fire

MDPI

Article

Inferring Wildfire Ignition Causes in Spain Using Machine Learning and Explainable AI

Clara Ochoa [1,*], Magí Franquesa [2,3], Marcos Rodrigues [4,5] and Emilio Chuvieco [1]

[1] Environmental Remote Sensing Research Group, Department of Geology, Geography and the Environment, Universidad de Alcalá, Calle Colegios 2, 28801 Alcalá de Henares, Spain; emilio.chuvieco@uah.es
[2] Instituto Pirenaico de Ecología, Consejo Superior de Investigaciones Científicas (IPE-CSIC), 50059 Zaragoza, Spain; magi.franquesa@ipe.csic.es
[3] Laboratorio de Climatología y Servicios Climáticos (LCSC), CSIC-Universidad de Zaragoza, 50018 Zaragoza, Spain
[4] Department of Geography and Land Management, University of Zaragoza, 50009 Zaragoza, Spain
[5] GEOFOREST Group University Institute of Research in Environmental Sciences—IUCA, University of Zaragoza, 50009 Zaragoza, Spain
* Correspondence: mclara.ochoa@uah.es

Abstract

A substantial proportion of wildfires in Mediterranean regions continue to be recorded without information about the cause or source of ignition, limiting our ability to understand ignition drivers and design effective prevention strategies. In this study, we develop a spatially harmonised wildfire database for mainland Spain by integrating ignition records from the Spanish General Fire Statistics (EGIF) with fire perimeters generated from satellite images. We then apply a Random Forest classifier to infer ignition causes for events lacking cause attribution. To interpret model behaviour, we use Shapley Additive Explanation (SHAP) values at both global and local scales. Results indicate that human-caused ignitions are dominant, with intentional and negligence-related fires accounting for 52.13% of all known events, although they are associated with contrasting climatic and land-use settings. Negligence-related fires tend to occur under hot, dry and windy conditions, often in agricultural interfaces, whereas intentional fires are more frequent under cooler and wetter conditions and in areas with higher population density and land-use change. Lightning-caused fires represent a small fraction of total ignitions (3%) but exhibit a distinct climatic signature, occurring primarily in sparsely populated areas, under intermediate moisture conditions, and often leading to larger burned areas. Despite strong overall model performance (F1-score = 0.82), minority classes (e.g., lightning and fire rekindling, 0.17%) remain challenging to classify, reflecting both data imbalance and uncertainty in causal attribution. Overall, the combined use of machine learning and explainable AI provides a coherent spatial characterisation of wildfire ignition drivers across mainland Spain, highlights systematic differences among ignition causes, and identifies key limitations in existing fire cause records. This framework represents a practical step towards improving fire cause information by integrating remote sensing products with field-based fire reports, thereby supporting more targeted and evidence-based fire risk management.

Keywords: wildfire dataset; ignitions causes; Spain; random forest; SHAP

Academic Editors: Moulay A. Akhloufi and Mozhdeh Shahbazi

Received: 8 February 2026
Revised: 8 March 2026
Accepted: 11 March 2026
Published: 24 March 2026

1. Introduction

Wildfires represent one of the most pressing environmental challenges in Mediterranean regions, yet the causes behind many of these events remain uncertain. Globally, a

considerable proportion of recorded wildfires continue to be classified as having 'unknown' or 'undetermined' origins, revealing a significant gap in information on fire occurrence [1]. Across Europe, about 50% of all wildfires documented between 1999 and 2016 were reported with unknown causes [2], while in Spain between 1991 and 2000, around 21% of fire events lacked an identified ignition source (Spanish General Fire Statistics, 2002). This uncertainty is central to contemporary wildfire management because incomplete or inaccurate attribution limits the ability of researchers and practitioners to understand ignition drivers, model fire occurrence, and design effective prevention strategies. According to previous studies [1,3–5], the relevance of investigating unknown causes can be summarised with three main aspects: the magnitude of the missing data problem, its implications for scientific modelling, and its critical role in fire prevention and mitigation.

Understanding how wildfires vary according to their ignition source—whether natural or human-caused, and within the latter, accidental or intentional—is essential for designing effective prevention policies. Although less frequent in number, natural ignitions can produce some of the largest and most severe fires in Mediterranean regions, particularly when they occur in remote, forested, and mountainous areas where human presence is limited. Their ignition efficiency depends strongly on meteorological conditions that influence both lightning flashes and fuel moisture content [6]. Their impact highlights the need for advanced strike monitoring and targeted fuel management in wilderness zones [3]. Human-caused ignitions dominate the contemporary fire regime in southern Europe, accounting for over 95% of recorded fires [1]. Intentional ignitions are often associated with socio-economic factors, land-use conflicts, or opposition to conservation measures; in Spain, they tend to concentrate in autumn and winter near the wildland–urban and agricultural interfaces [3]. Accidental ignitions—linked to agricultural burning, machinery, and power line sparks, or other human-related activities—follow marked seasonal patterns and commonly occur near settlements and transport corridors [3]. By extending the ignition season beyond the peak fire weather period (e.g., agricultural burning in early spring), human activities widen the temporal window of risk while increasing the complexity of prevention.

Machine learning (ML) has become a cornerstone for advancing the understanding, modelling, and management of wildfires, offering strong capabilities for handling the complexity and nonlinearity of large environmental datasets [7]. Compared to traditional statistical methods, ML algorithms such as Random Forest (RF) have shown higher predictive accuracy in modelling human-caused fire occurrences than standard regression analysis, achieving AUC values of 0.75 and 0.73 in studies conducted in Spain [8]. These approaches also allow researchers to assess the relative importance of climatic, vegetation, and anthropogenic variables, improving the diagnostic understanding of ignition dynamics and spatial variability [3]. Beyond prediction, ML has been instrumental in addressing the long-standing issue of fires with unknown causes, enabling the classification of ignition sources with notable accuracy [2]. Such applications highlight the potential of ML not only to predict fire occurrence but also to uncover the underlying drivers of ignition patterns, while explainable artificial intelligence (XAI) techniques such as Shapley Additive Explanations (SHAP) improve interpretability and stakeholder confidence by revealing how specific variables influence model outputs [2,9,10].

The increasing use of ML in wildfire science and management has therefore highlighted the need for XAI techniques, which address both the complexity of ML models and the relevance of decisions derived from them [2,10,11]. Many algorithms, while highly effective at detecting complex patterns in large datasets, are often considered "black boxes" because the internal decision-making processes can be opaque [10]. Techniques such as SHAP enable researchers to evaluate the primary drivers of wildfire ignitions across terri-

tories to quantify the influence of specific variables on global and local predictions, and improve the credibility of model outputs for prevention planning [11]. Importantly, XAI facilitates the design of targeted prevention strategies by revealing not only where a fire is likely to occur but also why, allowing interventions to focus on key environmental and anthropogenic factors.

This study pursues three main objectives. First, it aims to harmonise the ESFire30 burned area perimeters derived from remote sensing with the EGIF database, which contains field-based information on wildfire causes. By spatially matching fire perimeters from ESFire30 with ignition records from EGIF, this objective seeks to assign observed fire causes from EGIF to the corresponding burned area perimeters in ESFire30. Second, the study aims to infer the probable causes of wildfires for which no causal information is available by using machine learning methods. This objective addresses the high proportion of fires with unknown causes and responds to the critical importance of fire causality information for wildfire management. Third, the study aims to apply explainable artificial intelligence (XAI) techniques to identify the environmental and anthropogenic variables that most strongly contribute to a better understanding of cause-specific fire patterns and help identify areas that are more vulnerable to different types of wildfire ignitions.

2. Methodology

2.1. Study Area

The study area comprises mainland Spain, which covers approximately 493,000 km^2, excluding the Balearic and Canary Islands. The region is characterised by a complex topography that includes extensive plateaus (i.e., the Meseta Central, including the northern and southern subplateaus), several major mountain ranges such as the Pyrenees, the Cantabrian Mountains, the Iberian System, and Sierra Nevada, as well as a diverse set of coastal plains. Peninsular Spain experiences a predominantly Mediterranean climate, marked by hot, dry summers and mild, wet winters, although substantial spatial variability exists. The northern Atlantic regions (e.g., Galicia, Asturias, Cantabria and the Basque Country) have a temperate oceanic climate with higher rainfall, while the southern and eastern areas (e.g., Andalusia, Murcia, and Valencia) experience semi-arid conditions. This climatic and topographic diversity supports a wide range of ecosystems and land-use types, from Mediterranean shrublands and pine forests to dense broadleaved forests, and from rainfed agricultural areas in the hinterlands to densely populated coastal zones.

2.2. Input Fire Datasets

Two databases were used to conduct this study, each providing complementary information on wildfire occurrence. The ESFire30 contains annual fire perimeter data for mainland Spain from 1985 to 2023, including all burned areas larger than 5 ha. It was derived from Landsat satellite imagery using a semi-automatic processing algorithm, with post-processing refinements (Ochoa et al., 2026, under review). In total, ESFire30 contains 119,498 fire perimeters, accounting for 5,977,591 ha of burned area [12].

Ignition information was obtained from the official Spanish fire database, designated as EGIF (Estadística General de Incendios Forestales: Spanish General Fire Statistics). This database is maintained by the Ministry for the Ecological Transition and the Demographic Challenge (MITECO). The database integrates data reported by regional forest and emergency services, ensuring national consistency through a unified reporting protocol. The EGIF consists of a comprehensive and standardised record of wildfire events across Spain since the 1960s, with a systematic national coverage beginning in 1968. It compiles detailed information for each fire event, including the date, ignition cause, size, land cover affected, suppression resources, and environmental conditions at the time of ignition. The database

is currently available in Excel format. Since 2012, fire event records include spatial coordinates for ignition locations; however, for earlier years—and for a substantial portion of the dataset—these coordinates are not available. Consequently, the primary spatial reference system used in EGIF is a 10 × 10 km grid.

Fire causes are categorised into several groups, including lightning, negligence or accident, intentional fires, fire rekindling, and unknown causes. Due to its temporal coverage, the EGIF is one of the most valuable sources for analysing wildfire patterns, drivers, and trends in Spain. The dataset is regularly updated and publicly available at https://servicio.mapa.gob.es/incendios/Search/Publico, last accessed on 1 January 2026. While data are accessible up to 2023 at the provincial level, the most comprehensively verified national dataset dates to 2016.

2.3. Spatialisation of EGIF Ignition Records and Attribution of Fire Cause Information to ESFire30 Perimeters

Although highly valuable, the EGIF database presents important spatial limitations. For fires recorded before 2012, geographic coordinates—representing the reported location of ignition—are often missing or, when included, frequently contain positional errors, likely introduced during data transfer to the database. Prior to 2012, fire locations were referenced using an outdated sheet and grid system (sheets of 110 × 150 km and grids of 10 × 10 km) derived from the 1:250,000 military map, which is still used for official fire reporting (Figure 1a). To associate EGIF ignition information with ESFire30 fire perimeters, we applied a methodology based on Rodrigues and Riva [8] to spatialise the ignition records. Based on the sheet and grid system (Figure 1a), we first created a 'burnable area' layer by masking urban areas, water bodies, and a 100 m buffer around them to avoid placing ignitions on fully non-burnable surfaces. Random points were then generated within each grid cell, with the number of points matching the number of recorded EGIF observations for the corresponding sheet and grid (Figure 1b).

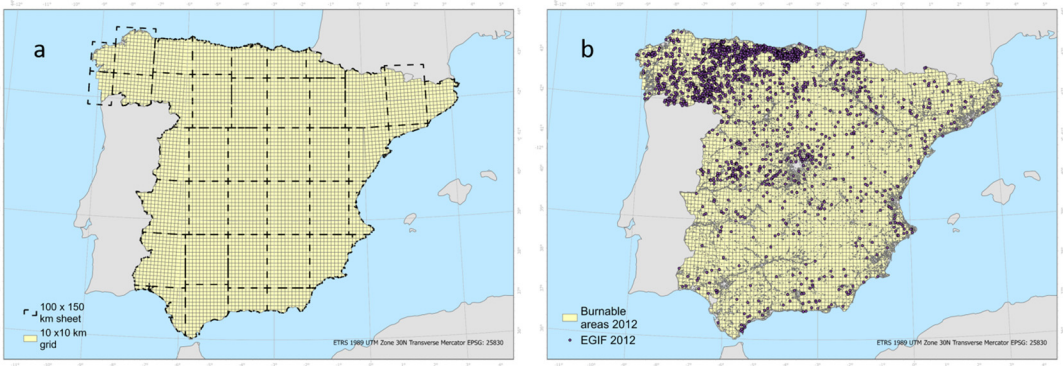


Figure 1. Spatialised ignition points from EGIF database: (**a**) example of the sheet grid system used to locate ignitions; (**b**) resulting spatialised ignition points over burnable areas.

Once the EGIF ignition points were generated, they were integrated into the ESFire30 dataset. However, there were various inconsistencies between the two datasets. In many grid cells, the number of fire perimeters and ignition point records did not coincide. Some cells contained more perimeters than ignition points, others had more ignition points than perimeters, and in some cases only one of the two datasets contained records. To assign a cause to each fire perimeter, we matched EGIF ignition points and ESFire30 fire perimeters within the same grid cell when their burned areas were similar (±10 ha). When such a match was found, the cause from the source ignition point was assigned to the corresponding polygon. If no ignition point with a comparable burned area was available, the predominant cause recorded in that grid cell was used instead, reflecting the

strong spatial clustering and local predominance typically observed in fire cause patterns (MITECO, 2019). To evaluate the reliability of this approach, we calculated the proportion of fires in each grid cell corresponding to the assignment majority cause, confirming that, in most cells, the predominant cause represents the majority of events. Finally, if a grid cell contained ESFire30 polygons but no EGIF ignition points, the polygon was classified as having an unknown cause. The database records the origin of each assignment in a field labelled 'Method', where a value of 1 indicates a similarity-based burned area match and a value of 2 indicates a majority-based assignment. Before modelling, both spatial and comparative analyses were carried out to evaluate the consistency of the integrated dataset.

2.4. Model

In this study, we developed a model to infer wildfire causes for those fire perimeters that did not have an associated cause in the EGIF database. In this way, we should be able to better understand fire drivers, and provide additional insights to improve fire prevention. We used a Random Forest classifier designed to infer the causes of fire events with unknown origin and to identify the environmental and demographic variables that best explained these patterns.

Prior to modelling, spatial analyses were conducted to identify where wildfires were most prevalent according to their cause, and to calculate the proportion of each cause category to establish an expected baseline distribution. These exploratory steps were essential for understanding the underlying spatial and statistical behaviour of the causes in our dataset, ensuring that the results obtained after the modelling process were coherent and meaningful. By characterising these patterns in advance, we were able to better interpret the model outputs and verify that they aligned with the observed dynamics in the study area.

Before beginning the classification stage, it was necessary to address the class imbalance in the target variable. Some cause categories were clear minorities, which can lead to biased model predictions. To alleviate this issue, we applied the Synthetic Minority Oversampling Technique (SMOTE), a widely used method in ecological and environmental modelling to handle class imbalance [2,11,13]. SMOTE generates synthetic samples of the minority classes rather than simply duplicating existing observations. It does so by selecting a minority class instance and creating new synthetic samples along the line segments connecting that instance with its k-nearest minority neighbours. This approach increases the number of minority samples in a more realistic and informative way, helping the model learn more accurate decision boundaries between classes.

In our case, the dataset was split into training (70%) and testing (30%) sets implementing a stratification procedure based on the original class proportions. SMOTE was applied only to the training set (k_neighbours = 2), ensuring that the test data remained independent for validation. A small number of neighbours was selected to ensure that the synthetic samples remained close to the original minority class observations, preserving the local structure of the feature space. SMOTE resampling improves the classifier's ability to correctly identify minority class events. While applying SMOTE in our model does not significantly boost the overall F1-score (82%), it does enhance the performance metrics for individual classes. Importantly, when the trained classifier is applied to the unseen test set, it performs inference in the machine learning sense: predicting the likely ignition cause for new wildfire instances based on parameters learned during training. This inference should not be interpreted as established causal relationships, but rather as generating predictions for data not used during model fitting.

To optimise model performance and prevent overfitting, cross-validation techniques were applied to determine the optimal set of hyperparameters. A grid search proce-

dure (GridSearchCV from the scikit-learn library [14]) was implemented using a five-fold cross-validation scheme, testing combinations of parameters such as the number of trees (n_estimators), the maximum tree depth (max_depth), the minimum number of samples required for node splitting (min_sample_split), the minimum number of samples per leaf (min_sample_leaf), and the maximum number of features considered at each split (max_features). The model achieved a mean cross-validated accuracy of 0.79 $\pm$ 0.004 (95% CI: 0.788–0.792), indicating stable performance across repeated folds. The final model configuration corresponded to the combination of parameters that achieved the highest mean cross-validation accuracy.

Once the classes were balanced, we trained an RF classifier to model wildfire causes. The RF algorithm is an ensemble learning method that builds multiple decision trees during training and combines their outputs to produce more accurate and robust predictions. The model was configured with the following parameters:

- n_estimators = 200, ensuring a sufficiently large number of trees to reduce variance.
- max_depth = None, allowing trees to grow fully and capture complex patterns.
- max_features = 'sqrt', selecting a subset of variables at each split to increase model diversity and reduce overfitting.
- min_samples_split = 5 and min_samples_leaf = 1, controlling how trees split and ensuring sufficient flexibility to learn from the data.
- random_state = 42, ensuring reproducibility.
- class_weight = 'balanced', giving more weight to minority classes to reduce bias toward majority categories.

To interpret the model and better understand the factors related to wildfire causes, we applied SHAP. This approach quantifies the contribution of each predictor variable to individual predictions, providing a transparent and interpretable way to identify which variables exert the strongest influence on each fire cause classification.

In addition to this global interpretation, we performed a local spatial analysis to determine which variables exert the greatest influence in specific geographic regions. This approach makes it possible to detect how the importance of individual drivers varied across the landscape, highlighting areas where specific climatic or environmental factors played a more dominant role.

2.4.1. Dependent Variable

For this model, the dependent variable is the ignition cause. The 119,498 fires recorded in the ESFire30 dataset were classified into five categories as reported in the EGIF (see Table 1):

Table 1. Dependent variable metrics.

Type	Class in ESFire30	EGIF Code	Total Number of Fires	Total Burned Area (ha)	Maximum Patch Size (ha)	Percentile 90 (ha)	Proportion (%)
Lightning	1	100	2966	309,417	19,978	85	3
Negligence	2	210 to 399	11,809	563,429	25,675	58	10
Intentional	3	400	50,344	2,021,564	14,060	68	42.13
Unknown	4	500	54,167	2,228,235	28,765	64	45.32
Fire rekindling	5	600	212	21,341	11,773	74	0.17

2.4.2. Independent Variables

To identify relevant human-related covariates, we drew on established methodologies and findings from prior research [5,8,15–17]. We also considered regional- and European-scale studies [11,18–20], as well as insights from a review on fire occurrence modelling [1].

The selected variables represent widely recognised indicators of fire occurrence and follow the typology of factors and drivers described by [21,22]. Some variables required transformations prior to integration into the model, whereas others were used in their original form. Table 2 summarises all selected variables. The transformation procedures are described below:

- Road density: To assess spatial accessibility in relation to wildfires, we generated secondary road density maps using a kernel density approach applied to linear vector data. Secondary roads were selected because they are closely associated with fire suppression access and human presence, both of which influence ignition risk [1]. The road network was converted into regularly spaced points (every 10 m) and rasterised at 30 m resolution, with each cell representing road density values. This continuous variable serves as a proxy for spatial accessibility: lower values correspond to remote or difficult to access areas, while higher values indicate easier access and greater human presence. Because historical road data are limited, we used a single road layer (https://centrodedescargas.cnig.es/CentroDescargas/redes-transporte, last accessed on 1 January 2026) supported by evidence that rural and secondary road networks in Spain have changed little in the past three decades [23]. For each burned area perimeter, we extracted the mean road density value to ensure that the metric reflects average accessibility conditions and is not biased by polygon size.
- Interfaces: Interfaces are spatial zones representing areas of contact or interaction between forest or wildland vegetation and other land-use categories. They delineate the proportion of the landscape where human activities or different land-cover types meet flammable natural vegetation, creating conditions that may influence both the likelihood of ignition and the behaviour of wildfires.

In this study, interfaces were derived from the SIOSE (https://centrodedescargas.cnig.es/CentroDescargas/siose, last accessed on 1 January 2026) and Corine Land Cover (https://centrodedescargas.cnig.es/CentroDescargas/corine-land-cover, last accessed on 1 January 2026) datasets for their corresponding years. A 250 m buffer was applied around forest and wildland classes and overlapping areas with the corresponding land-use class of interest (e.g., urban and agricultural) were identified. The 250 m buffer distance was selected to capture the immediate spatial context surrounding for each class while maintaining consistency with the spatial resolution of the burned area dataset (30 m). Similar distances have been used in studies analysing wildland–urban interfaces, as they represent the local zone where interactions between vegetation and anthropogenic land uses can influence ignition probability and fire dynamics [24,25]. The resulting variables quantify the total interface area (ha) within each burned area perimeter, representing the cumulative area of contact that burned. Specifically, five interface variables were generated to characterise different types of land-cover contact: the forest–agricultural interface (FAI), wildland–urban interface (WUI), wildland–agricultural interface (WAI), wildland–forest interface (WFI), and forest–urban Interface (FUI). This approach produces continuous variables, which improve the performance of RF models in representing gradients of human presence, land-use interaction, and potential ignition pressure. In general, interfaces capture the spatial coupling between vegetation and human-related land covers, functioning as indicators of socio-ecological processes.

- Rural areas (1985–1999 and 2000–2021): From 1985 to 1999, rural areas were derived from the Global Human Settlement Population Layer (GHS: https://human-settlement.emergency.copernicus.eu/ghs_pop.php, last accessed on 1 January 2026). The 1985 layer was assigned to 1985–1989, the 1990 layer to 1990–1995, and the 1995 layer to 1995–1999, following the five-year structure of the dataset [26]. The GHS dataset, developed by the Copernicus Emergency Management Service, provides

globally consistent gridded estimates of residential population density by combining census data with satellite-derived built-up area information. From 2000 to 2021 rural areas were obtained using the WorldPop unconstrained population density dataset at a 1 km spatial resolution (https://www.worldpop.org/, last accessed on 1 January 2026) [27]. WorldPop redistributes census counts using an RF-based dasymetric modelling approach to generate high-precision gridded population estimates. For 2021, the layer 2020 was used.

Following Law 45/2007 on Sustainable Rural Development, grid cells were classified as rural when population density ranged between 1 and 100 inhabitants per km^2, consistent with the definition of rural municipalities <30,000 inhabitants and <100 inhabitants/km^2 [28]. This method delineates sparsely populated areas based on actual population distribution rather than administrative boundaries, thereby improving spatial accuracy. Urban areas were assigned a value of zero—not to indicate absence of population, but to represent absence of rural population specifically. For each burned area perimeter, rural population values were summed across all intersecting pixels.

- Land-use change: Land-use change was derived from the Land Use, Land-Use Change and Forestry (LULUCF) database: https://www.eea.europa.eu/en/analysis/indicators/greenhouse-gas-emissions-from-land, (last accessed on 1 January 2026) which spans 1970–2021 [29]. We extracted pixels representing land abandonment, defined as land-cover transitions to wildlands or forests. Since LULUCF data are organised in multi-year intervals, each burned area perimeter was assigned to the interval containing its fire year to ensure temporal consistency. The total area of land-use change within each burned area perimeter was then calculated.
- Grazing: To account for livestock pressure, we employed the 2015 global livestock-density layer expressed in heads/km^2 (https://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1236449/, last accessed on 1 January 2026). Although this dataset represents a single year, it provides a suitable proxy for the entire study period given the lack of comparable high-resolution historical spatial datasets. As highlighted by Serrano-Zulueta, Gómez-Sal [30], traditional pastoral systems in Spain have undergone a sustained decline since the mid-20th century, driven by rural depopulation, structural shifts in livestock production, and policies favouring intensification. Similarly, Acebes, Iglesias-González [31] emphasise that grazing systems—particularly extensive and transhumant practices—are currently facing collapse in many regions. Given this long-term downward trend, using the 2015 livestock density layer likely represents an upper bound or stabilised snapshot of grazing pressure, offering a conservative and consistent estimate across the study period. For each burned area perimeter, we extracted the mean livestock density value.
- Climatic variables: Climatic variables were obtained from gridded meteorological datasets provided by the Spanish Meteorological Agency (AEMET) and other operational sources used in the Spanish meteorological drought monitoring system. This system is based on quality-controlled station observation and spatial interpolation at approximately 1.1 km resolution and provides long-term climatic information for Spain (https://monitordesequia.csic.es/historico, last accessed on 1 January 2026) [32]. The variables used in this study were potential evapotranspiration (eto), relative humidity (hr), incoming solar radiation (in_), precipitation (pr), dew point temperature (td), maximum temperature (tmax), minimum temperature (tmin), vapour pressure deficit (vpd), and wind speed (w). For each burned area perimeter, summer mean values (June, July, and August) were calculated for each climatic variable.

- Standardised Precipitation–Evapotranspiration Index (SPEI): To characterise summer climatic wetness and dryness in the months preceding each fire event, we computed the 3-month Standardised Precipitation–Evapotranspiration Index (SPEI-3). SPEI-3 was calculated for each fire perimeter using the climatic variables described above (precipitation and reference evapotranspiration), following the standard formulation based on the monthly climatic water balance and fitted to a log-logistic distribution as proposed by Vicente-Serrano, Beguería [33]. For each fire perimeter and year from 1985 to 2021, we extracted the August SPEI-3 value, reflecting drought conditions accumulated over June, July, and August—the main period of fuel drying in Mediterranean climates. This metric captures seasonal moisture anomalies relevant to fire occurrence and provides a standardised measure of short-term climatic dryness or wetness prior to ignition.
- Normalised Difference Vegetation Index (NDVI): Seasonal vegetation conditions were derived from the VIIB NDVI database, which provides biweekly observation at a 1.1 km resolution for Peninsular Spain from 1981 to the present [34]. For each burned area perimeter, NDVI values were first extracted for all 1044 biweekly dates available between July 1981 and December 2024 using area-weighted averaging, producing a continuous NDVI time series representative of mean vegetation conditions within each polygon. These biweekly values were then aggregated by calendar year to obtain seasonal means for spring (MAM: March–May) and summer (JJA: June–August), typically comprising six observations per season.

To account for interannual variability relative to historical norms, the resulting seasonal NDVI series were standardised following procedures commonly used for indices such as SPI and SPEI. For each perimeter and year, the variable included in the model was ndvi_mam, capturing vegetation conditions during the key growing season preceding and shaping summer fuel flammability.

Table 2. Biophysical and human explanatory variables.

Variable	Source	Temporal Extent	Resolution
Population density 1985–1999	[26]	1985; 1990; 1995	1 km
Population density 2000–2021	[27,35]	2000–2020	1 km
Secondary Roads density	[36]	2020–2025	Minimum mappable unit 5 m
Forest–agricultural interface (FAI)	[37,38]	SIOSE: 2005, 2009, 2011, 2014. CORINE: 1990, 2000, 2006, 2012, 2018	SIOSE: Minimum mappable unit 1 ha CORINE: Minimum mappable unit 25 ha
Wildland–urban interface (WUI)	[37,38]	SIOSE: 2005, 2009, 2011, 2014. CORINE: 1990, 2000, 2006, 2012, 2018	SIOSE: Minimum mappable unit 1 ha CORINE: Minimum mappable unit 25 ha
Wildland–agricultural interface (WAI)	[37,38]	SIOSE: 2005, 2009, 2011, 2014. CORINE: 1990, 2000, 2006, 2012, 2018	SIOSE: Minimum mappable unit 1 ha CORINE: Minimum mappable unit 25 ha
Wildland–forest interface (WFI)	[37,38]	SIOSE: 2005, 2009, 2011, 2014. CORINE: 1990, 2000, 2006, 2012, 2018	SIOSE: Minimum mappable unit 1 ha CORINE: Minimum mappable unit 25 ha

Table 2. *Cont.*

Variable	Source	Temporal Extent	Resolution
Forest–urban interface (FUI)	[37,38]	SIOSE: 2005, 2009, 2011, 2014. CORINE: 1990, 2000, 2006, 2012, 2018	SIOSE: Minimum mappable unit 1 ha CORINE: Minimum mappable unit 25 ha
Rural areas 2000–2021	[27,35]	2000–2020	1 km
Rural areas 1985–1999	[26]	1985; 1990; 1995	1 km
Land change	[29]	1985–2021	25 m
Grazing (goats, sheep and cattle /km^2)	[39]	2015	1 km
NDVI March, April, and May (MAM)	[40]	1985–2021	1 km
Climatic variables Jun, July, and August (JJA) eto: evapotranspiration (potential) hr: relative humidity in_: incoming solar radiation pr: precipitation td: dew point temperature tmax: maximum temperature tmin: minimum temperature vpd: vapor pressure deficit w: wind speed	[32]	1985–2021	1 km
Spei 3_august and (Evapotranspiration Standardised Precipitation)	computed as in [33] using de SPEI R package	1985–2021	1 km

A correlation matrix was then calculated for all predictor variables to determine whether any were collinear and thus providing redundant information. The analysis revealed clear patterns of association among climatic variables, given that related atmospheric and environmental processes influence many of them. As a result, variables representing different aspects of the climate system tended to vary together in a consistent manner. Importantly, even if some variables were correlated, retaining them in the model was still valuable for classifying wildfires causes, as each variable may capture unique aspects of environmental conditions and contribute to improving the model's accuracy and robustness.

3. Results

3.1. Spatial Distribution and Proportion of Wildfires by Causes

Figure 2a (EGIF) and 2b (ESFire30) show that the proportion of fires attributed to natural causes and fire rekindling remains relatively stable throughout the study period. However, several notable differences emerge between the two datasets. In ESFire30 (Figure 2b), the category of unknown causes dominates in nearly all years, followed by intentional and negligence-related fires, which show moderate fluctuations over time. In contrast, EGIF (Figure 2a) displays a lower proportion of unknown causes and a higher relative share of intentional fires during most of the period. Negligence-related fires appear comparatively more stable and less variable in EGIF than in ESFire30. Interannual variability is also more pronounced in the ESFire30 series, particularly for the intentional and unknown categories, whereas EGIF exhibits smoother year-to-year changes.

Figure 2c shows the distribution of fire causes in ESFire30 after the classification. The proportion of the different cause categories become more similar to those observed in EGIF. This is especially visible in 2020 and 2021. In these years, ESFire30 originally had a very large share of fires recorded as "unknown", mainly due to missing information. The

classification led to a distribution of fire causes in ESFire30 that was closer to the baseline established by EGIF.

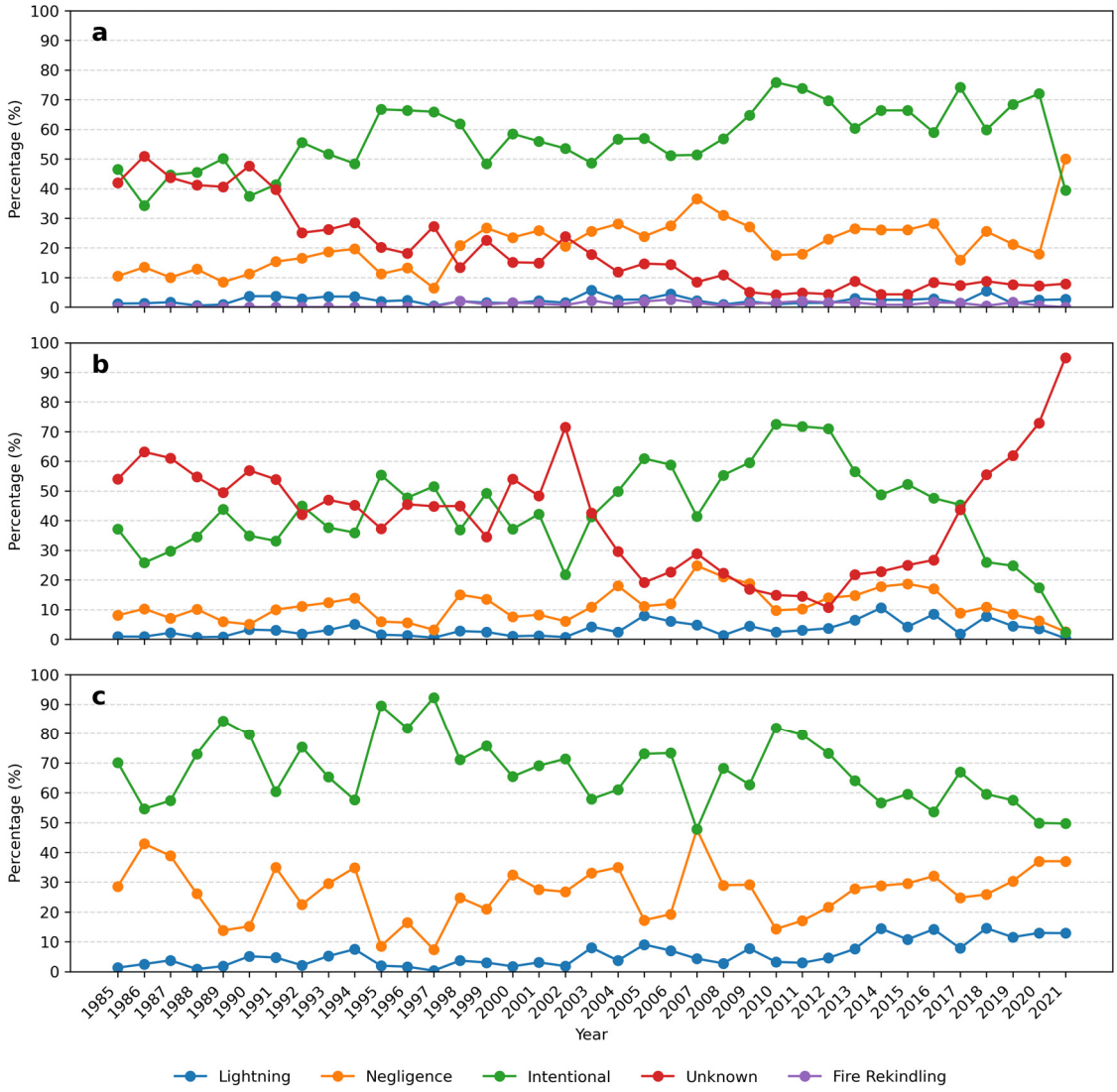


Figure 2. Annual percentage of wildfires by cause from 1985 to 2021: (**a**) shows EGIF data, (**b**) shows ESFire30 data, and (**c**) shows ESFire30 after classification.

The spatial distribution of wildfire causes reveals clear and contrasting patterns across Peninsular Spain (Figure 3). The maps display the density of wildfires by cause at a 30 m resolution, corresponding to the native resolution of the ESFire30 dataset (Ochoa et al. 2026: under review). The delineated regions correspond to the pyro-regions defined by Trigo, Sousa [41], who applied a cluster analysis to normalised burned area data from 66 administrative units across mainland Spain to identify regions with similar fire regimes. This analysis yielded four ecologically coherent clusters—northwest (NW), north (N), southwest (SW), and east (E)—which reflect distinct patterns in topography, vegetation cover and climatic conditions.

Human-caused wildfires (Figure 3a) are predominantly concentrated in the NW pyro-region, particularly along the border with Portugal. A notable density is also observed in the N pyro-region, with lower concentrations in the SW and E regions. In contrast, natural cause wildfires are primarily located within the eastern pyro-region. Figure 3b shows that wildfires of natural origin tend to cluster in the E region. Figure 3c displays wildfires of

unknown cause, which are widely distributed across the territory, with particularly high densities in the NW and N pyro-regions and smaller clusters in the SW and E regions.

Given the high proportion of wildfires with unidentified causes, we completed the attribution of causes using the RF model that was previously described, inferring the most likely causes of each fire perimeter from the auxiliary variables of the model.

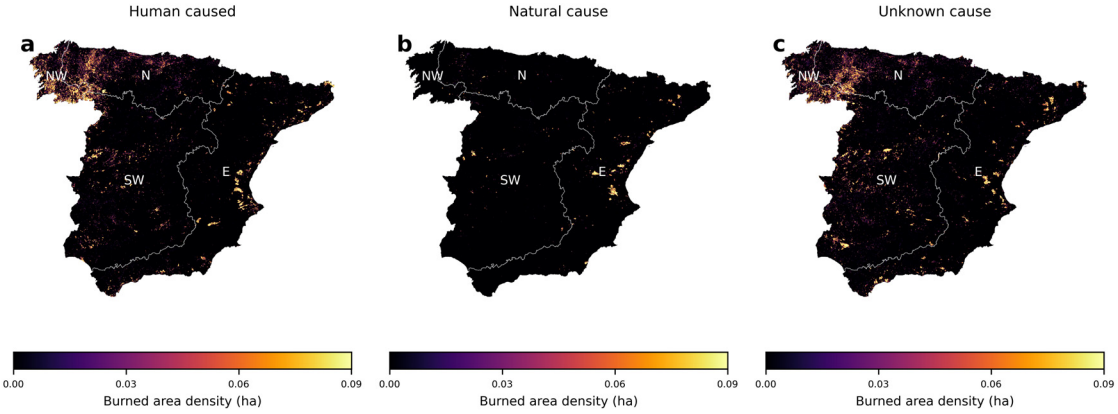
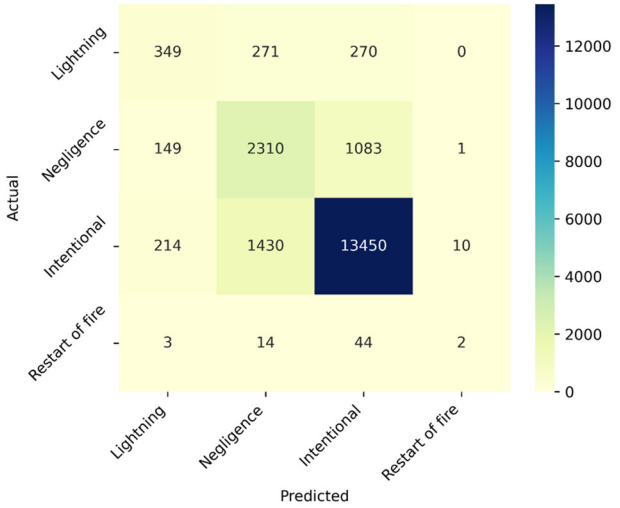


Figure 3. Wildfire causes from 1985 to 2021 as presented by ESFire30.

3.2. Model Performance

The performance metrics of the RF estimations (Figure 4) summarises the classification results across the four fire cause categories. The confusion matrix shows that the intentional class has the highest number of correctly predicted cases, with relatively few misclassifications. In contrast, the negligence and lightning classes exhibit greater dispersion, with a noticeable number of instances misclassified as intentional. The fire rekindling category contains very few observations overall and shows substantial overlap with other categories, indicating limited correct identification.



Class	Precision	Recall	F1-Score
Lightning	0.49	0.39	0.43
Negligence	0.57	0.65	0.61
Intentional	0.91	0.89	0.90
Restart of fire	0.15	0.03	0.05
Accuracy global			0.82
F1-score global			0.82
Recall global			0.82
Precision global			0.82

Figure 4. Confusion matrix and performance parameters per class.

The intentional class achieves the highest values across all metrics, with a precision of 0.91, recall of 0.89 and F1-score of 0.90. Negligence shows consistent values (precision = 0.57, recall = 0.65, and F1-score = 0.61), while lightning records lower but still consistent values (precision = 0.49, recall = 0.39, and F1-score = 0.43). Fire rekindling presents very low precision (0.15) and recall (0.03), resulting in a minimal F1-score (0.05). Overall, the model achieves a global accuracy and F1-score of 0.82, indicating strong general

predictive performance. However, performance varies notably among classes, with the model showing great difficulty in identifying minority classes despite the application of class balancing techniques.

3.3. Variable Importance

Based on the SHAP analysis, distinct environmental and anthropogenic patterns were identified for each fire cause (Figure 5). High evapotranspiration, high precipitation, and large burned areas, combined with low NDVI, low population density, and low grazing activity, characterised lightning-caused fires. These fires were also associated with medium maximum temperatures, medium wind speeds, a strong presence of the WFI, greater distances to roads, and minimal land-use change.

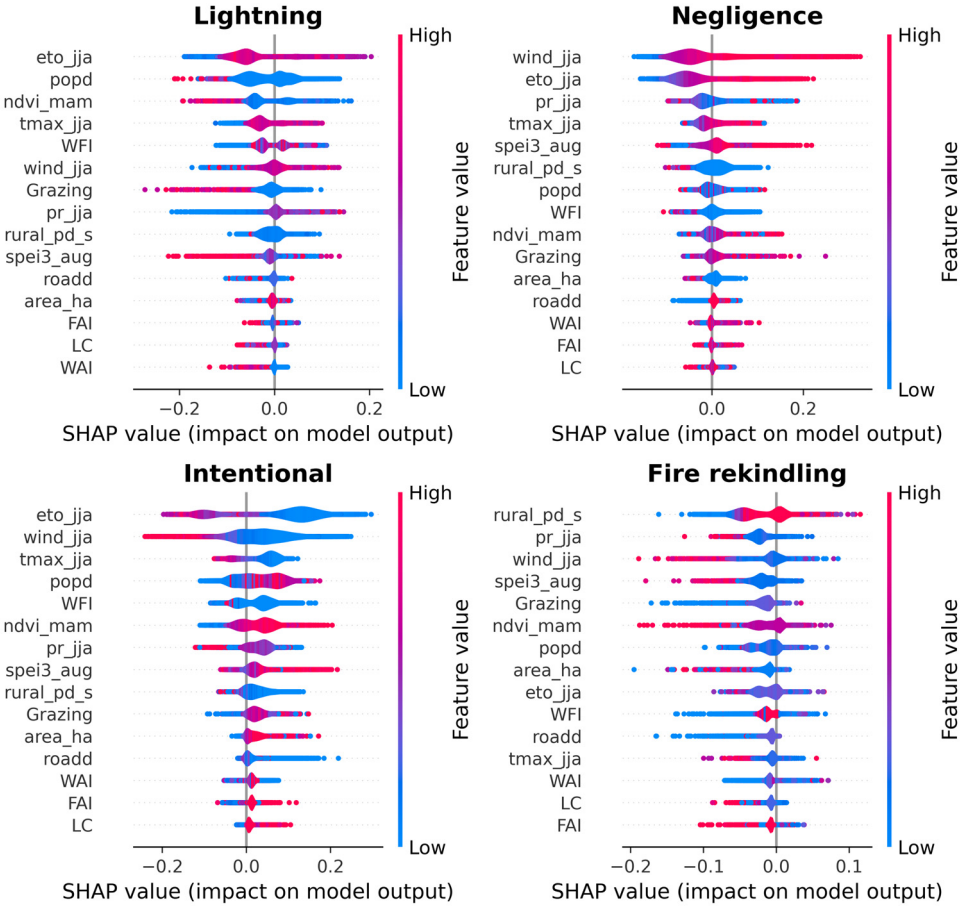


Figure 5. SHAP summary plots showing the importance and effect of predictor variables for each ignition cause. The colour of each dot represents the feature value and its position along the row represents the SHAP value for a given feature and instance. Positive values (right of the centred line) indicate a higher probability of belonging to a specific class, while negative values (left of the centred line) indicate a lower probability of belonging to a specific class.

Negligence-related fires exhibited a markedly different profile, with high wind speeds, high evapotranspiration, high maximum temperatures, and high SPEI values, coupled with low precipitation, low rural population, and low population density. These events showed high NDVI and grazing values, greater distances to roads, and were linked to high WAI and FAI, but low land-use change.

Intentional fires contrasted sharply with negligence-related ones. They occurred under low evapotranspiration, low wind speeds, and low maximum temperatures, yet in areas with high population density and high NDVI. These fires were associated with high

precipitation and large burned areas, high FAI, low WAI, and high land-change values, but with low rural population, high grazing activity, and high SPEI.

Finally, fire rekindling events were characterised by high rural population and high grazing activity, but low precipitation, low wind speeds, low evapotranspiration, and low SPEI. These events generally occurred in areas with medium NDVI, low population density, and low values across all interface and land-change metrics.

Overall, the most notable contrast emerged between intentional and negligence-related fires: while both are linked to human-related factors, negligence fires occurred under hotter, drier, and windier conditions with high evapotranspiration and high SPEI, whereas intentional fires were associated with cooler, wetter environments and denser populations. This distinction highlights different underlying mechanisms and contextual drivers behind these two anthropogenic fire causes.

An additional observation concerns the role of the burned area across causes. Lightning and intentional fires were associated with large burned areas, indicating that these causes tend to result in more extensive events. In contrast, negligence and fire rekindling showed comparatively low burned area values.

3.4. SHAP's Local Variable Importance

In the eastern regions, lighting-caused fires were primarily influenced by spring NDVI, precipitation, and population density (Figure 6a). The southwestern region showed a similar pattern, with NDVI and precipitation emerging as the dominant predictors, indicating that vegetation greenness and moisture availability are key drivers of natural fire occurrence in this area. In contrast, the N and the NW regions exhibited a more complex structure; NDVI, precipitation, and population density remained important, but additional variables—such as wind speed and maximum temperature—also contributed meaningfully to the predictions. This suggests that a broader combination of environmental factors governs lightning-caused fire occurrence in these regions.

The analysis of intentional fire causes (Figure 6b) across different regions revealed distinct patterns of influence. In the N and NW regions, intentional fires were primarily associated with evapotranspiration, population density, and wind speed, with additional but smaller contributions from land-use variables (interfaces, grazing and land-use change), NDVI and rural population. In the E region, evapotranspiration, population density, land-use factors, and precipitation emerged as the most relevant predictors. In the SW region, evapotranspiration and wind speed were the dominant factors, while population density, land-use variables, and burned area also showed secondary but notable contributions.

Negligence-related fires (Figure 6c) exhibited region-specific patterns of influence. In the central area, where the NW, N, and SW regions intersect, these fires were primarily associated with population density, land-use variables, rural population, and climatic factors. In contrast, across the broader NW, SW, N and eastern regions, evapotranspiration, wind speed and spring NDVI were the most influential variables, emphasising the role of environmental and vegetation conditions in shaping fire occurrence in these areas.

The regional aggregation of SHAP contributions (Figure 6d) reveals that evapotranspiration and wind speed emerge as the dominant predictors across the study area, reflecting their overarching control on fuel moisture and atmospheric conditions. However, notable regional distinctions persist: in the N, NW, and SW regions, population density contributes substantially to the model's output, whereas in the E region precipitation accounts for the largest number of instances. Importantly, this aggregated representation is strongly shaped by the predominance of intentional fire records in the dataset, which drives much of the overall signal.

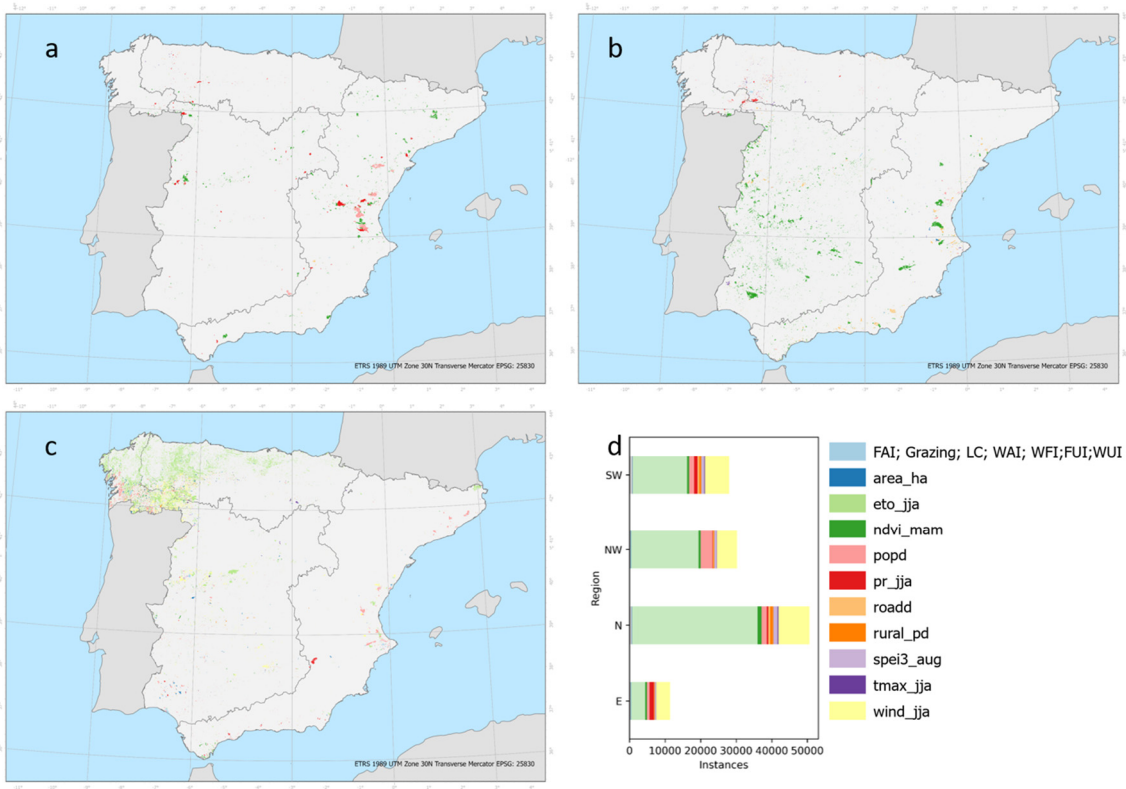


Figure 6. SHAP's local importance for each fire cause category: (**a**) natural, (**b**) negligence, and (**c**) intentional. Panel (**d**) shows the regional aggregated SHAP contribution.

4. Discussion

In this study, we developed a spatially harmonised wildfire database for mainland Spain by integrating ignition records from EGIF with fire perimeters provided by ESFire30, enabling a more consistent attribution of fire causes and a clearer assessment of their spatial distribution. Using this consolidated dataset, we applied an RF classifier to infer the origin of fires lacking an identified cause and to evaluate the environmental and anthropogenic factors most strongly associated with each ignition type. Finally, we employed SHAP-based explainability techniques at both global and local scales to disentangle the differing mechanisms underpinning fire causes, thereby providing an analytical and spatial framework that enhances our understanding of wildfire ignition patterns across the region.

A key methodological distinction concerns how the potential ignition locations of EGIF fires are defined. Previous spatialisation approaches [16] refined ignition areas by intersecting the sheet and grid system within municipal boundaries and forested perimeters, which was appropriate for studies focused specifically on forest-related fire dynamics. In our case, however, representing all fire occurrences is essential, as the complementary dataset used (ESFire30) does not distinguish fires by land-use category. For this reason, our spatialisation procedure avoids restricting ignition locations solely to forested areas and instead retains the full spectrum of burnable environments recorded in EGIF, including those classified as "Other land covers". By masking only clearly non-burnable surfaces—such as urban areas, water bodies, and a surrounding buffer—we ensure that the spatialised EGIF dataset remains compatible with ESFire30 and captures the complete set of fire events relevant to our analysis.

Despite applying SMOTE to balance the training data, the model still learned the dominant classes more effectively, reflecting the underlying class structure resulting from our spatialisation and integration procedures. This is evident in the performance metrics,

where minority categories, such as lightning-caused fires and fire rekindling, show notably lower predictive accuracy. Moreover, the underlying class imbalance can bias the SHAP analysis, as the model assigns stronger explanatory weight to features associated with dominant classes, potentially underrepresenting factors relevant to minor categories. Compared to Bountzouklis, Fox [2]—where lightning ignitions exhibit the highest classification performance—the contrast is striking. In that study, lightning-caused fires were easier to identify and showed very low confusion with other categories, largely because natural ignitions follow a distinct profile that is well captured by reliable field observations. Our results therefore highlight the importance of using input data with higher certainty and better-defined causal labels. Accurate and well-documented field information is essential for training robust models capable of recognising ignition causes effectively, particularly for classes that are naturally less frequent in national datasets. It is worth noting that some categories, such as "unknown" or rekindled fires, are inherently less reliable due to reporting limitations, which can influence both classification performance and interpretability analyses. These uncertainties in label assignment mean that the model's predictions and the subsequent SHAP- based interpretations may be more reliable for well-documented classes than for those with higher reporting uncertainty. Despite these uncertainties, we see that the overall performance of our model is high and even exceeds that reported by Bountzouklis, Fox [2].

For wildfire risk prevention, predicting human-caused fires is challenging due to the complexity of social behaviour, yet research consistently identifies a set of robust predictors that shape their spatial patterns [1]. Accessibility and human presence are among the strongest drivers: proximity to roads, railways, forest tracks, and human–vegetation interfaces (FAI, WUI, and WAI) reliably increases ignition likelihood. Socio-economic conditions such as population density, rural abandonment, unemployment, and ageing population also modulate fire occurrence by shaping how people use and move through the landscape. Climate acts as an enabling factor, with drought and dry weather conditions amplifying the probability that human activities result in ignitions [3,11]. Together, these factors highlight a complex interplay between people, land-use, and environmental conditions, providing essential context for interpreting the patterns observed in our own results.

Building on these established drivers, our SHAP analysis reveals clear differences in the conditions under which the main anthropogenic fire causes arise. Fires attributed to negligence align closely with the combination of enabling factors described in the literature: dry and windy weather, high evapotranspiration, and agricultural activity near flammable vegetation. These conditions suggest that only a minimal ignition source is needed for a fire to start and spread. Intentional fires, by contrast, tend to occur under cooler and wetter environmental conditions, indicating that they do not rely on favourable fire weather but are instead initiated independently of climatic triggers. This divergence in climatic context provides a meaningful temporal dimension for interpreting fire causes and aligns with the idea that human-caused ignitions respond differently to environmental constraints. A further distinction is reflected in burned area patterns: intentional fires are more often associated with larger fire events, whereas negligence-related fires tend to remain smaller. Hantson, Andela [42] suggests that human-caused ignitions can generate more extreme fire behaviour when they coincide—either unintentionally or deliberately—with periods of heightened fire weather conditions, allowing fires to scape initial suppression efforts. While the analysis by Hantson, Andela [42] treated all human ignitions as a single category, our results provide a more nuanced perspective by showing that intentional and negligent ignitions are associated with different climatic contexts and consequently lead to distinct fire outcomes.

Previous research has shown that lightning-caused fires tend to occur in remote areas with limited human presence, where suppression is less immediate, and their occurrence is strongly modulated by weather conditions [2,11,43,44]. Conditions such as moderate fuel moisture, intermediate productivity, and topographic features that influence storm development are known to shape the spatial patterns of these events. In line with this understanding, our results indicate that lightning-caused fires are associated with landscapes of low human density and limited infrastructure, occurring predominantly in areas influenced by WFI and under intermediate wind speed and evapotranspiration conditions. These fires also tend to be linked to larger burned areas, although the contribution of this variable to the model prediction is comparatively modest, suggesting that, while lightning ignitions may grow large, their spatial signature is shaped more by biophysical context than by fire size itself. Overall, the patterns captured by our model align with the broader characterisation of lightning-driven fire regimes: ignition emerging in sparsely populated, fuel-rich environments where atmospheric conditions are sufficient—but not necessarily extreme—to enable fire ignitions and spread.

While Bountzouklis, Fox [2] used ignition point coordinates as predictors to estimate overall variable importance, our analysis advances this approach by computing importance values at the observation level. This enables spatial mapping and reveals how the relevance of specific predictors varies across the study area when distinguishing among different fire types. Our results indicate that human-caused fires are concentrated primarily in the NW region, fires of natural origin occur more frequently in the E region, and fires of unknown cause are distributed across multiple regions, with a notable density in the NW—an area that also exhibits the highest overall fire incidence.

Local importance analyses further allow us to assess whether the model's predictive behaviour is spatially consistent with these patterns and provide insight into whether particular drivers exert stronger influence in some regions than in others. Despite its advantages, local interpretability analysis has seen limited application in wildfire cause modelling studies. Ochoa, Bar-Massada [11], who conducted a comparable analysis at the European scale, reported that climatic variables play a substantial role in shaping fire occurrence patterns in the Mediterranean. As shown in Figure 6d, the aggregated representation is strongly shaped by the predominance of intentional fire records in the dataset, which drives much of the overall signal. Consequently, finer spatial nuances characterising individual fire cause categories become partially obscured when observations are pooled at the regional scale, blurring the distinct environmental and socio-ecological mechanisms that operate within each fire cause. This illustrates the need for disaggregated or cause-specific analyses to fully resolve the spatial variability underpinning fire occurrence.

Interpreting causality patterns across datasets requires careful consideration of the methodological choices and data limitations that shape fire statistics. Differences observed between ESFire30 and EGIF become more evident once fire causes from EGIF are transferred to the ESFire30 dataset. Several methodological and data-related factors may explain these discrepancies. First, ESFire30 consistently reports a higher total number of fires than EGIF (Ochoa et al., 2026: under review), which inevitably results in a larger proportional of events classified as unknown. Second, EGIF shows a marked decline in data availability after 2016, potentially affecting the relative distribution of causes in recent years. Finally, the majority-based approach used to assign causes to the fire perimeters tends to emphasise dominant categories—most notably intentional and unknown—while reducing the representation of less frequent causes such as negligence or natural ignitions.

Overall, our findings demonstrate the importance of integrating detailed causal information, robust spatialisation procedures, and local interpretability tools to better understand the mechanisms driving different types of fire ignitions. By aligning model outputs

with well-established fire drivers and highlighting areas where uncertainties or data limitations persist, this work contributes to a more refined characterisation of fire causation across Peninsular Spain. The combination of global and local interpretability provides a more complete view of how climatic and human factors interact spatially to shape fire patterns. Together these methodological steps improve our ability to distinguish among ignition causes and provide a robust, spatially explicit basis for developing evidence-based fire risk management approaches.

5. Conclusions

This study demonstrates that integrating ignition information from EGIF with satellite-derived fire perimeters from ESFire30, combined with machine learning classification, provides an effective approach for exploring wildfire ignition causes in Peninsular Spain.

Although the approach involves several sources of uncertainty—especially for minority classes and years with limited data—the model produces spatial and environmental patterns that are consistent with the patterns shown in the EGIF. The use of SHAP values, both globally and locally, provides additional context for interpreting how different drivers relate to each ignition type, while also highlighting areas where data limitations remain important. Overall, the framework presented here is not intended to deliver causal explanations; rather, it provides a structure and spatially explicit way to integrate heterogeneous wildfire information and reduce uncertainties associated with fire cause attribution. By aligning the ESFire30 dataset with the EGIF reference distribution, improving the representation of previously unknown causes, and linking fire occurrences to distinct environmental and anthropogenic drivers, this approach reveals consistent regional patterns and contrasts between fire cause types—particularly between intentional and negligence-related fires, and between human and natural ignitions. The spatial and regional SHAP analyses further highlight how the relative importance of climatic, vegetation, and socio-territorial variables shifts across pyro-regions, while also showing the dominant influence of evapotranspiration and wind speed at the broader scale. In this way, the framework helps organise and interpret complex fire cause dynamics, contributing to a more coherent understanding of wildfire drivers in mainland Spain, and supporting a more targeted prevention and management planning.

In addition, this study builds upon and improves the ESFire30 dataset by integrating all relevant environmental and socio-economic attributes discussed in this paper, including land-use, road, grazing, and interface data. The dataset, available as shp (SHP), now links detailed burned area perimeters with these explanatory variables, enabling a more comprehensive analysis of fire patterns and causes. This enhanced dataset provides a valuable resource for developing robust models to identify predominant fire causes and supporting future research on fire dynamics and risk assessment in the region.

Author Contributions: Conceptualization, C.O., M.F., M.R. and E.C.; Methodology, C.O., M.F. and M.R.; Software, C.O.; Formal analysis, C.O.; Investigation, C.O., M.F. and E.C.; Data curation, C.O. and M.F.; Writing—original draft, C.O.; Writing—review & editing, C.O., M.F., M.R. and E.C.; Supervision, E.C.; Project administration, E.C.; Funding acquisition, E.C. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is funded by the H2020 Project FirEUrisk (Grant Agreement Number: 101003890). The FirEUrisk project contributions reported in this paper were carried out by Rodrigues, M., Ochoa, C. and Chuvieco, E. Franquesa, M. was supported by the Grant JDC2022-048710-I, funded by MCIN/AEI/10.13039/501100011033, and by the European Union NextGenerationEU/PRTR.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The complete database used in this study has been made publicly available and is accessible through the following DOI: https://doi.org/10.5281/zenodo.18449006.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Costafreda-Aumedes, S.; Comas, C.; Vega-Garcia, C. Human-caused fire occurrence modelling in perspective: A review. *Int. J. Wildland Fire* **2017**, *26*, 983. [CrossRef]
2. Bountzouklis, C.; Fox, D.M.; Bernardino, E.D. Predicting wildfire ignition causes in Southern France using eXplainable Artificial Intelligence (XAI) methods. *Environ. Res. Lett.* **2023**, *18*, 044038. [CrossRef]
3. Jones, M.W.; Abatzoglou, J.T.; Veraverbeke, S.; Andela, N.; Lasslop, G.; Forkel, M.; Smith, A.J.; Burton, C.; Betts, R.A.; van der Werf, G.R. Global and regional trends and drivers of fire under climate change. *Rev. Geophys.* **2022**, *60*, e2020RG000726. [CrossRef]
4. Martinez, J.; Vega-Garcia, C.; Chuvieco, E. Human-caused wildfire risk rating for prevention planning in Spain. *J. Environ. Manag.* **2009**, *90*, 1241–1252. [CrossRef]
5. Rodrigues, M.; Jiménez-Ruano, A.; Peña-Angulo, D.; de la Riva, J. A comprehensive spatial-temporal analysis of driving factors of human-caused wildfires in Spain using Geographically Weighted Logistic Regression. *J. Environ. Manag.* **2018**, *225*, 177–192. [CrossRef]
6. Pineda, N.; Rigo, T. The rainfall factor in lightning-ignited wildfires in Catalonia. *Agric. For. Meteorol.* **2017**, *239*, 249–263. [CrossRef]
7. Chicas, S.D.; Nielsen, J.Ø. Who are the actors and what are the factors that are used in models to map forest fire susceptibility? A systematic review. *Nat. Hazards* **2022**, *114*, 2417–2434. [CrossRef]
8. Rodrigues, M.; la Riva, J.D. Assessing the effect on fire risk modeling of the uncertainty in the location and cause of forest fires. In *Advances in Forest Fire Research*; International Association of Wildland Fire (IAWF): Missoula, MT, USA, 2014; pp. 1061–1072.
9. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
10. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A review of machine learning applications in wildfire science and management. *Environ. Rev.* **2020**, *28*, 478–505. [CrossRef]
11. Ochoa, C.; Bar-Massada, A.; Chuvieco, E. A European-scale analysis reveals the complex roles of anthropogenic and climatic factors in driving the initiation of large wildfires. *Sci. Total Environ.* **2024**, *917*, 170443. [CrossRef] [PubMed]
12. Ochoa, C.; Khairoun, A.; Coello, J.A.R.; Chuvieco, E. ESFire30. 2026. Available online: https://zenodo.org/records/18451006 (accessed on 1 January 2026).
13. Quan, Y.; Zhong, X.; Feng, W.; Chan, J.C.-W.; Li, Q.; Xing, M. SMOTE-Based Weighted Deep Rotation Forest for the Imbalanced Hyperspectral Data Classification. *Remote Sens.* **2021**, *13*, 464. [CrossRef]
14. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
15. Rodrigues, M.; Jiménez, A.; de la Riva, J. Analysis of recent spatial–temporal evolution of human driving factors of wildfires in Spain. *Nat. Hazards* **2016**, *84*, 2049–2070. [CrossRef]
16. Rodrigues, M.; de la Riva, J. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environ. Model. Softw.* **2014**, *57*, 192–201. [CrossRef]
17. Chuvieco, E.; Aguado, I.; Jurdao, S.; Pettinari, M.L.; Yebra, M.; Salas, J.; Hantson, S.; de la Riva, J.; Ibarra, P.; Rodrigues, M. Integrating geospatial information into fire risk assessment. *Int. J. Wildland Fire* **2014**, *23*, 606–619. [CrossRef]
18. Nunes, A.; Lourenço, L.; Meira, A.C. Exploring spatial patterns and drivers of forest fires in Portugal (1980–2014). *Sci. Total Environ.* **2016**, *573*, 1190–1202. [CrossRef]
19. Nunes, A.N. Regional variability and driving forces behind forest fires in Portugal an overview of the last three decades (1980–2009). *Appl. Geogr.* **2012**, *34*, 576–586. [CrossRef]
20. Padilla, M.; Vega-García, C. On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human-caused fire occurrences in Spain. *Int. J. Wildland Fire* **2011**, *20*, 46–58. [CrossRef]
21. Leone, V.; Lovreglio, R.; Martín, M.P.; Martínez, J.; Vilar, L. Human factors of fire occurrence in the Mediterranean. In *Earth Observation of Wildland Fires in Mediterranean Ecosystems*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 149–170.
22. Leone, V.; Koutsias, N.; Martínez, J.; Vega-García, C.; Allgöwer, B.; Lovreglio, R. The human factor in fire danger assessment. In *Wildland Fire Danger Estimation and Mapping: The Role of Remote Sensing Data*; World Scientific: Singapore, 2003; pp. 143–196.
23. Escolano, C.L.; Campos, Á.P.; Vidal, R.P.; Logroño, M.P.A. Valoración y representación cartográfica de la accesibilidad viaria en la España peninsular: 1960–2014. *GeoFocus* **2016**, *18*, 169–189.
24. Bar-Massada, A.; Stewart, S.I.; Hammer, R.B.; Mockrin, M.H.; Radeloff, V.C. Using structure locations as a basis for mapping the wildland urban interface. *J. Environ. Manag.* **2013**, *128*, 540–547. [CrossRef] [PubMed]

25. Bar-Massada, A. A comparative analysis of two major approaches for mapping the wildland-urban interface: A case study in California. *Land* **2021**, *10*, 679. [CrossRef]
26. Pesaresi, M.; Schiavina, M.; Politis, P.; Freire, S.; Krasnodębska, K.; Uhl, J.H.; Carioli, A.; Corbane, C.; Dijkstra, L.; Florio, P. Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data. *Int. J. Digit. Earth* **2024**, *17*, 2390454. [CrossRef]
27. WoldPop. *Population Density—WoldPop Project Categories*; University of Southampton, WorldPop—School of Geography and Environmental Science: Southampton, UK, 2025.
28. Ministerio de Agricultura, Pesca y Alimentación. *Demografía en la Población Rural en España in Análisis y Prospectiva 2020*; Ministerio de Agricultura, Pesca y Alimentación: Madrid, Spain, 2020.
29. European Environment Agency. *Net Greenhouse Gas Emissions of the LULUCF Sector*; European Environment Agency: Copenhagen, Denmark, 2025.
30. Serrano-Zulueta, R.; Gómez-Sal, A.; Pauné, F.; Velado-Alonso, E.; Garzón, J.; del Prado, A.; Herrera, P.M.; Majadas, J.; Pasetti, F.; Prada-Llorente, E.; et al. A Classification of Pastoralism in Spain: Understanding the Past to Address Present Challenges. *Nomadic Peoples* **2024**, *28*, 242–274. [CrossRef]
31. Acebes, P.; Iglesias-González, Z.; Muñoz-Galvez, F.J. Do Traditional Livestock Systems Fit into Contemporary Landscapes? Integrating Social Perceptions and Values on Landscape Change. *Agriculture* **2021**, *11*, 1107. [CrossRef]
32. Vicente-Serrano, S.M.; Tomas-Burguera, M.; Beguería, S.; Reig, F.; Latorre, B.; Peña-Gallardo, M.; Luna, M.Y.; Morata, A.; González-Hidalgo, J.C. A high resolution dataset of drought indices for Spain. *Data* **2017**, *2*, 22. [CrossRef]
33. Vicente-Serrano, S.M.; Beguería, S.; López-Moreno, J.I. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *J. Clim.* **2010**, *23*, 1696–1718. [CrossRef]
34. Franquesa, M.; Reig-Gracia, F.; Serrano, S.M.V. *Vegetation Indices for the Iberian Peninsula and Balearic Islands (VIIB) Database*; Universidad de Zaragoza: Zaragoza, Spain, 2024.
35. Stevens, F.R.; Gaughan, A.E.; Linard, C.; Tatem, A.J. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE* **2015**, *10*, e0107042. [CrossRef]
36. IGN. *Redes de Transporte del CNIG*; Instituto Geográfico Nacional (IGN): Madrid, Spain, 2024.
37. Centro Nacional de Información Geográfica (CNIG). *Sistema de Información sobre la Ocupación del Suelo en España (SIOSE): Series 2005–2020*; Ministerio de Transportes y Movilidad Sostenible, Gobierno de España: Madrid, Spain, 2020.
38. European Environment Agency (EEA). *CORINE Land Cover (CLC) Dataset Series: 1990, 2000, 2006, 2012, 2018*; European Environment Agency: Copenhagen, Denmark, 2018.
39. Gilbert, M.; Nicolas, G.; Cinardi, G.; Van Boeckel, T.P.; Vanwambeke, S.O.; Wint, G.R.W.; Robinson, T.P. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci. Data* **2018**, *5*, 180227. [CrossRef]
40. Franquesa, M.; Reig, F.; Arretxea, M.; Adell-Michavila, M.; Halifa-Marín, A.; Vilas, D.; Beguería, S.; Vicente-Serrano, S.M. Near-real-time vegetation monitoring and historical database (1981–present) for the Iberian Peninsula and the Balearic Islands. *Earth Syst. Sci. Data* **2025**, *17*, 5885–5902. [CrossRef]
41. Trigo, R.M.; Sousa, P.M.; Pereira, M.G.; Rasilla, D.; Gouveia, C.M. Modelling wildfire activity in Iberia with different atmospheric circulation weather types. *Int. J. Clim.* **2016**, *36*, 2761–2778. [CrossRef]
42. Hantson, S.; Andela, N.; Goulden, M.L.; Randerson, J.T. Human-ignited fires result in more extreme fire behavior and ecosystem impacts. *Nat. Commun.* **2022**, *13*, 2717. [CrossRef] [PubMed]
43. Rodrigues, M.; Jiménez-Ruano, A.; Gelabert, P.J.; de Dios, V.R.; Torres, L.; Ribalaygua, J.; Vega-García, C. Modelling the daily probability of lightning-caused ignition in the Iberian Peninsula. *Int. J. Wildland Fire* **2023**, *32*, 351–362. [CrossRef]
44. Johnson, E.; Miyanishi, K.; Bridge, S. Wildfire regime in the boreal forest and the idea of suppression and fuel buildup. *Conserv. Biol.* **2001**, *15*, 1554–1557. [CrossRef]