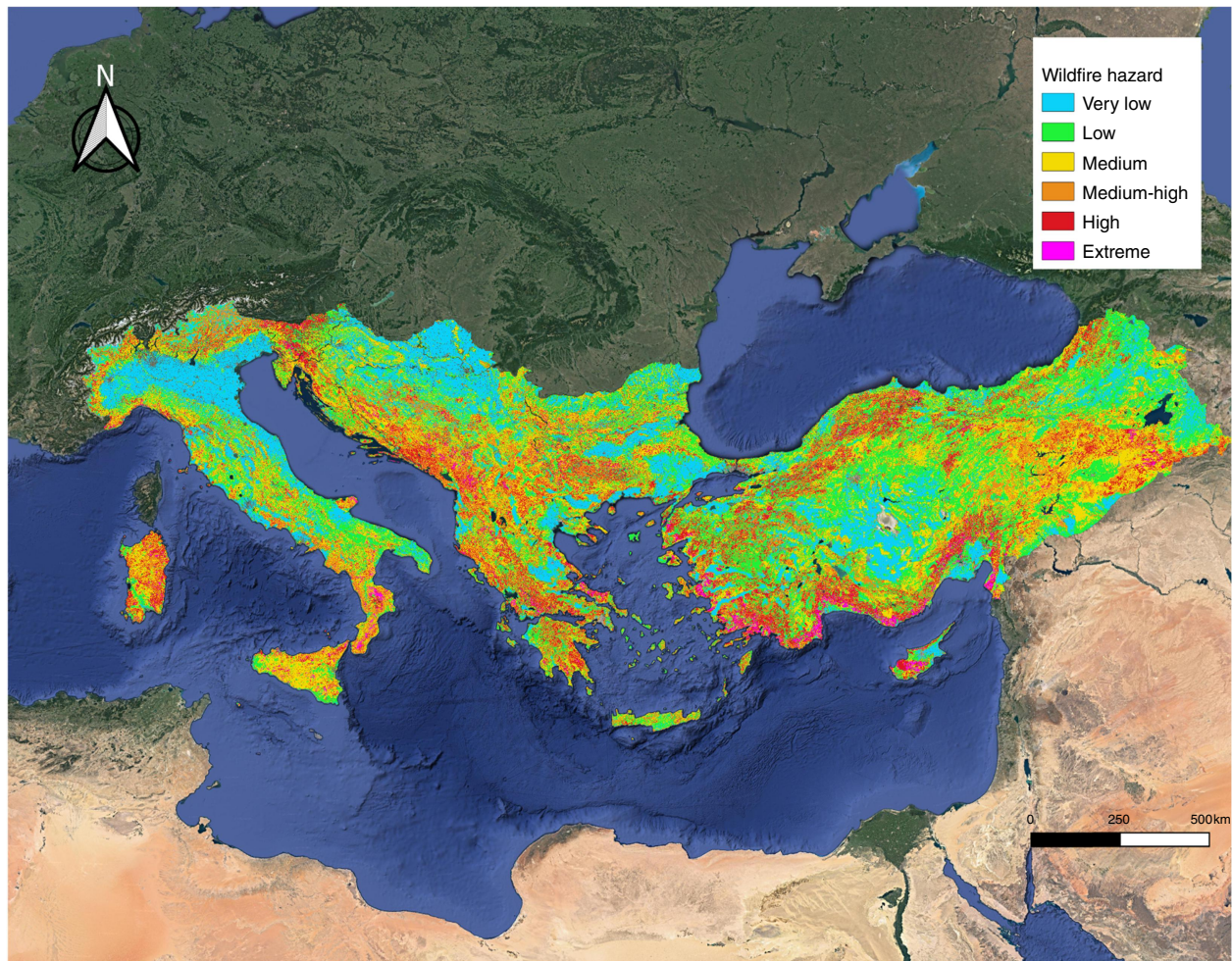
susceptibility is aggregated into the classes of Fig. 3. In general, the more test burned pixels fall in the upper fifth percentile of the susceptibility distribution, the more reliable the map is in identifying fire-prone areas. Fig. 9 reports the results of this analysis.

Testing the susceptibility mapping over the most recent burned areas gives makes it possible to understand whether the model is working as expected, reaching a good level of generality and the ability to overcome

issues related to overfitting on the training dataset. Here, most of the burned areas were characterised by a medium or high level of susceptibility, with more than 60% of pixels falling in the two highest susceptibility classes.

### Ranking drivers' importance

Ranking the importance of the driving variables is a crucial step to gain a fully explainable Artificial Intelligence (XAI)



**Fig. 6.** Hazard map of the study area. It is a categorical map with six classes, from Very Low to Extreme. Red and violet pixels cover the areas in which a severe wildfire has a high probability of occurring.

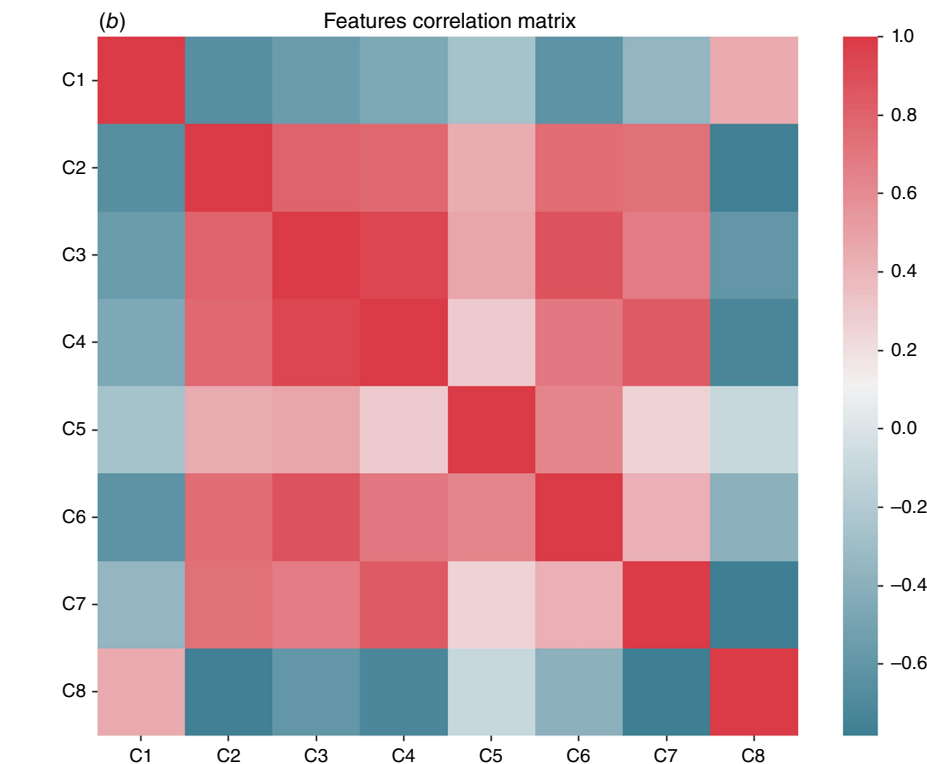
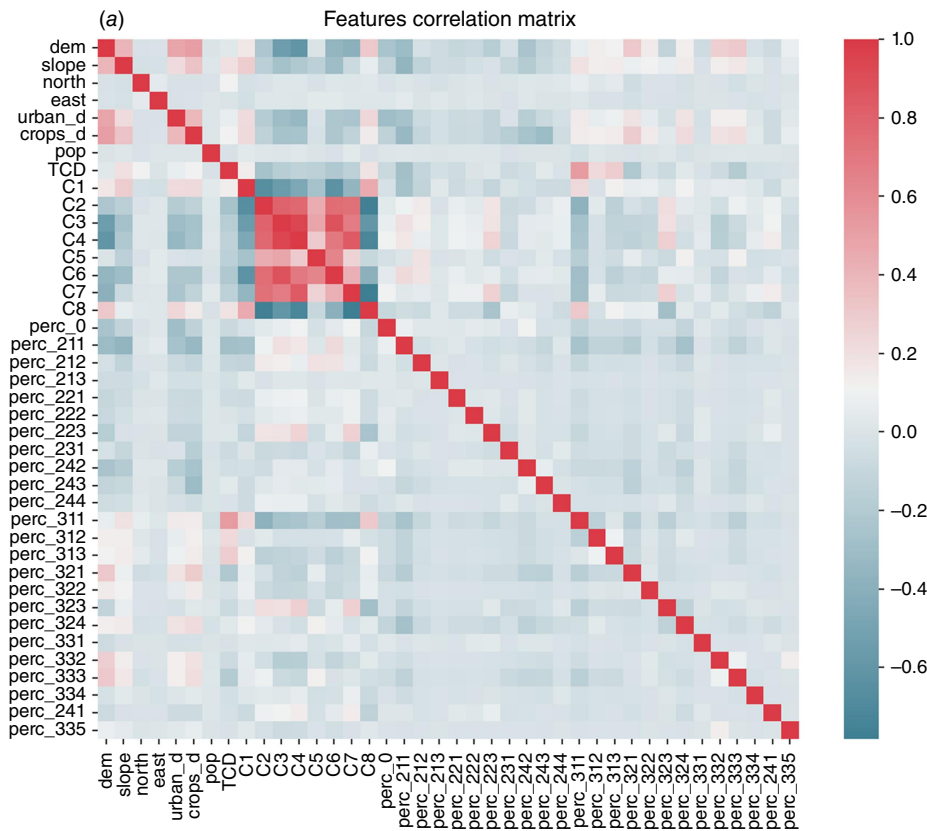
**Table 7.** The performance indicators evaluated using both the test dataset that include a random selection of pixels in the adopted dataset corresponding to burned areas 2008–2019, and the validation dataset of burned area values for 2020–2021.

Model	Experiment code	Test pixels of training burnt polygons		Pixels of training years burnt area 2020–2021	
		MSE	ROC AUC	MSE	ROC AUC
All variables	E1	0.182	0.789	0.282	0.833
No neighbouring percentages vegetation	E2	0.183	0.787	0.292	0.822
No climate	E3	0.203	0.742	0.250	0.767
Simplified climate	E4	0.192	0.770	0.285	0.819

The indicators refer to the four model configurations adopted.

framework, as it is of paramount importance that the results of the ML techniques can be understandable by scientists, end-users and practitioners. XAI is the opposite of the concept of the ‘black box’ in ML where even its designers cannot explain why an AI procedure led to a specific decision (Barredo Arrieta *et al.* 2020). Analysis of the drivers’

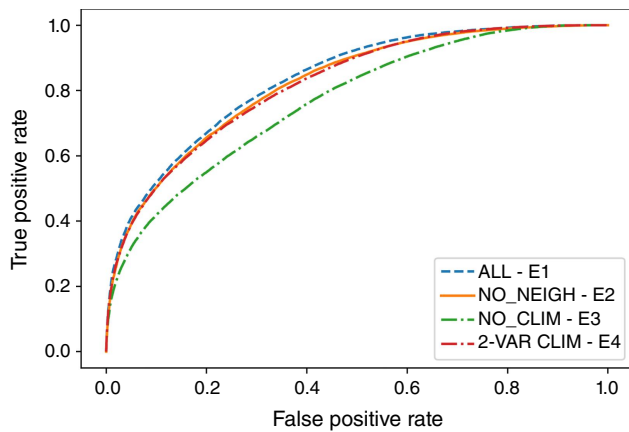
importance mainly relies on the Gini index (Breiman 2001) computed via the sklearn Python package (Pedregosa *et al.* 2011). The post-processing on the trained RFs allows an estimation of variable importance in defining the predicted label based on the Gini impurity, which is evaluated at the different nodes in the tree estimator’s



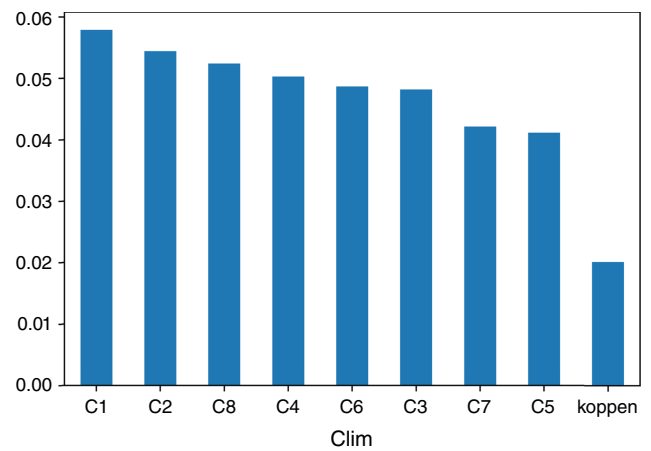
**Fig. 7.** In this figure, the Pearson correlation coefficient matrix is plotted. Blue, red and white cells respectively represent positive, negative and no correlation between the entries. In (a), all the predisposing factors are considered, while (b) portrays a subset of the whole matrix showing the correlation between the climate factors. The analysis was carried out using all the information of the training dataset (see Materials and methods section). C1 stands for mean precipitation, C2 for the maximum number of consecutive dry days, C3 for the maximum temperature, C4 for the mean temperature, C5 for the annual number of hot days, C6 for the number of summer days, C7 for the number tropical nights, C8 for the soil moisture.

branches. The feature importance histograms show the normalised impact of each driving variable on the model outcome, so that the sum of all the importances is always one.

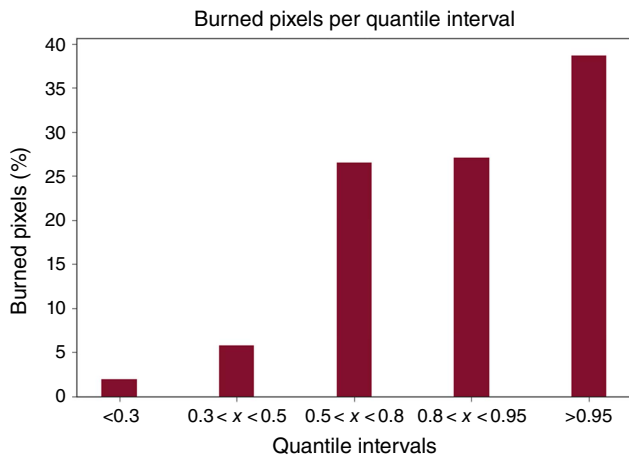
Several importance graphs are plotted. In particular, Fig. 10 shows the relative importance of the main categories, separating climate (green), topographic (red), vegetational



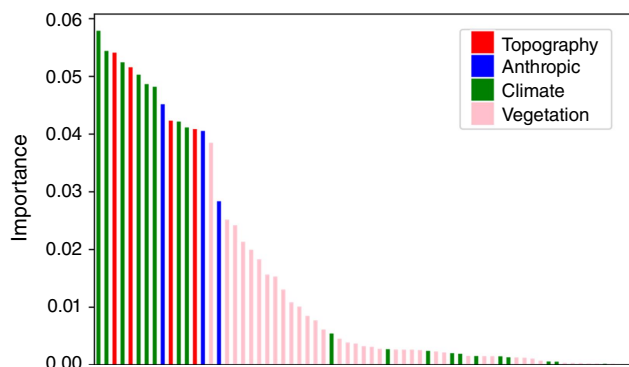
**Fig. 8.** ROC curves for the three model configurations: ALL represents the model with the full ensemble of predisposing factors, NO\_NEIGH does not include the neighbouring vegetation while NO\_CLIM does not include the climatic variables. 2-VAR CLIM includes just precipitation and temperature climatic variables.



**Fig. 11.** Importance barplot for climatic variables of Fig. 10, with the different Köppen–Geiger one-hot-encoded variables that have been summed into the category ‘koppen’. From more to less important, C1 stands for mean precipitation, C2 for the maximum number of consecutive dry days, C8 for the soil moisture, C4 for the mean temperature, C6 for the number of summer days, C3 for the maximum temperature, C7 for the number of tropical nights, C5 for the annual number of hot days.



**Fig. 9.** Histogram showing the frequency of burned pixels in the years 2020–2021 for each susceptibility class. More than 85% of the pixels fall in the 50% most susceptible areas (including the Medium, High and Very High classes) while more than 60% of pixels are found in High to Very High susceptibility classes.



**Fig. 10.** Importance bar plot for all the predisposing factors for Model configuration E1, which accounts for the whole set of inputs.

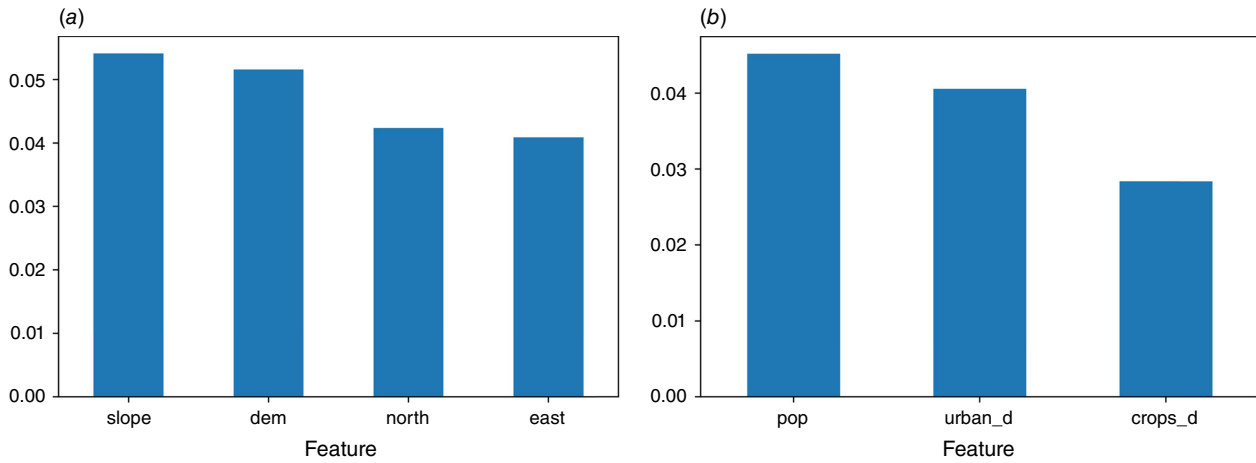
(pink) and anthropic (blue) components. In this choice, the categorical variables that are split into several predisposing factors through one-hot encoding (Potdar *et al.* 2017) are represented with their own importance, without performing any aggregation. The results indicate that climate and topography are definitely the most important factors, followed by anthropic factors and finally by vegetation.

The plots shown in Figs 11–13 report the importance of the various parameters for each category.

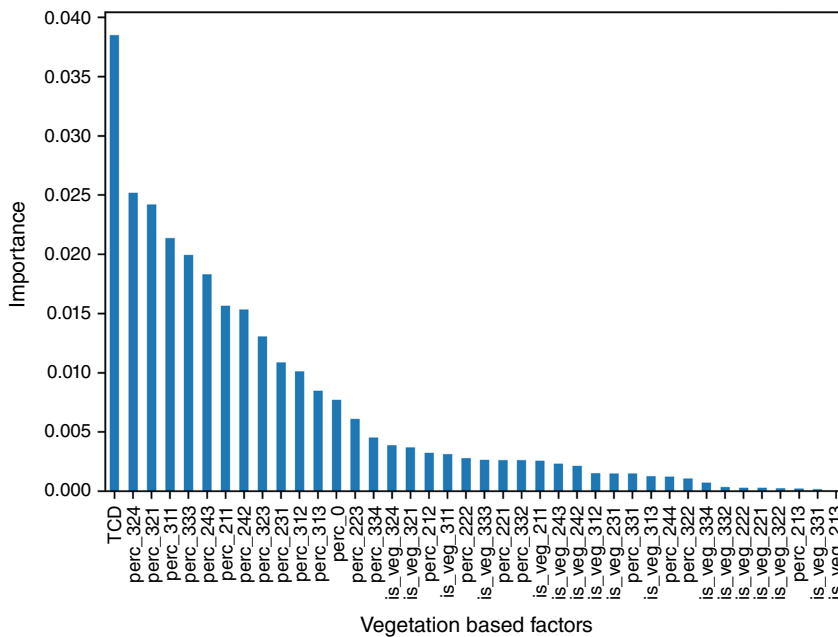
Concerning climate, the first three parameters in the ranking are all related to precipitation, followed by parameters related to air temperature. The eight climate indices from the Köppen–Geiger climate classification are in the lowest part of the ranking (in the plot, we show the sum of the importance of each Köppen index that comes from the one-hot encoding of the related categorical variable). Despite the highest resolution, these variables are almost unused in the classification.

For topography, slope and elevation are more relevant than exposure, whereas considering anthropic factors, population density and the distance from urban areas are more important than the distance from crops (Fig. 12a, b).

For vegetation, TCD is the first parameter in the ranking of Fig. 13. Most fires in the study region spread to rural areas and shrublands where the TCD is very low. In addition, high-density forests are mostly represented by old mature forests where the frequency of fires is extremely low compared with rural areas and shrubland. Importantly, vegetation continuity of the different vegetation classes remains among the most important factors. All parameters related to the class of vegetation itself in each pixel are almost unused in the



**Fig. 12.** Topography (a) and anthropic (b) variable importance for experiment EI (detail of Fig. 10).



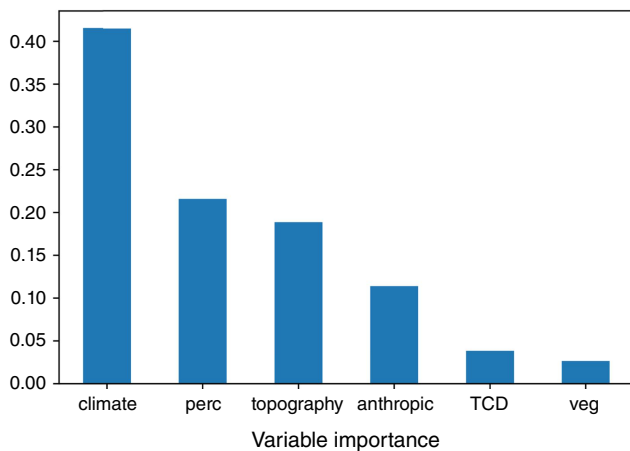
**Fig. 13.** Vegetational variable for the experiment EI (detail of Fig. 10). TCD stands for Tree Cover Density; perc\_XXX for the percentage of neighbouring variables related to Corine Land Cover class XXX; is\_veg\_XXX stands for the class of the analysed pixel (not accounting for the neighbourhood). The local vegetation variables originated from the categorical ‘vegetation’ class after a one-hot encoding procedure.

algorithm if compared with the respective, spatially averaged, neighbouring percentage variables.

Almost all parameters characterised by the highest ranking are related to the continuity of fine and flammable vegetation from transitional woodland/shrubland, shrublands, rural areas, grassland and sclerophyllous vegetation. However, the third position in the ranking is represented by broadleaf forest. In this ranking, high importance of the parameter does not necessarily mean high frequency of wildfire but only that in many cases highly discriminating decisions of the random trees are taken considering that parameter. This is in line with the results obtained at national scale and is even more important for the larger-scale study region considered here.

Important information comes from considering the total importance of each category after opportune aggregation into groups (see Fig. 14). Climate variables emerge as the most important category, confirming that climate plays a crucial role for estimating susceptibility at the supranational scale of the study area. The second position is taken by the neighbouring percentage vegetation (perc), which has been considered separately from TCD and the vegetation itself (veg). The ranking continues with the topographic and anthropic categories. TCD and the vegetation of the single pixel are less important.

Given its importance, it is thus important to analyse in detail the role of neighbouring vegetation: from a modelling point of view, the dataset includes a number of features



**Fig. 14.** Gini importances aggregated by category and summed for display purposes. Climate, climatic variables; perc, neighbouring vegetation variables; topography, topographic variables; anthropic, anthropic variables; TCD, Tree Cover Density; veg, vegetation of the single pixel that has been encoded into several binary variables.

equal to the number of vegetation types, as explained in the Materials and methods section, whereas the important contribution comes from the sum of all such features. However, these features represent partial information, because they give solely the percentage of a specific type of vegetation per pixel and not the whole composition of vegetation in the surrounding area. In contrast, the other categories represent an ensemble of different variables, even if they belong to the same more general category (i.e. climate). Having clarified this aspect, it is thus important to note that neighbouring vegetation always has importance higher than 20%, while the vegetation of a single pixel is always under 5%. The continuity of vegetation, then, truly impacts the choices of the ML algorithm, giving much more information than just the local land cover data. This result highlights the role of the neighbouring vegetation shown in the first part of this section, offering quantitative evidence for its impact on the model. Climate is indeed very important in discriminating if a pixel has a higher or lower probability of burning. However, as shown by Fig. 11, Table 1, climate information in the analysed experiment includes nine variables, which also have a potential high degree of correlation. In fact, they show a very similar importance, where each value has a weight ranging from 5 to 6%.

The importance plots for experiments E2, E3 and E4 follow similar trends to the ones shown in this paper, with the obvious removal of some columns from Fig. 10 (which expresses the ranking for the complete set of inputs) but respecting, with few limited exceptions, the same ranking.

In conclusion, both climate and neighbouring vegetation add important information to the dataset, helping the model to get more accurate results. The variable describing vegetation continuity is considerably the most relevant, expressing the full potential if coupled to at least two climatic

variables. The vegetation of an individual pixel does not improve model classification when vegetation continuity is not included. In that case, the climatic and topographic variables have the largest impact on the final classification.

## Performance of hazard mapping

A method for testing the hazard map is proposed in this section. Validating the hazard map is not a trivial task, as it gives information on wildfire probability and severity that cannot be directly tested either on the whole historical set of wildfires or on a random sampling of it. To cope with the limits of data availability on fire severity for the period 2008–2019, validation focused on the most severe wildfires occurring in summer 2021 in southern Turkey. For the event considered, the Union Civil Protection Mechanism (UCPM) was activated.

The situation was summarised by ERCC (Emergency Response Coordination Centre) through a dedicated European Civil Protection and Humanitarian Aid Operations (ECHO) daily map (Fig. 15a) (ERCC - Emergency Response Coordination Centre 2021).

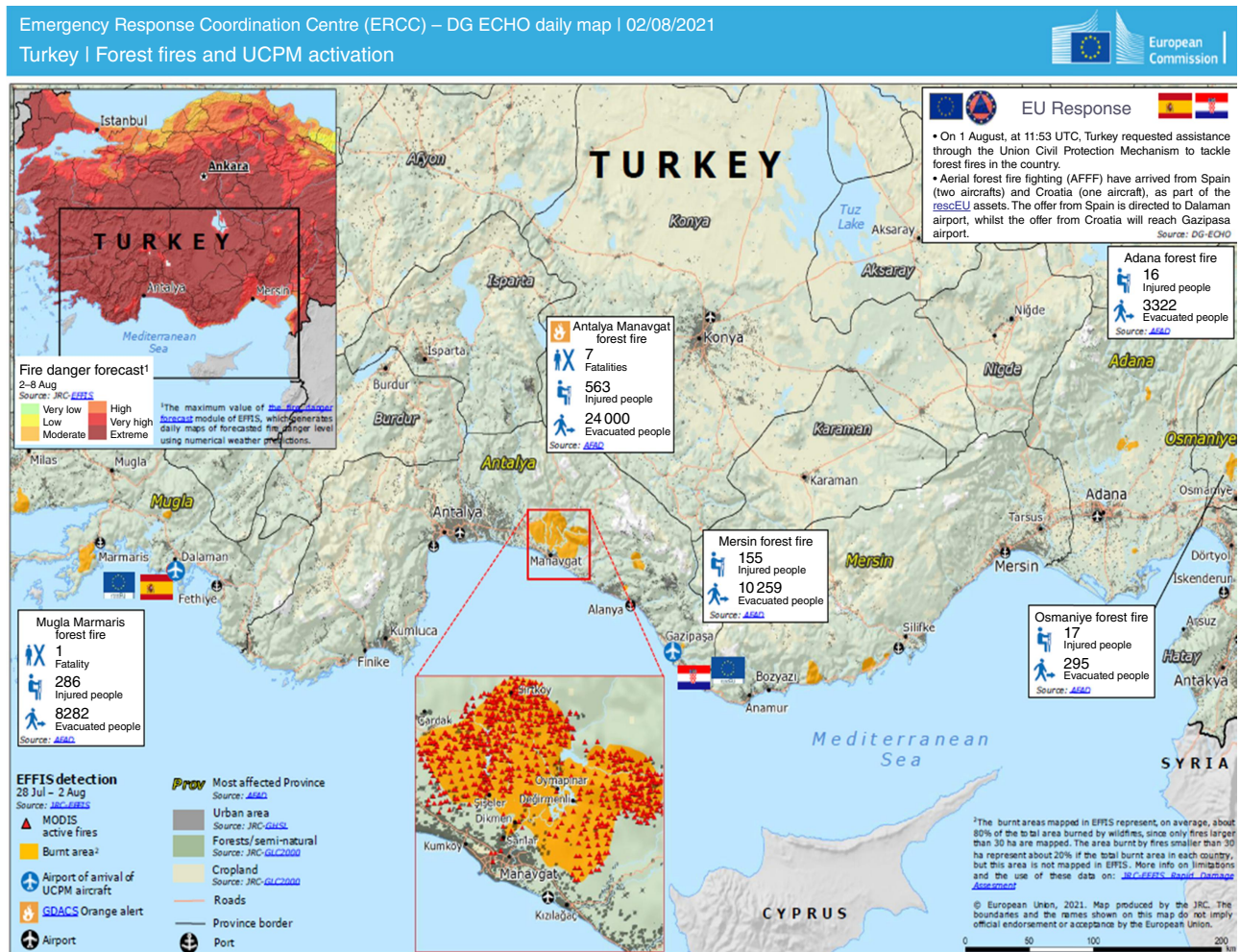
The frequency of the hazard classes belonging to each pixel was estimated in the areas burned in this event, as reported by the ECHO map, showing that ~70% of the burned area was characterised by high and extreme hazard classes; see Fig. 15b.

The results obtained in such hazard distribution histograms show that these wildfires took place in very hazardous areas. The majority of the burned pixels are characterised by hazard classes equal to or greater than 5, which indicates that High and Extreme wildfire hazard was expected in such areas. These results offer concrete support for a positive validation of the hazard map, underlining the ability to identify the portion of the territory that is most frequently affected by intense wildfires.

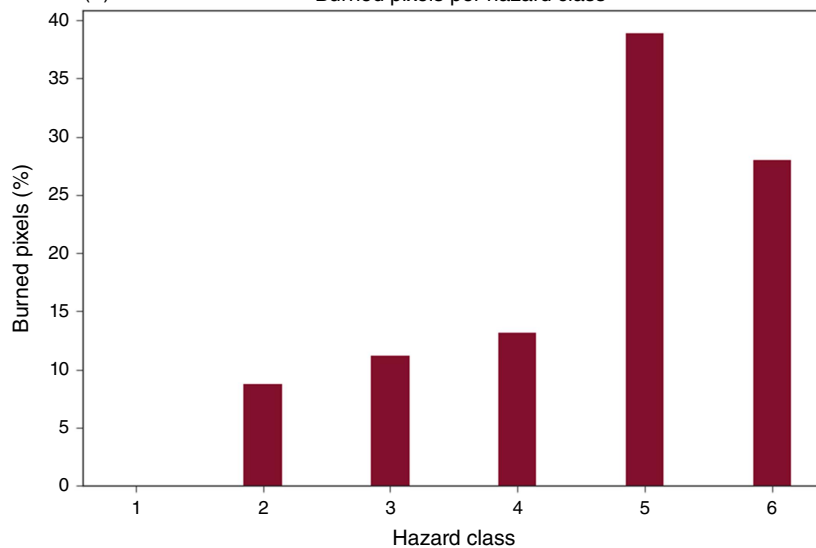
## Discussion

By embracing a wide-scale study area characterised by different climates, we found that climate indices play a major role in wildfire hazard assessment. In this case, climate helps to discriminate, within the same class of vegetation, those areas that are potentially more susceptible to wildfire. For instance, in terms of wildfire hazard, it is essential to discriminate between Mediterranean pines and mountain spruces within the class conifers. However, shrubs and sclerophyllous vegetation are not always characterised by very high hazard depending on the current climatic conditions, but climate change could increase their level of hazard in the future. For these reasons, it is urgent to implement prevention activities by encouraging the establishment of shared priorities based on objective and reliable wildfire hazard assessment and mapping. A key opportunity for

(a)



(b) Burned pixels per hazard class



**Fig. 15.** (a) ECHO Daily Map of 2 August 2021. Burned area retrieved by EFFIS from the severe wildfires that occurred between 28 July and 2 August 2021, which required the activation of the Union Civil Protection Mechanism (<https://reliefweb.int/map/turkey/turkey-forest-fires-and-ucpm-activation-dg-echo-daily-map-02082021>). (b) Distribution of the hazard values in the burned areas that required an UCPM activation between 28 July and 2 August 2021, from Very Low to Extreme.



discriminating between priorities is to consider wildfire risk scenarios by including the exposed elements, in terms of their vulnerability and value, thus identifying areas characterised by very-high-risk scenarios.

Among the four categories of drivers considered here, vegetation and anthropic features are the only drivers that can be managed by landscape planners and fire managers through specific interventions such as fuel treatment in highly populated areas and expansion of broadleaf forests.

As an additional remark, the validity of the maps produced holds as long as no major change affects the input layers, such as land cover, anthropic layers, or the climatic conditions. In order to better reflect changes in climatic conditions and land use, such maps should be updated when input data change significantly, also providing new data for the burned areas that constitute the observed variable in the susceptibility modelling.

## Conclusions

This work analysed the factors controlling susceptibility to wildfires in a large region of ~1 612 500 km<sup>2</sup>, encompassing eastern Mediterranean countries from Italy to Turkey, using open-data sources on burned area, climate, topography, anthropic features and vegetation. The susceptibility map highlighted the importance of climate and vegetation continuity in susceptibility assessments. Vegetation continuity has a twofold role: on the one hand, continuity of high flammability fuels is the main factor responsible for severe fire events; on the other hand, continuity of native broad-leaved forests may limit the propagation of large wildfires.

The preliminary results of the risk mapping process and of the susceptibility maps discussed in this work are currently being refinement through: (i) use of more accurate local data, (ii) consideration of the seasonality of the phenomena, (iii) implementation of *ad hoc* spatial validation procedures. A more thorough factor importance analysis, with a special focus on the impact of different vegetation types, is ongoing.

In conclusion, we believe that the preliminary results of the introduced risk mapping process open a relevant path to restoration and adaptation strategies, fostering the objectives of the European Union (EU) Biodiversity Strategy for 2030.

The methodology presented in this work has been adopted for the development of wildfire risk mapping guidelines through the ongoing European program IPA Floods and Fires (IPAFF), which targets the Western Balkans (Albania, Kosovo<sup>2</sup>, Montenegro, Serbia, North Macedonia, Bosnia–Herzegovina) and Turkey. The authors of this work are currently developing, together with national and local experts of a regional working

group across the Western Balkans and Turkey, advanced tools and methodologies to build capacities in wildfire risk assessment and mapping, considering static hazard and risk mapping across national boundaries.

## References

- Ascoli D, Moris J, Marchetti M, Sallustio L (2021) Land use change towards forests and wooded land correlates with large and frequent wildfires in Italy. *Annals of Silvicultural Research* **46**(2), 177–188. doi:10.12899/asr-2264
- Bachmann A, Allgower B (2000) The need for a consistent wildfire risk terminology. In 'Joint Fire Science Conference and Workshop Proceedings: Crossing the Millennium: Integrating Spatial Technologies and Ecological Principles for a New Age in Fire Management. Vol. 1', Boise, Idaho. (Eds LF Neuenschwander, KC Ryan, GE Gollberg) pp. 67–77. (University of Idaho and the International Association of Wildland Fire: Moscow, ID and Fairfield, WA)
- Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* **58**, 82–115. doi:10.1016/j.inffus.2019.12.012 ISSN 1566-2535.
- Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF (2018) Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data* **5**, 180214. doi:10.1038/sdata.2018.214
- Breiman L (2001) Random Forests. *Machine Learning* **45**, 5–32. doi:10.1023/A:1010933404324
- ERCC - Emergency Response Coordination Centre (2021) ECHO Daily Map of 02 August 2021 - Turkey - Forest Fire - August 2021. Available at <https://erccportal.jrc.ec.europa.eu/ECHO-Products/Maps#/maps/3782> [last accessed 30 June 2022]
- European Environment Agency (2002) Europe's biodiversity – biogeographical regions and seas. Biogeographical regions in Europe. Introduction. EEA Report No 1/2002. Available at [https://www.eea.europa.eu/publications/report\\_2002\\_0524\\_154909](https://www.eea.europa.eu/publications/report_2002_0524_154909) [last accessed 5 December 2022]
- Gholamnia K, Gudiyangada Nachappa T, Ghorbanzadeh O, Blaschke T (2020) Comparisons of Diverse Machine Learning Approaches for Wildfire Susceptibility Mapping. *Symmetry* **12**, 604. doi:10.3390/sym12040604
- Ghorbanzadeh O, Blaschke T, Gholamnia K, Aryal J (2019a) Forest Fire Susceptibility and Risk Mapping Using Social/Infrastructural Vulnerability and Environmental Variables. *Fire* **2**, 50. doi:10.3390/fire2030050
- Ghorbanzadeh O, Blaschke T, Gholamnia K, Meena SR, Tiede D, Aryal J (2019b) Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sensing* **11**, 196. doi:10.3390/rs11020196
- Hardy CC (2005) Wildland fire hazard and risk: Problems, definitions, and context. *Forest Ecology and Management* **211**, 73–82. doi:10.1016/j.foreco.2005.01.029
- Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz Sabater J, Nicolas J, Peubey C, Radu R, Rozum I, Schepers D, Simmons A, Soci C, Dee D, Thépaut J-N (2018) 'ERA5 hourly data on single levels from 1959 to present.' (Copernicus Climate Change Service (C3S) Climate Data Store (CDS)) doi:10.24381/cds.adbb2d47 [accessed 14 April 2021]
- Langanke T, Büttner G, Dufourmont H, Iasillo D, Probeck M, Rosengren M, Sousa A, Strobl P, Weichselbaum J (2013) GIO land (GMES/Copernicus initial operations land) High Resolution Layers (HRLs) – Summary of product specifications. In 'European Environment Agency Copernicus Report'. (European Environmental Agency: Copenhagen, Denmark) Available at <https://land.copernicus.eu/>

<sup>2</sup>This designation is without prejudice to position on status, and is in line with United Nations Security Council resolution (UNSCR) 1244 and the International Court of Justice (ICJ) Opinion on the Kosovo declaration of independence.

- user-corner/technical-library/gio-land-high-resolution-layers-hrls-2013-summary-of-product-specifications [accessed 18 January 2021]
- Ministry of Forests (MOF) (1997) 'Glossary of Forest Terms.' (Ministry of Forests: Province of British Columbia, Canada)
- Ministry of Forests and Range (MOF) (2008) Glossary of Forestry Terms in British Columbia. Available at <https://www.for.gov.bc.ca/hfd/library/documents/glossary/glossary.pdf> [accessed 5 December 2022]
- Nachappa TG, Tavakkoli Piralilou S, Gholamnia K, Ghorbanzadeh O, Rahmati O, Blaschke T (2020a) Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology* **590**, 125275. ISSN 0022-1694/ISSN 0022-1694/ISSN 0022-1694doi:10.1016/j.jhydrol.2020.125275/ISSN 0022-1694
- Nachappa TG, Ghorbanzadeh O, Gholamnia K, Blaschke T (2020b) Multi-Hazard Exposure Mapping Using Machine Learning for the State of Salzburg, Austria. *Remote Sensing* **12**, 2757. doi:10.3390/rs12172757
- National Wildfire Coordinating Group (NWCG) 2003 Glossary of Wildland Fire Terminology. <https://www.nwcg.gov/publications/pms205>
- Novara A, Gristina L, Sala G, Galati A, Crescimanno M, Cerdà A, Badalamenti E, La Mantia T (2017) Agricultural land abandonment in Mediterranean environment provides ecosystem services via soil carbon sequestration. *Science of the Total Environment* **576**, 420–429. doi:10.1016/j.scitotenv.2016.10.123 ISSN 0048-9697
- Olaya V (2009) Basic Land-Surface Parameters. In 'Geomorphometry. Concepts, Software, Applications. Developments in Soil Science. Vol. 33'. (Eds Hengl T, Reuter H) pp. 141–169. (Elsevier: Amsterdam, The Netherlands)
- Parente J, Pereira MG (2016) Structural fire risk: the case of Portugal. *Science of the Total Environment* **573**, 883–893. doi:10.1016/j.scitotenv.2016.08.164
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830.
- Plieninger T, Hui C, Gaertner M, Huntsinger L (2014) The Impact of Land Abandonment on Species Richness and Abundance in the Mediterranean Basin: A Meta-Analysis. *PLoS One* **9**(5), e98355. doi:10.1371/journal.pone.0098355
- Potdar K, Pardawala TS, Pai CD (2017) A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. *International Journal of Computer Applications* **175**, 7–9. doi:10.5120/ijca2017915495
- Pradhan B, Dini Hairi Bin Suliman M, Arshad Bin Awang M (2007) Forest fire susceptibility and risk mapping using remote sensing and geographical information systems (GIS). *Disaster Prevention and Management* **16**, 344–352. doi:10.1108/09653560710758297
- ReliefWeb (2021) Algeria: Forest Wildfires - Emergency Appeal No: MDRDZ007, Posted: 2 November 2021, Originally published 30 October 2021. Available at <https://reliefweb.int/report/algeria/algeria-forest-wildfires-emergency-appeal-mdrdz007>
- San-Miguel-Ayanz J, Durrant T, Boca R, Maianti P, Libertá G, Artés-Vivancos T, Oom D, Branco A, de Rigo D, Ferrari D, Pfeiffer H, Grecchi R, Nuijten D (2022) 'Advance Report on Forest Fires in Europe, Middle East and North Africa 2021'. JRC128678. EUR 31028 EN. (Publications Office of the European Union: Luxembourg) doi:10.2760/039729 ISBN 978-92-76-49633-5.
- Scott JH, Thompson MP, Calkin DE (2013) A wildfire risk assessment framework for land and resource management. Gen. Tech. Rep. RMRS-GTR-315. 83 p. (USDA Forest Service, Rocky Mountain Research Station) doi:10.2737/rmrs-gtr-315
- Tavakkoli Piralilou S, Einali G, Ghorbanzadeh O, Nachappa TG, Gholamnia K, Blaschke T, Ghamisi P (2022) A Google Earth Engine Approach for Wildfire Susceptibility Prediction Fusion with Remote Sensing Data of Different Spatial Resolutions. *Remote Sensing* **14**, 672. doi:10.3390/rs14030672
- 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
- Tedim F, Leone V, Amraoui M, Bouillon C, Coughlan MR, Delogu GM, Fernandes PM, Ferreira C, McCaffrey S, McGee TK, Parente J, Paton D, Pereira MG, Ribeiro LM, Viegas DX, Xanthopoulos G (2018) Defining Extreme Wildfire Events: Difficulties, Challenges, and Impacts. *Fire* **1**, 9. doi:10.3390/fire1010009
- Tonini M, D'Andrea M, Biondi G, Degli Esposti S, Trucchia A, Fiorucci P (2020) A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy. *Geosciences* **10**(3), 105. doi:10.3390/geosciences10030105
- Trucchia A, Meschi G, Fiorucci P, Gollini A, Negro D (2022) Defining Wildfire Susceptibility Maps in Italy for Understanding Seasonal Wildfire Regimes at the National Level. *Fire* **5**(1), 30. doi:10.3390/fire5010030
- Turco M, Bedia J, Di Liberto F, Fiorucci P, von Hardenberg J, Koutsias N, et al. (2016) Decreasing Fires in Mediterranean Europe. *PLoS One* **11**(3), e0150663. doi:10.1371/journal.pone.0150663
- Turco M, Rosa-Cánovas JJ, Bedia J, Jerez S, Montávez JP, Llasat MC, Provenzale A (2018) Exacerbated fires in Mediterranean Europe due to anthropogenic warming projected with non-stationary climate-fire models. *Nat Commun* **9**(1), 3821. doi:10.1038/s41467-018-06358-z
- United Nations Environment Programme (2022) Spreading like Wildfire – The Rising Threat of Extraordinary Landscape Fires. A UNEP Rapid Response Assessment. (Nairobi) Available at <https://www.unep.org/resources/report/spreading-wildfire-rising-threat-extraordinary-landscape-fires>
- 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
- World Bank Group (2021) Metadata Climate Change Knowledge Portal (CCKP). Available at <https://climateknowledgeportal.worldbank.org/media/document/metatag.pdf> [accessed 5 December 2022]
- Yamazaki D, Ikeshima D, Tawatari R, Yamaguchi T, O'Loughlin F, Neal JC, Sampson CC, Kanae S, Bates PD (2017) A high accuracy map of global terrain elevations. *Geophysical Research Letters* **44**, 5844–5853. doi:10.1002/2017GL072874

**Data availability.** All the work was developed with open data whose source is cited in detail throughout the paper.

**Conflicts of interest.** The authors declare no conflicts of interest.

**Declaration of funding.** A. P. acknowledges support from the EU H2020 project FirEUrisk, Grant Agreement No. 101003890. P. F., U. P., G. M., A. T. acknowledge support from the European Commission, DG ECHO project IPA Floods and Fires 'EU Support to Flood Prevention and Forest Fires Risk Management in the Western Balkans and Turkey', Agreement No. ECHO/BI/2020/IPA/834998. P. F., G. M., A. T. acknowledge support from the EU-funded programme 'Prevention, preparedness and response to natural and man-made disasters in Eastern Partnership countries – Phase 3 (PPRD East 3)' Ref. Ares (2019)5721249 – 12 September 2019.

#### Author affiliations

<sup>A</sup>CIMA Research Foundation, I-17100 Savona, Italy.

<sup>B</sup>Istituto di Geoscienze e Georisorse del CNR, Via Moruzzi 1, 56124 Pisa, Italy.

<sup>C</sup>Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment, University of Lausanne, CH-1015 Lausanne, Switzerland.

<sup>D</sup>University of Rome 'La Sapienza', Scuola di Ingegneria Aerospaziale, Via Salaria 851, 00138, Rome, Italy.