



**HAL**  
open science

# Assessing spatiotemporal forest fire dynamics in southern France: A geospatial and machine learning approach

Christos Bountzouklis

► **To cite this version:**

Christos Bountzouklis. Assessing spatiotemporal forest fire dynamics in southern France: A geospatial and machine learning approach. Geography. Université Côte d'Azur, 2023. English. NNT: 2023COAZ2027. tel-04268364

**HAL Id: tel-04268364**

**<https://theses.hal.science/tel-04268364>**

Submitted on 2 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ  
CÔTE D'AZUR



ÉCOLE DOCTORALE  
SOCIÉTÉS, HUMANITÉS,  
ARTS ET LETTRES

# THÈSE DE DOCTORAT

Évaluation de la dynamique  
spatiotemporelle des incendies de  
forêt dans le sud de la France

**Christos BOUNTZOUKLIS**

ESPACE - Etude des Structures, des Processus d'Adaptation et des  
Changements de l'Espace

**Présentée en vue de l'obtention  
du grade de docteur en Géographie  
d'Université Côte d'Azur**

**Dirigée par :** Dennis Fox

**Co-encadrée par :** Elena Di Bernardino

**Soutenue le :** 5 juillet 2023

**Devant le jury, composé de :**

Etienne Cossart, Professeur des  
universités, Université Jean Moulin Lyon 3  
Lea Wittenberg, Professeure, University of  
Haifa

Marielle Jappiot, Ingénieure de Recherche,  
INRAE

Dennis Fox, Professeur des universités,  
Université Côte d'Azur

Elena Di Bernardino, Professeure des  
universités, Université Côte d'Azur



UNIVERSITÉ  
CÔTE D'AZUR



INTERDISCIPLINARY INSTITUTE  
FOR ARTIFICIAL INTELLIGENCE  
**3IA CÔTE D'AZUR**





# Évaluation de la dynamique spatiotemporelle des incendies de forêt dans le sud de la France

Une approche basée sur l'analyse géospatiale et l'apprentissage automatique

## Assessing spatiotemporal forest fire dynamics in southern France

A geospatial and machine learning approach

### **Jury :**

#### **Président du jury**

Etienne Cossart, Professeur, Université Jean Moulin Lyon 3

#### **Rapporteurs**

Lea Wittenberg, Professeure, University of Haifa

#### **Examineurs**

Marielle Jappiot, Ingénieure de Recherche, INRAE

Dennis Fox, Professeur des universités, Université Côte d'Azur

Elena Di Bernardino, Professeure des universités, Université Côte d'Azur



# **Évaluation de la dynamique des incendies de forêt dans le sud de la France : une approche basée sur l'analyse géospatiale et l'apprentissage automatique**

---

## **Résumé**

Les incendies de forêt représentent une menace importante pour les écosystèmes et le bien-être humain à l'échelle mondiale, en particulier dans la région méditerranéenne, où les étés chauds et secs favorisent l'allumage et la propagation des feux. La saison des incendies de 2022 a été la deuxième plus grave en termes de superficie brûlée et de nombre de feux en Europe depuis 2006. Les sécheresses récurrentes et l'augmentation des températures dans le contexte du changement climatique devraient entraîner une augmentation de 200% de la superficie brûlée dans la région méditerranéenne d'ici 2090.

L'analyse géospatiale est une approche puissante qui permet une compréhension approfondie des interactions, relations, tendances et modèles complexes entre les phénomènes spatiaux, en fin de compte, en informant une meilleure prise de décision et résolution de problèmes. De plus, l'apprentissage automatique (ML) s'est révélé être un outil précieux dans la science des incendies de forêt, mais son efficacité dépend de la disponibilité de vastes ensembles de données de haute qualité.

Les relations spatiales entre l'activité des incendies et les facteurs environnementaux changent avec le temps en raison du changement climatique, des modifications des biomes et des pratiques de gestion des incendies. La cartographie des schémas spatiaux et des forces motrices principales influençant la distribution des incendies fournit des informations précieuses pour les agences de protection civile, leur permettant de mieux allouer les ressources et de mettre en œuvre des mesures de prévention efficaces. Bien que la suppression des incendies soit souvent priorisée, la prévention, l'atténuation et la préparation communautaire sont des composantes essentielles de la gestion des incendies de forêt. De plus, la prévention ciblée des incendies de

forêt nécessite la compréhension et la documentation des causes des incendies de forêt, cependant, il existe un pourcentage important d'incendies de cause inconnue en Europe et en France.

Cette thèse de doctorat examine l'évolution spatiotemporelle des zones brûlées dans le sud de la France sur une période de 50 ans (1970-2019) ; elle se concentre sur les interactions avec la topographie et les types de végétation en utilisant des bases de données d'incendies. Les résultats sont divisés en deux périodes de 25 ans en raison de la mise en œuvre d'une nouvelle politique de suppression des incendies en 1994. Au cours des 25 dernières années, la superficie brûlée a considérablement diminué et la répartition géographique des incendies a changé, en particulier dans les régions où se produisent de grands incendies. Les versants orientés au sud sont devenus plus sensibles aux incendies au cours de la deuxième période, tandis que les faibles inclinaisons de pente étaient de plus en plus évitées après 1994. La majorité des zones brûlées étaient fortement associées à l'emplacement des groupements de végétation sclérophylle (maquis), qui sont très sensibles aux incendies et s'étendent avec le temps.

En outre, cette thèse présente un modèle basé sur le ML qui utilise un cadre explicatif pour prédire la cause des départs de feu dans le sud de la France en fonction des caractéristiques environnementales et anthropiques. Les résultats indiquent que la source des incendies peut être prédite avec différents niveaux de précision : les incendies naturels ont la précision la plus élevée (score F1 de 0,87) par rapport aux incendies causés par l'homme, tels que les incendies accidentels (score F1 de 0,74) et les incendies criminels (score F1 de 0,64). Les propriétés spatiotemporelles et les caractéristiques topographiques sont considérées comme les caractéristiques les plus importantes pour déterminer la classification des incendies de cause inconnue dans la zone d'étude.

En exploitant de grandes bases de données d'incendies, une analyse avancée des données géospatiales et des techniques de ML, cette recherche démontre comment les approches

géospatiales peuvent quantifier les tendances spatio-temporelles de la dynamique des incendies dans un paysage et comment les techniques de ML peuvent être efficacement utilisées pour combler les lacunes dans les causes d'allumage des incendies. Il souligne également la nécessité d'une meilleure collecte harmonisée des données, car cela constitue un élément clé qui renforcera la puissance des techniques de ML pour fournir des informations utiles pour les stratégies de gestion des terres.

---

**Mots clés :** Feux de forêt, Géomatique, Analyse géospatiale, Apprentissage automatique, Modèles spatio-temporels

---

## **Assessing spatiotemporal forest fire dynamics in southern France: a geospatial and machine learning approach**

---

### **Abstract**

Forest fires pose a significant threat to ecosystems and human welfare globally, especially in the Mediterranean region, where hot dry summers facilitate fire ignition and propagation. The fire season of 2022 was the second most severe in terms of burned area and number of fires in Europe since 2006. Recurrent droughts and higher temperatures in a context of climate change are expected to increase burned area by 200% in the Mediterranean region by 2090.

Geospatial analysis is a powerful approach that enables a deeper understanding of the complex interactions, relationships, patterns, and trends between spatial phenomena, ultimately informing better decision-making and problem-solving. In addition, Machine Learning (ML) has emerged as a valuable tool in wildfire science, but its effectiveness depends on the availability of extensive, high-quality datasets.

Spatial relationships between fire activity and environmental factors change over time due to



climate change, biome alterations, and fire management practices. Mapping spatial patterns and the primary driving forces impacting fire distribution provides valuable insights for civil protection agencies, enabling them to better allocate resources and implement effective prevention measures. Although fire suppression is often prioritized, prevention, mitigation, and community preparedness are crucial components of wildfire management. In addition, targeted wildfire prevention necessitates understanding and documenting forest fire causes, however, there is a substantial percentage of unknown-caused fires both in Europe and in France.

This Ph.D. thesis examines the spatiotemporal evolution of burned areas in southern France over a 50-year period (1970-2019); it focuses on the interactions with topography and vegetation types using fire geodatabases. Results are divided into two 25-year periods due to the implementation of a new fire suppression policy in 1994. In the last 25 years, the burned area significantly decreased, and the geographic distribution of fires changed, particularly in regions with large fires. South-facing slopes became more fire-prone in the second period, whereas low slope inclinations were increasingly avoided after 1994. The majority of burned areas were strongly associated with the location of sclerophyllous vegetation clusters (shrublands), which are highly fire-prone and expand over time.

Furthermore, this thesis presents an ML-based model that uses an explainable framework to predict the cause of fire ignitions in southern France based on environmental and anthropogenic features. Results indicate that the source of fires can be predicted with varying accuracy levels: natural fires have the highest accuracy (F1-score 0.87) compared to human-caused fires such as accidental (F1-score 0.74) and arson (F1-score 0.64). Spatiotemporal properties and topographic characteristics are deemed the most important features for determining the classification of unknown-caused fires in the study area.

By leveraging large fire geodatabases, advanced geospatial data analysis, and ML techniques, this research demonstrates how geospatial approaches can quantify spatio-temporal trends in

fire dynamics in a landscape and how ML techniques can be effectively used to fill gaps in fire ignition causes. It also underscores the need for better and more harmonized collection of data as this is a key component that will enhance the power of ML techniques in providing useful information for land management strategies.

---

**Keywords:** Forest Fires, Geomatics, Geospatial analysis, Machine learning, Spatiotemporal patterns

# Acknowledgments

It is with great pleasure that I acknowledge the invaluable contributions of all those who have supported and encouraged me throughout the last three and a half years of my Ph.D. journey.

First and foremost, I would like to express my sincere gratitude to my director, Prof. Dennis Fox, for his unwavering support, guidance, and mentorship throughout the course of my entire Ph.D. journey. His expertise and dedication have been instrumental in shaping my research and academic growth and I am grateful for the countless hours he spent offering invaluable insights and constructive feedback which have been instrumental in achieving the goals of this thesis. I have tremendous admiration and respect for your professional and personal qualities. It has been a absolute privilege and a pleasure to collaborate with you, I look forward to the opportunity to collaborate with you again on future projects.

Next, I'd like to thank my co-supervisor Prof. Elena Di Bernardino for providing me with the opportunity to explore and engage the exciting field of Artificial Intelligence, but also having the chance to create interdisciplinary collaborations through her affiliation with the 3iA Côte d'Azur - Interdisciplinary Institute of Artificial Intelligence. In addition to providing the financial support that enabled me to carry out this research, 3iA Côte d'Azur facilitated meaningful exchanges with other Ph.D. students and postdocs from different scientific backgrounds. These interactions have been crucial in broadening my understanding this field, as they allowed me to engage with individuals from different disciplines, each bringing unique perspectives and insights to the table.

I would also like to express my appreciation to the members of my thesis committee, Etienne Cossart, Lea Wittenberg and Marielle Jappiot. I am grateful for their interest in evaluating and improving my work as well as the time and effort they dedicated to this process.

I would like to extend my gratitude to the institutions that have made their data openly available, which has been vital to the completion of this work. Particularly, I would like to recognize the contribution of the National Forestry Office (Office National des Forêts, ONF) for providing the valuable fire geodatabase that was essential in conducting this study.

My sincere appreciation goes to all of my colleagues and fellow Ph.D. students at UMR ESPACE for their collaboration and intellectual stimulation throughout this journey. I am particularly grateful to David, Lise, Andrea, Iris, Lolita, Juliette and Munir for their friendship which has been invaluable emotional support. Their kindness, humor, and positive attitudes have been a constant source of inspiration and motivation, and I am proud to be part of such a dynamic and supportive group. Thank you for making my Ph.D. journey a truly memorable and rewarding experience. I wish to each of you the best of luck in your future endeavors, and I look forward to maintaining the connections we have built as we move forward in our academic careers.

Finally, I would like to express my deepest gratitude to my family, especially my parents, for their unconditional support, encouragement, and belief in my abilities throughout my academic career. Their love and support has been a great source of strength and motivation, and I dedicate this thesis to them.

## Table of Contents

Acknowledgments.....	10
General Introduction .....	14
Chapter 1 : Literature review .....	20
Part 1 : Factors affecting forest fires with a focus on Mediterranean environments .....	20
1.1 Introduction .....	20
1.2 Environmental factors.....	21
1.3 Anthropogenic factors .....	29
Part 2 : Artificial intelligence in wildfire science .....	30
1.4 Introduction .....	30
1.5 Applications of AI in wildfire science.....	32
1.6 Challenges and limitations of AI.....	33
1.7 Conclusion .....	36
Chapter 2 : Environmental Factors Affecting Wildfire Burned Area In South-Eastern France, 1970-2019 .....	37
2.1 Introduction .....	37
2.2 Fire database .....	40
2.3 Environmental variables .....	41
2.4 Fire history 1970-2019 .....	44
2.5 Methods .....	45
2.6 Results .....	49
2.7 Discussion.....	61
2.8 Conclusion .....	65
Chapter 3 : Arson fire cause prediction in Southeastern France.....	67
3.1 Introduction .....	67
3.2 Fire database .....	70
3.3 Features.....	71
3.4 Methods .....	73
3.5 Results .....	74
3.6 Conclusion .....	76
Chapter 4 : Unknown wildfire ignition cause prediction using machine learning .....	79
4.1 Introduction .....	79
4.2 Fire database .....	83
4.3 Considered features .....	85
4.4 Methods .....	87
4.5 Results .....	90
4.6 Discussion.....	96

4.7 Conclusion .....	99
General Discussion .....	101
Part 1 : Spatiotemporal changes of burned area in SE France in the context of climate change .....	101
Part 2 : The role of ML in determining fire ignition causes: data challenges and Explainable AI .....	106
Part 3 : Leveraging geospatial data and ML to advance wildfire research .....	108
General Conclusion.....	112
References.....	115
Appendix.....	131

# General Introduction

Forest fire is a common and important element of the Earth system (Bond & Keeley, 2005) that disturbs natural ecosystems and threatens human welfare and wellbeing throughout much of the globe. The Mediterranean climate is characterized by hot dry summers which favor fire ignition and propagation. Consequently, wildfires are particularly active around the Mediterranean basin, and fires in the Mediterranean climate zones are considered to have a wide range of environmental and socioeconomic impacts (Ganteaume, Camia, et al., 2013; Miller et al., 2009; San-Miguel-Ayanz, Moreno, et al., 2013).

In Europe, the fire season of 2022 was the second most severe in terms of burned area (BA) and number of fires, preceded only by the record set in 2017, since records began in 2006 (San-Miguel-Ayanz et al., 2023). In southern European countries, which maintain longer fire records, 2021 was the second highest fire season in terms of average fire size since 1986 despite having the lowest total number of fires recorded, indicating a reduction in the number of fires but an increase in their size (San-Miguel-Ayanz et al., 2022). The current situation is likely to deteriorate with climate change since dry conditions and high temperatures are increasing. Several studies (Bowman et al., 2011; Dupuy et al., 2020; Halofsky et al., 2018; Keeley & Syphard, 2016) point out that climate change is already affecting fire seasons worldwide and that fire weather is expected to be more severe in the future. If there is no adaptation in the challenges that climate change is presenting, the annual average burnt area in the Mediterranean region is predicted to increase by about 200% by 2090 in comparison to the period between 2000 to 2008 (Khabarov et al., 2016).

Forest fires burn an average of 440 000 ha each year in the Euro–Mediterranean region, and this corresponds to about 85 % of the total BA in Europe (San-Miguel-Ayanz et al., 2020). Of the five principal Euro–Mediterranean countries concerned by forest fires (Portugal, Spain,

France, Italy, Greece), France has the lowest amount of BA (San-Miguel-Ayanz et al., 2020). It also has the smallest potential burnable area, bond since only the southern Mediterranean fringe is affected by forest fires. France, Spain, Italy, and Greece all show similar trends in decreasing decadal BA in 1980–2010, and only Portugal experienced a progressive increase during this interval (San-Miguel-Ayanz et al., 2020). It should be noted that BA is generally decreasing despite increases in summer temperatures throughout the Euro–Mediterranean zone (Pokorná et al., 2018; Rodrigues et al., 2020) and this can be attributed to more efficient firefighting strategies (Fox et al., 2015; Ganteaume & Barbero, 2019; Turco et al., 2016).

Rapid response and effective resource allocation in the initial stages of a fire are key to the success of firefighting efforts which ensure fires do not escalate into large-scale disasters (Pyne et al., 1996). Indeed, the sharp decrease in terms of BA in France is observed after the implementation of a new fire suppression strategy that focuses on timely and aggressive suppression of all fire ignitions, regardless of the weather conditions (Direction de la Sécurité Civile, 1994). In addition, the change in fire policy had a notable influence on fire activity in southern France, leading to a weakened correlation between fire and weather (Ruffault & Mouillot, 2015). Despite its effectiveness, there are concerns about the sustainability of the rapid suppression strategy during extreme weather conditions when firefighting resources are thinly spread across a larger number of fire ignitions, as was the case in 2003 and 2016 (Curt & Frejaville, 2018). Considering the recent European State of the Climate (ESOTC) report by the Copernicus Climate Change Service (C3S), the concerns about the sustainability of firefighting strategies are particularly relevant, given that Europe saw its second warmest year on record in 2022, with annual temperatures 0.9°C above average. Summer temperatures beat historical records, with temperatures reaching 1.4°C above average, in addition to widespread and extensive drought conditions.



As alterations in climate, biomes, and fire management practices change spatial relationships between fire activity and environmental factors over time, it is essential to map and comprehend these trends to assess the efficacy of firefighting strategies and create appropriate policies (Bowman et al., 2017). Identifying spatial patterns and the primary driving forces that impact fire distribution offers valuable insights to fire and civil protection agencies, thereby enabling them to allocate firefighting resources appropriately and formulate effective prevention measures (Moreira et al., 2011).

Although fire suppression can be an essential aspect of firefighting, it often receives significant attention and support from politicians and the media due to its visually impressive nature and the perception of immediate action. However, there should be less emphasis on fire suppression (Tedim et al., 2022) as relying solely on it could lead to neglecting other crucial components of wildfire management, such as prevention, mitigation, and community preparedness. Prevention and mitigation strategies, including prescribed burning, vegetation management, and the development of fire-resistant infrastructure, can play a significant role in reducing wildfire risks and minimizing their impact. These proactive measures may not be as attention-grabbing as fire suppression, but they can be more cost-effective and sustainable in the long run (Fernandes & Botelho, 2003).

Another approach of ensuring targeted prevention is through understanding, documenting, and mapping the spatiotemporal patterns of forest fire causes (Rodrigues et al., 2014). Greater attention and resources must be allocated towards these, as prioritizing this aspect is crucial to ensure that attempts to modify the habits, customs, and behaviors of the human environment are accurate and effective (Oliveira et al., 2012a; Tedim et al., 2022). The importance of understanding the complex causes of forest fires as a social phenomenon is further emphasized in the context of France and the Euro-Mediterranean region, where 95% of wildfires are caused by human activity (Ganteaume, Camia, et al., 2013; Ganteaume & Jappiot, 2013). By

identifying and addressing the complex human and environmental factors contributing to these fires, targeted prevention efforts can be implemented to reduce the occurrence and impact of wildfires in these regions.

Machine Learning (ML) has emerged as a valuable tool to identify complex relationships between various factors contributing to forest fires, gaining traction in the fire community (Jain et al., 2020). The effectiveness of these algorithms, however, depends largely on the availability of extensive, high-quality datasets. Furthermore, ML models have often faced criticism for their "black box" nature, as it can be challenging to decipher the reasoning behind their predictions or the identification of specific patterns (Loyola-Gonzalez, 2019), though new approaches such as eXplainable Artificial Intelligence (XAI) are emerging to tackle this issue. This lack of transparency makes it difficult for practitioners and decision-makers to fully trust and understand the models' outputs, limiting the practical application of ML in various fields, including wildfire management. Despite these challenges, the continuous growth in size and quality of geospatial datasets and improvements in the interpretability of results will inevitably make ML algorithms even more potent tools for analyzing ignition occurrences and causes and improve management decision support (Bot & Borges, 2022).

By leveraging large fire geodatabases, advanced geospatial data analysis and ML techniques, this study aims to enhance the understanding of fire dynamics and inform effective resource allocation and wildfire management prevention strategies to mitigate the impacts on human welfare and ecosystems in southeastern France.

More specifically, this study aims to achieve the following objectives:

- i) Analyze the spatiotemporal dynamics and patterns of forest fires over a 50-year period (1970-2019), in the context of a rapid shift in fire suppression strategies in the early 1990's.

- ii) Examine fire selectivity and the spatiotemporal relationship of BA and environmental factors such as topography and vegetation type.
- iii) Develop ML-based models to predict the source of ignition for unknown causes through a combination of environmental and anthropogenic drivers of fire ignition applied on various spatial scales and fire databases.
- iv) Exploit XAI frameworks to evaluate the significance and the impact of environmental and anthropogenic factors on fire ignition causes.

## **Thesis Outline**

**Chapter 1** – Provides the theoretical background to the thesis and is divided into two parts. Part 1 gives an overview of the major environmental and anthropogenic factors that are affecting wildfires while Part 2 introduces the main aspects of artificial intelligence and describes its applications, limitations and future prospects in wildfire science.

**Chapter 2** – Identifies potential changes in spatial occurrence of fires by analyzing the spatiotemporal dynamics and patterns of forest fires in southeastern France over a 50-year period (1970-2019) and investigates the interactions between burned area and environmental variables such as topography and vegetation type. This chapter (Bountzouklis et al., 2022b) investigates these associations through the lens of fire selectivity and spatiotemporal correlations between burned area and environmental factors. The findings are analyzed and presented for two 25-year periods one prior to and one following the implementation of France's new fire suppression policy that was put into place in 1994.

**Chapter 3** – Investigates whether ML techniques can predict ignition cause for wildfires with unknown origins and more particularly distinguish between arson and non-arson. Furthermore, this chapter aims to assess the significance of various environmental and anthropogenic factors in determining the ignition source between arson and non-arson. This examination is based on

a point fire ignition dataset of high geographic accuracy but limited in terms of quantity, to conduct the analysis within a small-scale study area situated in the southeastern region of France. Finally, this chapter (published in proceedings of the IX International Conference on Forest Fire Research, Bountzouklis et al., 2022a) presents the shortcomings and prospects of the conducted analysis and utilizes these findings as feedback for the subsequent chapter.

**Chapter 4** – Presents the development of an ML-based model that can predict the ignition source of wildfires such as accidental, arson, lightning, private and professional negligence. The model described in this chapter (published in ERL, Bountzouklis et al., 2023) builds upon the outcomes presented in the previous chapter and provides enhanced functionality, accuracy, and a more comprehensive understanding of the effect that explanatory factors hold in determining the classification of each ignition cause. The analysis is based on a significantly larger database and spans over the entire Mediterranean France.

# **Chapter 1 : Literature review**

## **Part 1 : Factors affecting forest fires with a focus on Mediterranean environments**

### **1.1 Introduction**

Forest fire spatial distribution, size, and frequency are associated with several interacting factors that can be categorized into two main groups, namely (i) environmental and (ii) anthropogenic factors. Environmental factors generally include fuel characteristics (e.g., type and water content), topography (e.g., slope inclination, altitude, and aspect), and weather conditions (e.g., temperature, relative humidity, and wind speed); anthropogenic factors include the characteristics of the transitional zone between wildland vegetation and artificial areas (wildland-urban interface, WUI) like population and housing density as well as proximity to artificial features such as roads and buildings.

A wide range of methods and variables have been tested to predict the probability of fire occurrence (Mhaweji et al., 2015), but there is no universal rule stating which factors are most important since the relative level of importance can vary from one region to another and according to the scale of the study (Ganteaume & Long-Fournel, 2015; Laforteza et al., 2013; Moritz et al., 2005). Identifying spatial patterns and the main driving forces that determine fire distribution in a region can provide valuable information to fire and civil protection agencies to assist in allocating appropriate firefighting resources and in designing proper prevention actions, especially in the Mediterranean area (Moreira et al., 2011). Subsequent sections below summarize the major factors and their relationship with wildfires.

## **1.2 Environmental factors**

The following subsections discuss the environmental factors that drive forest fires. These factors are organized into three primary categories: i) topography, ii) fuel, and iii) weather conditions.

### **1.2.1 Topography**

Among the environmental characteristics, several studies have demonstrated the existence of spatial correlations of forest fire probability and burned area (BA) to topography (Dickson et al., 2006; A. N. Nunes et al., 2016; Padilla & Vega-García, 2011). Incoming solar radiation is influenced by slope aspect, which can consequently dictate fuel type, fuel moisture, and fuel density, all of which affect flammability, as noted by Holden et al. (2009). In addition, aspect influences the degree of ecological change related to fire severity (Birch et al., 2015; Estes et al., 2017; Parks et al., 2018). In the northern hemisphere, south-facing slopes receive more solar radiation during the day than north-facing slopes, and this can enhance burn severity (Alexander et al., 2006; Oliveira et al., 2014a; M. G. Pereira et al., 2014) but the trend is not systematic (Broncano & Retana, 2004). In addition to higher fire severity, other studies Mouillot et al. (2003) have demonstrated that south-facing slopes in Corsica (France) can burn more frequently than other exposures. A recent study by Eskandari et al. (2020) utilized data-mining techniques to model fire risk based on a set of influencing factors including topography. The authors found that most fires occur in south and southwest facing slopes and ranked slope exposure as the highest in terms of fire danger importance in comparison to the rest of the topographic factors (elevation, slope angle). On the north shore of the Mediterranean, south-facing slopes frequently have more housing than north-facing slopes, and this may also contribute to a greater number of ignitions (Fox et al., 2018).

Steep slopes tend to have higher rates of spread as well as increased fatality rates over flat areas (Molina-Terrén et al., 2019). Csontos & Cseresnyés (2015) observed an exponential increase

in upslope fire spread with the increase in slope inclination whereas downslope fire spread velocity was unaffected by slope angle and was similar to rates detected on flat terrain. Slope and altitude tend to be correlated but their association with fires is often conflicting. For instance, Nunes et al. (2016) studied BA and ignition density on a municipal scale in Portugal and found both were positively correlated with elevation and slope. Similarly, Elia et al. (2019) showed that the probability of fire ignition increased with elevation and slope inclination in southeastern Italy. In contrast, Eskandari et al. (2020) noted that most fires occur on steep but low elevation slopes.

Work by Calviño-Cancela et al., (2016) focused on the interactive effects of topographical features, land use / land cover (LULC) types, and WUI characteristics. The investigation was conducted in the region of Galicia, situated in the north-western area of the Iberian Peninsula, and relied upon a fire database comprised of 26,838 fires spanning from 2006 to 2011. The results revealed a proclivity towards greater risk of fire ignition at lower elevations, although this pattern was not uniform across all LULC types situated either inside or outside of the WUI areas. It is worth noting that most fires within the WUI zone displayed a similar slope steepness compared to those outside the WUI. However, due to the flatter topography of the WUI zone, areas outside the WUI appeared relatively steeper than average.

No clear pattern associated with aspect and ignition points within the WUI areas was observed, while outside the WUI the percentage of fires occurring on north-facing slopes was lower than on other slopes.

### **1.2.2 Fuel**

Fuel, along with heat and oxygen, is one of the three elements that must be combined in the right proportions to facilitate the occurrence of a fire (Finney et al., 2021). In the context of forest fires, fuel refers to dead or live organic matter and can be classified into three categories: crown (tree canopies), surface (litter, short shrubs) and ground (peat) (Keeley et al., 2011).

Several features of organic matter can influence fire behavior, including fuel flammability, type, and continuity (Keeley et al., 2011). Unlike topography, fuel is a factor that can be regulated through human intervention. Furthermore, fuel characteristics tend to vary over time: for example, as a result of seasonal fluctuations in temperature that can affect water content and plant flammability (Fares et al., 2017).

Flammability is complex and difficult to define scientifically (Gill & Moore, 1996), and according to (Anderson, 1970) it consists of four components: i) ignitibility, which indicates how quick fuel will ignite when exposed to a heat source; ii) combustibility, which refers to the capacity of biomass to burn effectively once it has been ignited; iii) sustainability, which refers to the duration of time that the combustion process can be sustained; iv) consumability, which is defined as the amount of consumed biomass by fire. The components of flammability can vary significantly depending on plant species (Dimitrakopoulos, 2001; Etlinger & Beall, 2004; Simpson et al., 2016). Flammability of the most common native Mediterranean species has been assessed in several studies through laboratory experiments (Dimitrakopoulos & Papaioannou, 2001; Ganteaume et al., 2011; White & Zipperer, 2010). Studies like (Molina et al., 2017) focused on identifying which specific species should be used for landscaping in the WUI by homeowners in southern Spain based on their flammability. Similarly, Ganteaume et al. (2013) assessed the flammability of dead plant biomass (litter) of ornamental vegetation in the WUI of south-eastern France. Despite the extensive research conducted on flammability, it is worth mentioning that experiments in laboratory conditions are unable to fully predict the flammability of fuels in real conditions since plant exposure to heat in such experiments is often not comparable to actual field conditions (Fernandes & Cruz, 2012).

Overall, there is widespread agreement in the literature that shrublands are regarded as fire prone areas at multiple scales: regional (Carmo et al., 2011; Moreno et al., 2011), national (Nunes et al., 2016; Nunes et al., 2005) and continental (Mermoz et al., 2005; Moreira et al.,



2011; Nunes et al., 2016; Oliveira et al., 2014a). The probability of large fires is greater in dense shrublands than in forested ecosystems in the Mediterranean basin (Moreira et al., 2011; Ruffault & Mouillot, 2017). According to (Mermoz et al., 2005), fire proneness of shrublands could be related to their recovery rate since shrublands can regenerate faster and favor fuel accumulation in a short time unlike forests which take longer to recover and expand. In addition, Oehler et al., (2012) point out that shrubs are considered as a low suppressing priority by fire fighters due to the low cost of restoration.

In Europe, other vegetation types, such as grasslands, are also considered to be fire prone (Oliveira et al., 2014a). Cultivated areas are the least fire prone vegetation types mainly because of their low combustibility and their geographic proximity to built-up land covers which facilitates rapid fire detection and suppression (Moreira et al., 2011). Forested areas are found to be more fire prone than cultivated areas but less than shrublands (Moreira et al., 2011). More specifically, broad-leaved forests are usually less prone to burning than coniferous species which present a greater fire hazard (Moreira et al., 2009; Oliveira et al., 2014a).

Leaf size also influences flammability: large, well-ventilated leaves have been found to be more combustible and extinguish fire faster while smaller leaves sustain combustion for longer duration (Cornwell et al., 2015; Michelaki et al., 2020). Particularly in broadleaf forests, larger leaves form open, low-density fuel beds that facilitate increased airflow and burn quickly (Page et al., 2012). Leaves with a higher surface area-to-volume ratio ignite faster because they provide a larger contact area where pyrolysis can occur (Gill & Moore, 1996).

Fuel moisture content (FMC), defined as the mass of water per unit mass of dry material, is a temporally dynamic fuel characteristic which influences many fire processes like ignition and combustion (Keane, 2015). Ganteaume et al. (2009) observed that ground fuel flammability is reduced with higher moisture contents since the time required for a fuel particle to burn is increased. FMC is mainly controlled by weather conditions such as precipitation and

evapotranspiration, but topographic conditions can similarly impact fuel moisture by forming natural firebreaks through drainage systems such as rivers, streams and other waterways (Fang et al., 2018).

### **1.2.3 Weather**

Weather, like topography, is a naturally occurring factor that cannot be influenced by human actions. It is also more dynamic than fuel characteristics since weather conditions can undergo significant changes within a matter of hours as for example temperature varies throughout the day with the highest fire risk arising during the hottest hours.

Temperature, precipitation, wind and relative humidity are major weather variables affecting forest fires (Moriondo et al., 2006) and are crucial for estimating fire weather danger and applying risk indices (Ruffault & Mouillot, 2017). Temperature and precipitation in particular can directly influence the combustion process (Verdú et al., 2012) and are associated with the number of fires in the Mediterranean region (Turco et al., 2014). These variables control the water balance and fuel moisture of a region, affecting the ignition, propagation, and spread of wildfire severity (Mueller et al., 2020). Higher temperatures can evaporate water stored in fuels faster while heating them and thus leave fuels more susceptible to a natural or anthropogenic source of ignition. On longer temporal scales, temperature also influences the abundance of available fuel in addition to moisture content (Mueller et al., 2020). Based on predictions, climate change is likely to cause temperature increases in various regions, including the Mediterranean, where a rise of 2-4 °C is expected during the next century (Resco de Dios et al., 2007). An increase in temperature in future climates is expected to increase fire activity, severity, and intensity (Fares et al., 2017).

Precipitation is a crucial variable that influences wildfire behavior in various ways (Mueller et al., 2020). As the primary source of water input to a region, it plays a significant role in controlling the moisture content of fuels such as plants, leaves, and other organic materials on

the ground. Precipitation patterns and variability, such as the frequency and intensity of rainfall events, can impact wildfire risk (Keeley & Syphard, 2016). Understanding precipitation patterns and variability is essential for predicting wildfire risks and developing effective fire management strategies (Palheiro et al., 2006). When precipitation levels are high, it contributes to increasing the moisture content in fuels, making them less susceptible to ignition and reducing the likelihood of wildfires. However, when a region experiences low precipitation or prolonged periods of drought, the moisture content in fuels decreases, making them more prone to ignition and increasing the probability of wildfires (Varela et al., 2019). Absence of adequate precipitation can lead to drier conditions and create environments more favorable to fire spread as dry fuels burn more readily and at a faster rate than moist fuels.

Wind is an important force of fire propagation that influences the rate of spread and the direction of the flame front (Beer, 1991). Furthermore, wind can transport embers or firebrands from a burning area to unburned areas, potentially igniting new fires downwind from the original fire, a phenomenon known as spotting (Martin & Hillen, 2016). This process can make it more challenging for firefighters to contain and control wildfires, especially during periods of strong or gusty winds. Indirectly, high wind speeds, lower values of relative humidity, and increasing days without precipitation can result in the increased extent and severity of wildfires (Wu et al., 2018). This is particularly relevant in Mediterranean France where recurrent dry strong winds (“Mistral”) originating in the Rhone valley affect the regional fire regime (Ruffault et al., 2017).

Relative humidity refers to the amount of water content in the air (absolute humidity) relative to the maximum amount of water vapor the air can hold at a given temperature. Relative humidity has an inverse relationship with temperature, as the latter increases the air's capacity to hold moisture also increases, thereby causing relative humidity to decrease for constant moisture content. Low relative humidity values indicate drier air which contributes to reduced

fuel moisture levels, making vegetation more prone to ignition and fires more likely to spread (Wu et al., 2018).

When surface moisture is scarce, energy from the incoming solar radiation is converted to heat while in high surface moisture the energy from incoming solar radiation is used mainly for evaporation. Overall, the ideal conditions for a fire to spread faster are dry, hot and windy weather conditions (Holsinger et al., 2016), all of which are found in the Mediterranean region. Numerous attempts have been undertaken to relate weather factors to forest fire occurrence. (Vasilakos et al., 2009) linked fire occurrence with precipitation, temperature and wind speed. Padilla and Vega-García, (2011), in an effort to evaluate the relative performance of several meteorological variables (such as min/max temperature, cloudiness, relative humidity, solar radiation etc.) in Spain, showed that relative humidity and maximum daily temperature were the most important variables. At a European scale, maximum temperature and relative humidity have also been associated with fire occurrence (Oliveira et al., 2012b). Ruffault and Mouillot (2017), in a study that focused on the French Mediterranean area, provided proof that spatial patterns of fires are linked to fire weather, and they highlighted that climate change in the near future may affect the spatial patterns of fire hazard. Weather factors are particularly important for large fires since they are the main drivers (Ganteaume & Barbero, 2019; San-Miguel-Ayanz et al., 2013) and they play a critical role in fire suppression, either by improving and facilitating firefighting or overwhelming any undertaken effort (Fernandes et al., 2016). For instance, Koutsias et al. (2012) indicated that the very large and catastrophic fires that occurred in Greece in 2007 were influenced by the synergistic effect of weather factors and fuel accumulation.

Weather variables are used for operational fire risk assessment through Fire Danger Ratings (FDR) systems. These provide a quantitative measure of fire danger by considering the interplay of weather factors such as temperature, humidity, wind speed, and precipitation with fuel moisture conditions (Dimitrakopoulos et al., 2011). Fire weather indices aim to provide a

more accurate representation of the weather conditions that are favorable for wildfires to propagate, as for example the Canadian FWI (Van Wagner, 1987). The Canadian FWI is widely regarded as the most prevalent and commonly used system globally, and it provides a quantitative measure of fire danger by considering the interplay of weather factors such as temperature, humidity, wind speed, and precipitation with fuel moisture conditions. The Canadian FWI system consists of six components: 1- the Fine Fuel Moisture Code (FFMC) which measures the moisture content of surface litter and other small fuels, 2- the Duff Moisture Code (DMC) which measures the moisture content of deeper organic layers beneath the litter, 3- the Drought Code (DC) measures the moisture content of the large woody fuels, 4- the Initial Spread Index (ISI) which measures the rate of fire spread, 5- the Buildup Index (BUI) which measures the total amount of fuel available for combustion, and 6- the Fire Weather Index (FWI) which represents the cumulative effect of the previous five components and is used as an overall indicator of fire danger (Van Wagner, 1987). While the Canadian FWI system is designed to evaluate fire danger in Canadian vegetation, Viegas et al., (1999) found that the components of the FWI system exhibited strong correlations with fire activity in southern Mediterranean countries despite the significant differences in vegetation and Mediterranean climate compared to Canada. There are numerous attempts to provide adaptations of the Canadian FWI to specific regions in the Mediterranean zone (see Dimitrakopoulos et al., 2011), as for example the I85 and I85/90 indices (Carrega, 1991) developed for Southeastern France. Due to the popularity of the FWI, the European Forest Fire Information System (EFFIS) currently uses the FWI as the official index for medium-range operational fire danger forecasts, reflecting the widespread adoption of this index (San-Miguel-Ayanz, et al., 2013).

### **1.3 Anthropogenic factors**

In Europe most forest fires are caused by humans (Ganteaume, Camia, et al., 2013) and in the European Mediterranean zone, the majority of the fires are intentional and located where artificial areas intersect or intermingle with wildlands (WUI). There is a distinct concern in terms of wildfire management in those interfaces since fires tend to be more frequent but also pose a greater threat to human lives and infrastructures (see for example Lekkas et al., 2018). WUI area is expanding due to urban development and thus increases human pressure; for example, urban areas in France increase by about 25,000 ha each year each year at the cost of agricultural and natural areas (Bouillon et al., 2020).

In recent years, numerous human-related variables have been integrated in research studies as factors that influence the spatial patterns of fire occurrence, either directly or indirectly. These variables include population density (Elia et al., 2019; Nunes et al., 2016), housing density (Ruffault & Mouillot, 2017), road density (Padilla & Vega-García, 2011; Ruffault & Mouillot, 2017), proximity to roads (Ricotta et al., 2018; Vilar et al., 2016) proximity to built-up areas (Mancini et al., 2018; Padilla & Vega-García, 2011) and proximity to railways (Vilar et al., 2016). Other studies (Calviño-Cancela et al., 2016; Ricotta et al., 2018) implicate interactions of fire occurrence with LULC types and their position relative to the WUI (within or outside the interface).

Overall, there is higher fire incidence at shorter distances from roads and houses (Mancini et al., 2018; Ricotta et al., 2018) but that is not always the case (e.g., Marques et al., 2011), particularly in the case of bush-clearing fires by shepherds. Fire ignitions are more likely to happen near roads because of negligence like cigarettes or other accidents. Likewise, higher population or housing densities in the WUI increase human pressure in forested and semi-natural areas and therefore potentially increase the number of fires due to accidents or negligence. In some studies (e.g., Nunes et al., 2016) higher population densities are associated

with higher fire occurrence while others (e.g., Elia et al., 2019) exhibit an inverse relationship. In low density areas with high vegetation continuity, conditions are favorable to fire propagation but not ignition. In high density areas with low vegetation continuity, conditions are favorable to fire ignition but not propagation. There is therefore an intermediate zone that corresponds to characteristics frequently found in the WUI where both ignition and propagation are high.

Similarly to environmental factors, human variables can vary in importance depending on the scale or the region; for example, road proximity was a robust factor in predicting human-induced fire ignition in Southern California (Santa Monica Mountains) (Syphard et al., 2008) while not being significant at the scale of the entire county (Syphard et al., 2007). The dependence of local or regional studies to accurately examine the robustness of the variables involved is therefore evident here as well.

## **Part 2 : Artificial intelligence in wildfire science**

### **1.4 Introduction**

Wildfires continue to be a major threat to communities, ecosystems, and economies worldwide. Climate change, coupled with land use changes has led to an increase in the frequency, intensity, and severity of wildfires in many regions. This has put significant pressure on the management and suppression resources of fire agencies, and it has led to an increased interest in the application of new technologies and approaches to wildfire science and management.

Artificial Intelligence (AI) is a subfield of computer science that aims to create intelligent machines/algorithms that can perform tasks that typically require human intelligence (Mitchell,

1997). Machine Learning (ML) algorithms are a subset of AI that can learn from data and improve their predictive performance over time without being explicitly programmed (Bishop & Nasrabadi, 2006). In other words, ML algorithms can automatically learn and improve from experience, which makes them well-suited for tasks such as classification, prediction, and decision-making based on data. In wildfire science, ML algorithms can be used to classify and map fuel properties, predict fire occurrence and spread rates, and model fire risk and susceptibility, etc.

Deep Learning (DL) is a type of ML that uses artificial neural networks (an algorithm inspired by the structure and function of the human brain) to recognize patterns in data (LeCun et al., 2015). In contrast to traditional ML algorithms, which are typically designed to work with structured data and require explicit feature engineering, deep learning algorithms are capable of automatically learning complex patterns and features from unstructured data, such as images, audio, and text. DL algorithms are particularly well-suited for applications in image and speech recognition, natural language processing, and computer vision (LeCun et al., 2015). In wildfire science, DL techniques can be applied to a wide range of tasks, including the use of remote sensing data to map active or final burned areas and severity, detect fires and smoke, and forecast fire weather variables.

In recent years, there has been growing interest in the use of AI, ML, DL and other advanced technologies to improve the accuracy and effectiveness of wildfire detection, prediction, and management (Bot & Borges, 2022; Jain et al., 2020). However, despite advances in technology and management practices, wildfires remain a complex and challenging problem to address. Successful wildfire management requires a multidisciplinary approach that involves a wide range of stakeholders, including fire agencies, policymakers, landowners, and local communities. In addition, effective wildfire management requires ongoing research and monitoring to better understand the complex interactions between climate, vegetation, and fire



behavior, as well as to identify new strategies and approaches to reduce the risk and impact of wildfires. By leveraging the power of AI, researchers and practitioners can more effectively predict, prevent, and respond to wildfires.

### **1.5 Applications of AI in wildfire science**

In recent years, there has been a surge of interest in the application of AI in wildfire science, with researchers exploring a wide range of potential applications. AI techniques, including ML and DL can be used to classify and map fuel properties and quantities as input variables in fire behavior models in order to improve the accuracy of fire spread and growth rate predictions (López-Serrano et al., 2016; K. L. Riley et al., 2014). ML algorithms can also be employed for the rapid detection of fire and smoke, enabling a more effective and timely firefighting response (Ba et al., 2019; Langford et al., 2018; Zhao et al., 2018). Moreover, remote sensing data can be utilized to develop models for active or final burned area and severity mapping in order to inform suppression planning and damage assessment (Ban et al., 2020; Collins et al., 2018; A. Pereira et al., 2017). ML algorithms can also be applied to forecast fire weather variables, such as wind speed and direction, which can aid in the prediction of fire occurrence and growth (Bates et al., 2017; Lagerquist et al., 2017). In addition, ML models can be developed to predict fire occurrence, spread and growth rates, as well as burned area to facilitate more effective dissemination of suppression assets (Chetehouna et al., 2015; Dutta et al., 2016; Hodges & Lattimer, 2019; Mayr et al., 2018; Vecín-Arias et al., 2016). Finally, ML techniques can be used to model and map fire risk and susceptibility by utilizing fire databases in conjunction with various environmental and anthropogenic variables to identify potential drivers of fire ignition and/or burned area (Curt et al., 2016; Ghorbanzadeh et al., 2019; Molina et al., 2019). These applications highlight the potential of ML in wildfire science for improving the accuracy and effectiveness of wildfire management and mitigation efforts. Despite the fact that ML has been utilized in various aspects of wildfire research, it has not yet been employed to predict the

ignition cause of fires. In later sections, it will be discussed how ML can be leveraged to predict fire ignition causes, potentially filling a significant gap in our understanding and enabling more targeted wildfire prevention and management strategies.

## **1.6 Challenges and limitations of AI**

While the application of AI in wildfire science has shown great potential, there are several challenges and limitations that must be considered. The following parts describe shortcomings related to data volume and accuracy but also the need for AI systems to be transparent and understandable to humans.

### **1.6.1 Limitations of data quality and quantity**

AI techniques, particularly ML and DL, depend heavily on the quantity and quality of multimodal data which are utilized to train and validate models and provide inputs for prediction and decision-making. Ensuring data quality and availability is therefore crucial for the successful implementation of AI in wildfire management and prevention.

A significant challenge in data quality pertains to remote sensing data, which includes satellite imagery, aerial photography, and drone images, but also weather data and other similar sources. These data are frequently used to develop models for active or final burned area and severity mapping, classification and mapping of fuel properties and quantities as input variables in fire behavior models (Jain et al., 2020). However, the presence of noise, errors, and biases in these data sources can compromise the accuracy of resulting models (Olofsson et al., 2014). For instance, existing global BA products do not provide adequate accuracy to satisfy the standards of modelling experts (Mouillot et al., 2014). Large remotely sensed fire databases, including FireAtlas (Andela et al., 2018) and GlobFire (Artés et al., 2019), underestimate both burned area and number of fires (Galizia et al., 2020). As reported by Franquesa et al., (2022), this results in consistently under-representing cropland fires while boreal forest fires are detected

with higher accuracy due to their large spatial area and long-duration. These constraints need to be considered when employing such datasets in the context of fire management, policy formulation, and scientific research.

Datasets often contain missing values, and these can either hinder their usability or negatively impact the performance of models. For instance, the documentation of forest fire causes can vary significantly depending on the country, available resources, and the efforts of local authorities (Tedim et al., 2022). As a result, the quality and comprehensiveness of forest fire cause documentation can be inconsistent across different national fire databases. For example, in France the official forest fire database for the Mediterranean area (Prométhée.com) is lacking information on the ignition cause of nearly 70% of all fires in the database.

The lack or complete absence of data is being addressed through protocols for data standardization (e.g., Camia et al., 2013) as this facilitates the fusion of national fire databases with historical records and enables the creation of more extensive databases with enhanced modeling potential, especially in the context of ML. Several other georeferenced data standardization initiatives in Europe, including the European INSPIRE Directive and the European Geodata Infrastructure, aim to establish a common framework for the management and dissemination of geospatial data throughout Europe and are aligned in this direction.

Another challenge is the potential for bias in data used to train and validate AI models when using historical data; this is known as concept drift. This refers to the phenomenon where the underlying relationships between input variables and the target variable in a dataset change over time (Gama et al., 2014; Krawczyk et al., 2017). This can pose a challenge for predictive models, as the assumptions made during the training phase may no longer hold true when making predictions on new, unseen data. In the context of wildfires, fire risk and susceptibility models are commonly developed using historical wildfire data, such as records of past ignition locations, ignition causes and burned area (Bot & Borges, 2022). Concept drift could manifest

as changes in factors influencing fire risk and behavior, such as shifts in land cover patterns, climate change, or changes in fire management practices. As a result, AI models trained on historical wildfire data may not accurately capture these evolving relationships, and lead to decreased predictive performance or biased results, such as by over-predicting fire occurrences in areas with a known history of wildfires, potentially overlooking other areas that may have a greater risk.

### **1.6.2 Interpretability & explainability in AI**

Interpretability and explainability are two significant challenges in the AI field, not only in wildfire science but also in various other domains. Despite lacking a consensus within the AI community regarding the precise definition of these terms (Marcinkevičs & Vogt, 2023), interpretability often refers to the ability to understand and explain the decision-making process of a model, while explainability involves providing a reason or justification for the model's output (Doshi-Velez & Kim, 2017; Lipton, 2018). While AI models can be very effective at identifying patterns in complex datasets, they are often considered "black boxes" because it can be difficult to understand how the model arrives at its predictions or how certain patterns were identified (Al-Bashiti & Naser, 2022). This lack of interpretability can be a barrier to adoption as it may be difficult for stakeholders to trust AI models without understanding how they arrive at their conclusions. In recent years, various methods have been proposed to improve the interpretability and explainability of AI models, as for instance LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (Shapley Additive Explanations) (Lundberg & Lee, 2017; Ribeiro et al., 2016). These methods are aimed at providing clear and understandable explanations thereby making it easier for users to understand how the model arrives at its predictions or classifications. Interpretability and explainability are essential for building trust in AI systems and constant efforts for improvement in that regard are needed (Bot & Borges, 2022). This will ensure that decisions made by these systems are

understandable, fair and ethical. However, it is important to note that interpretability and explainability are not always achievable, particularly for complex AI models, and there may be trade-offs between model performance and interpretability (Doshi-Velez & Kim, 2017). Overall, while AI has been influential in wildfire science, it is important to address these challenges and limitations to ensure the development of accurate and reliable AI models for wildfire prediction and mitigation.

## **1.7 Conclusion**

Wildfires are a complex and dynamic phenomenon influenced by both environmental and anthropogenic factors operating at a range of spatio-temporal scales. A thorough understanding of these factors and their interplay is crucial for effective wildfire management and prevention. As we continue to collect and analyze geographic data and fire databases, we can extract further scientific knowledge and exploit opportunities to combat wildfires. As demonstrated in the literature, advanced technologies such as AI and ML are becoming more accessible and have the potential to substantially improve our ability to predict and mitigate the effects of wildfires. While AI applications in wildfire management are still in their early stages, the potential benefits are clear: AI can help to analyze and interpret large datasets, identify complex patterns and trends, and provide early warning systems for wildfire outbreaks. However, the limitations of AI, such as lack of transparency and explainability, but also data related concerns must also be acknowledged and addressed. Consequently, in order to mitigate the devastating impacts of wildfires it is essential to continue research and development of AI applications using explainable approaches in conjunction with effective wildfire management policies.

# Chapter 2 : Environmental Factors Affecting Wildfire Burned Area In South- Eastern France, 1970-2019

Reference to published article:

*Bountzouklis, C., Fox, D. M., & di Bernardino, E. (2022). Environmental factors affecting wildfire-burned areas in southeastern France, 1970–2019. Natural Hazards and Earth System Sciences, 22(4), 1181–1200. <https://doi.org/10.5194/nhess-22-1181-2022>*

## 2.1 Introduction

Spatial relationships between fire occurrence and environmental factors evolve over time due to changes in biomes and climate, but also as the result of fire management practices. Mapping and understanding these trends are crucial for evaluating the effectiveness of fire-fighting strategies and developing suitable policies (Bowman et al., 2017). There are numerous recent efforts that aim to analyze spatial and temporal trends of fire activity at a global, national and regional level. Otón et al., (2021) analyzed global trends of BA based on the FireCCILT11 database which is the longest available global BA dataset to date (1982-2018). At a national level Catarino et al., (2020) investigated the trends of annual BA in Angola between 2001 and 2019 using MODIS products (MCD64A1) and associated the significant trends to land cover, ecological regions and protected areas. In regional scale, Ganteaume & Barbero (2019) utilized a long-term (1957-2017) fire geodatabase to analyze spatio-temporal variations of large fires in terms of frequency and BA, in the French Mediterranean. Silva et al., (2019) used a satellite derived BA dataset covering a 39-year period over the Iberian Peninsula to study BA trends and explore the relationship between areas with significant BA trends and fire danger. Urbieto et al., (2019) studied the spatio-temporal trends in Spain between 1980 to 2013 with regard to

fire frequency, BA and fire size, and their relationship with changes in climate, land-use and land-cover, and fire suppression. Viedma et al., (2018) assessed the changing role of environmental and human-related factors in reference to fire activity, in west-central Spain from 1979 to 2008.

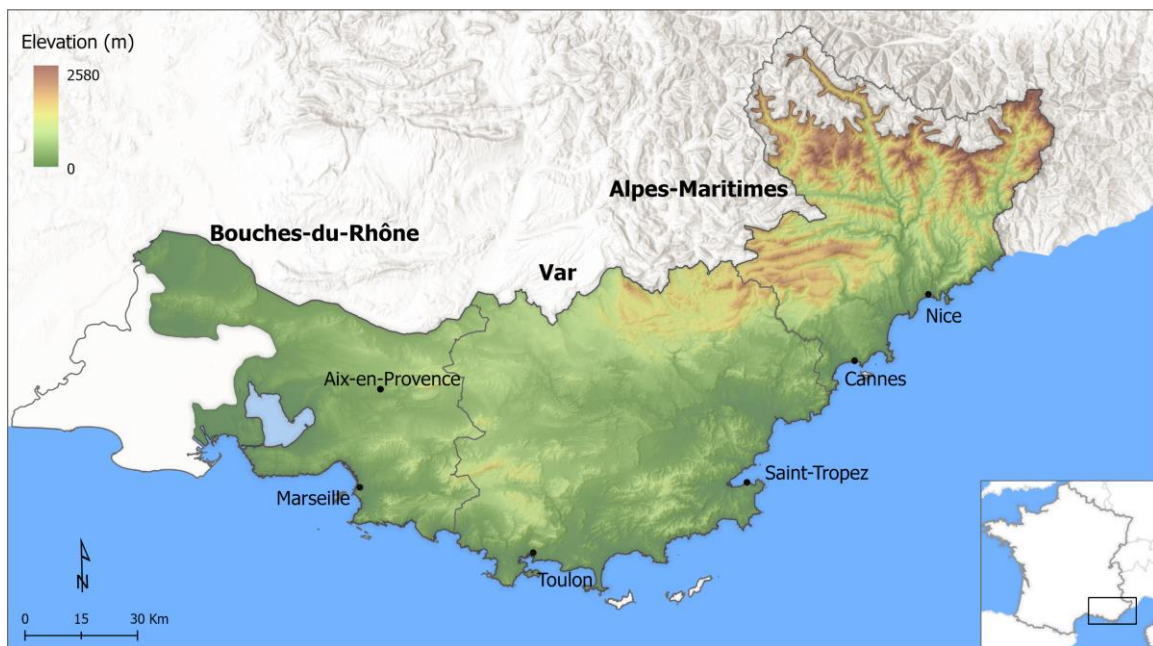
Fire suppression is an important factor that can influence fire spread. In France, as a response to the large fires that occurred between 1986 to 1990 a major change in fire suppression strategy was established in the 1990s; it focused on rapid suppression of fire ignitions regardless of the weather conditions in order to avoid fire propagation (Direction de la Sécurité Civile, 1994). The fire policy had a significant impact in fire activity in Southern France and weakened the fire-weather relationship (Ruffault & Mouillot, 2015). Despite the sharp decrease in BA after the full implementation of the fire management policy, its effectiveness on very large fires was not as successful as for smaller fires since changes in BA that correspond to large return periods are not significant (Evin et al., 2018). Although many studies have focused on determining relationships between fire behavior and driving factors (Mhawej et al., 2015), few studies have examined how fire suppression strategies impact the spatial distribution of BA. Identifying spatial patterns and the main driving forces that determine fire distribution provides useful information for fire and civil protection agencies, and it assists in allocating appropriate firefighting resources and in designing proper prevention actions (Moreira et al., 2011).

The objective of this section is to provide an overview of the quantitative changes in spatio-temporal BA patterns induced by a major shift in fire suppression strategy initiated in the early 1990s in South-eastern France. The time interval under study spans 5 decades (1970-2019) and includes the relation of BA with respect to environmental factors such as a) topography (Slope aspect and inclination) and b) Vegetation type. Although several studies have investigated the relationships between BA and environmental factors, very few have covered such a long-time

interval based on burn scar polygons, nor have they been explicitly related to changes in fire suppression methods.

### 2.1.1 Study area

The study area is comprised of a subset of the 3 administrative departments with the greatest BA in continental France (only Corsica has greater burned area) according to the French official forest fire database (promethee.com): Bouches-du-Rhône, Var, and Alpes-Maritimes (Table 1, Fig. 1). Areas within the departmental limits that were excluded represent surfaces that cannot burn such as marshlands in the westernmost part of Bouches-du-Rhône and high alpine mineral surfaces located in the northern part of Alpes-Maritimes.



**Figure 1: Map of south-eastern France showing the study area and the departmental limits overlaid on a 5 m Digital Elevation.**

**Table 1: Environmental characteristics of the study area per departmental unit**

	<b>Bouches-du-Rhône</b>	<b>Var</b>	<b>Alpes-Maritimes</b>
<b>Total area (km<sup>2</sup>)</b>	3456	6019	3495
<b>Forested area (km<sup>2</sup>)</b>	1530	4044	2727
<b>Ratio forest/total (km<sup>2</sup>)</b>	0.44	0.67	0.78
<b>Mean slope (°)</b>	8.8	11.9	24.3
<b>Median slope (°)</b>	5.7	9.6	25.2

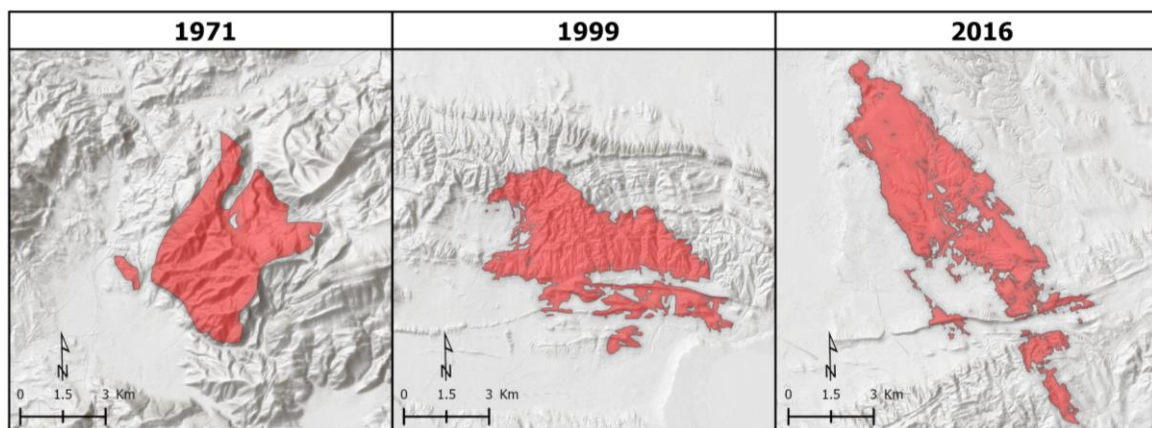


Topography varies from west to east (Fig. 1). The gentlest slopes are found in the west (Bouches-du-Rhône) and both altitude and slope inclination increase eastwards. The steepest slope inclinations are found in the northeastern part of the study area where the French Alps are located. Topography influences population distribution since much of the built area is concentrated along the coast or on shallow to intermediate slopes in the WUI. In the Bouches-du-Rhône, the western portion of the department has particularly low population densities due to the presence of the national park and wetlands mentioned above. Similarly, much of the population in the Alpes-Maritimes is concentrated in the southern portion of the department. The 2010 population densities of 388.8, 167.5, and 252.0 persons/km<sup>2</sup> for the Bouches-du-Rhône, Var, and Alpes-Maritimes, respectively, are approximative as they simply divide population by total area without accounting for geographic distributions. The order, however, is accurate and shows the greatest population density for Bouche-du-Rhône, and the lowest for the Var. Based on the demographic and environmental characteristics described above, the westernmost section (Bouche-du-Rhône) of the study area has low potential for fire ignition and propagation but increases when moving towards the eastern half of department. The central part of the study area (Var department) has a high potential for fire ignition and the greatest potential for fire propagation since it has a high forested area and a large continuous WUI area. Finally, the eastern section (Alpes-Maritimes department) has high ignition and propagation potentials in the southern portion of the department and low ignition / high propagation at higher altitudes.

## **2.2 Fire database**

Forest fire research in France is frequently based on the national database for forest fires in France ([www.promethee.com](http://www.promethee.com)) where fire location is defined as the municipality where fire ignition occurred. For this study, we used a fire Geographic Information Systems (GIS)

database provided by the National Forestry Office (Office National des Forêts, ONF) and the Delegation for the Protection of the Mediterranean Forest (Délégation à la Protection de la Forêt Méditerranéenne, DPFM). Even though the number of recorded fires is significantly lower than the Promethee database, the total area burned is almost identical; very small fires recorded in Promethee are not all digitized in the ONF database. To the best of our knowledge, this is only the second use of this geodatabase after Ganteaume & Barbero (2019). The dataset includes more than 3,000 digitized burn scar polygons for fires that occurred between 1970 and 2019. Due to the long temporal extent of the database, the accuracy and the methods used to define burn scars varied over time. In the 1970s, burn scars were mapped using field measurements with GPS devices, and the technique progressively evolved to integrate remote sensing data (satellite imagery, orthophotos). Although the description of how BA was defined is not recorded in the database, earlier polygons are clearly less accurate (coarse shapes with little detail) than burn scars after the advent of satellite imagery (Fig. 2).



*Figure 2: Evolution of digitized burn scar accuracy over the past decades.*

## 2.3 Environmental variables

### 2.3.1 Topography

Burn scar polygons were rasterized to a 5 m spatial resolution and overlain on a 5 m Digital Elevation Model (DEM) extracted from RGE-ALTI<sup>®</sup>, the official National Geographic

Institute (Institut Géographique National, IGN) database. The DEM was used to calculate Slope aspect and inclination. In the conversion of vector polygons to raster cells, BA polygons smaller than half the cell size (25 m<sup>2</sup>) were not defined as burned during rasterization, so BA for the Slope aspect and inclination analyses represent approximately 96 % of actual BA in the study area. Aspect was divided into 5 categories: Flat, North, East, South and West. Inclination was divided into 5 categories: 0°-10°, 10°-20°, 20°-30°, 30°-40° and >40°.

### 2.3.2 Vegetation type

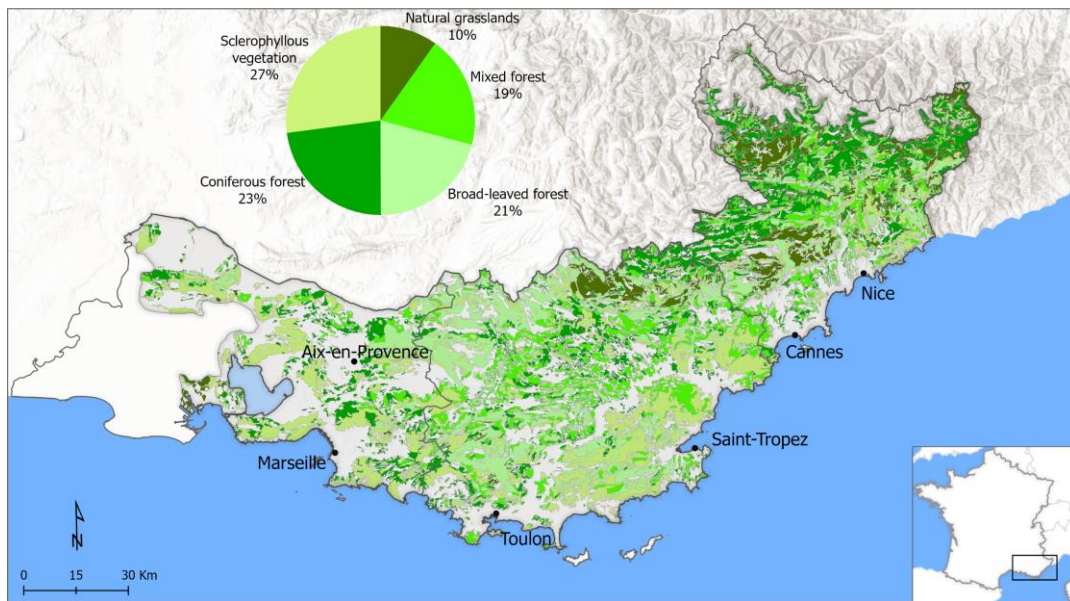
For the computation of the forested BA and the identification of fire-prone vegetation categories, GIS forest layers were extracted from the European CORINE land cover (CLC) database. The database includes five reference years 1990, 2000, 2006, 2012 and 2018. In addition to the CLC reference layers, it was considered best to backcast two additional forest cover layers for 1972 and 1980 to account for any transitions between forested and non-forested surfaces for the two decades preceding the CLC database. The methodology followed for the projection process is addressed in Subsection 2.5.1. The fire geodatabase was then matched with the CLC layer that was chronologically closest to the equivalent fire period (see Table 2).

**Table 2: Corine land cover layers and their respective fire periods.**

<b>Corine Land Cover</b>	<b>Fire period</b>
1972 (Predicted)	1970 – 1974
1980 (Predicted)	1975 – 1984
1990	1985 – 1994
2000	1995 – 2002
2006	2003 – 2009
2012	2010 – 2014
2018	2015 – 2019

The vegetation types that were used in the current study follow the CLC nomenclature: Broad-leaved forest, Coniferous forest, Mixed forest, Natural grasslands and Sclerophyllous vegetation (Fig. 3). Although Natural grasslands and Sclerophyllous vegetation are not forests,

the categories will be referred to collectively as wildland or forested areas indiscriminately for the sake of brevity.



*Figure 3: Distribution of vegetation types based on CLC 2018.*

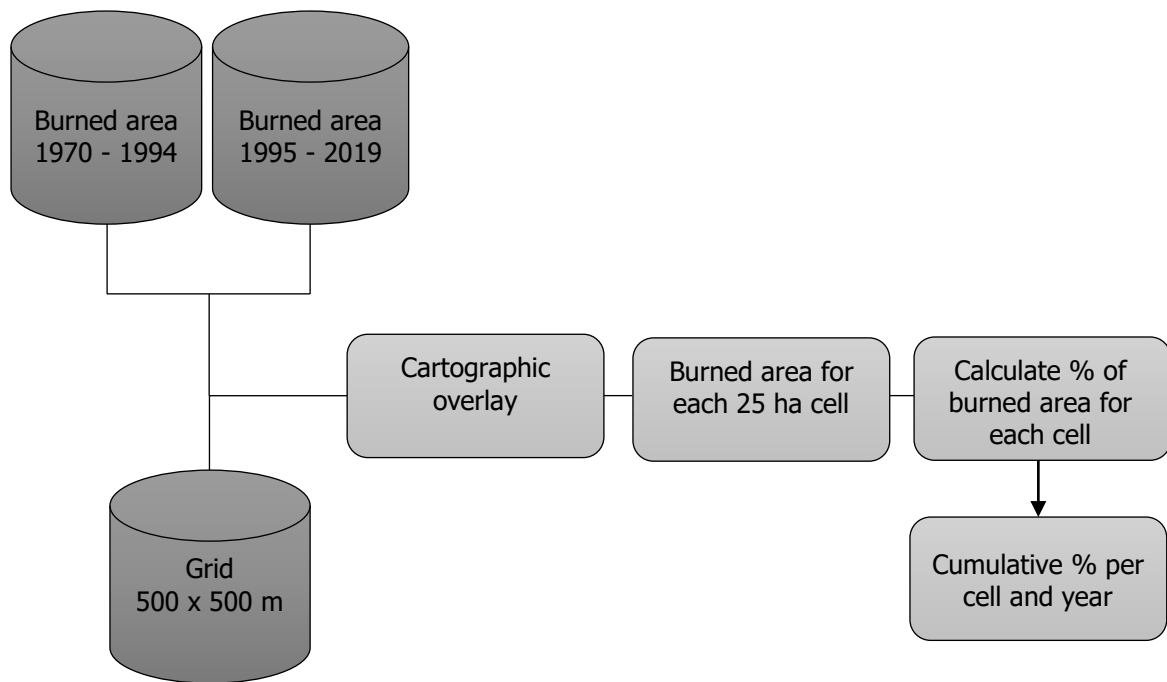
### 2.3.2.1 Forest layer projection

Although most urban growth occurred on agricultural land (Roy et al., 2015) and forest cover changed little, the Land Change Modeler (LCM) module of Terrset (Eastman 2020) was used to predict vegetation cover in 1972 and 1980. LCM is programmed to forecast change from an earlier to a later date, so going back in time (backcast) required the temporal inversion of filenames for the 1990 (renamed to 2000) and 2000 (renamed to 1990) CLC layers; in this way, land cover was simulated for 1980 and 1972. Land cover categories were simplified from the original CLC categories to the following: Built, Broad-leaved forest (Broad), Coniferous forest (Conifer), Mixed forest, Natural grasslands (Grass), Sclerophyllous vegetation (Bush), other, and water. Only transitions greater than 0.05 % of the landscape (14.3 km<sup>2</sup>) were modeled, and these included the following (From-To): Bush-Grass, Bush-Other, Built-Other, Grass-Other, Broad-Bush, Other-Grass, Bush-Conifer, Other-Bush, Bush-Broad, Bush-Mixed, Mixed-Bush, Other-Conifer, Mixed-Broad, Mixed-Other, Other-Broad, Other-Mixed, Broad-Other, Grass-Bush, Mixed-Conifer, Built-Mixed, Built-Bush, Conifer-Mixed. Note that these are the inverse

of historical trends, so the Built-Mixed transition actually backcasts the historical transition of Mixed forest to Built area. Explanatory variables used to predict land cover change were the following: Altitude, Slope inclination, Distance from Built area, Distance from Broad, Distance from Conifer, Distance from Mixed, Distance from Grass, Distance from Bush, Distance from Other and Distance from water. According to Eastman (2020), Cramer's V values of  $\geq 0.15$  for explanatory variables are useful and should be kept in the model, and all explanatory variables used here met this criterion. Accuracy rates to model transitions ranged from 65 % to 90 % with mean and median values of 78 % and 80 %, respectively.

#### **2.4 Fire history 1970-2019**

A 500x500 m grid (25 ha) was created and overlaid on the study area in order to measure the percentage of each cell that was burned each year between 1970 and 2019 (50 years) (Fig. 4). These percentage values were then summed to produce the cumulative percentage of BA for each cell. This approach facilitated the effort to identify clusters of cells/areas that have been burned multiple times and to give an overview of the spatial distribution of BA in the region. To better illustrate the impact of suppression strategies on fire occurrence, the method was applied to two 25-year subsets of the fire dataset i) 1970-1994, and ii) 1995-2019 as the mid-point break corresponds to the major shift in firefighting strategy and allocated resources in France.



**Figure 4:** Flow chart depicting the processing steps to generate the cumulative percentage of forested burned area per cell.

## 2.5 Methods

### 2.5.1 Spatio-temporal analysis – Contextual Mann-Kendall

In order to identify spatio-temporal trends within the entire time period (1970-2019), a modified version of the Mann-Kendall test was applied (Mann, 1945). The Mann-Kendall test is a non-parametric test which is used to statistically assess monotonic upward or downward trends for a variable through time. In this study we used the contextual Mann-Kendall (CMK) test which was introduced by (Neeti and Eastman, 2011), and it differs from the original test by evaluating trends at a 3x3 cell neighbourhood for each cell in a grid. The specific method has been used to assess trends in BA with satisfactory outcomes (Catarino et al., 2020; Otón et al., 2021; Silva et al., 2019). The CMK method was devised from Tobler’s First Law of Geography (Tobler, 1889) which states that “everything is related to everything else, but near things are more related than distant things.” By assuming that trends show signs of spatial autocorrelation between adjacent cells, the CMK test allows for greater confidence in identifying the presence of a trend (Neeti & Eastman, 2011). However, it requires observations

to be a set of independent random variables and thus applying the test on data that are temporally autocorrelated may lead to false rejection of the null hypothesis of no trend (Douglas et al., 2000). To assess the temporal autocorrelation in our dataset we applied the Durbin-Watson test (Durbin & Watson, 1950), and to remove it, the prewhitening procedure by Wang and Swail (Wang & Swail, 2001) which preserves the same temporal trend but without the autocorrelation (Fig. 5).



**Figure 5: Flow chart depicting the processing steps to estimate trend significance using the**

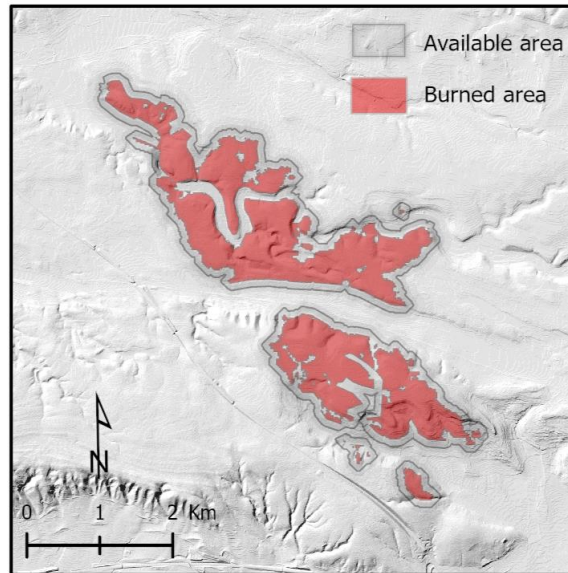
### 2.5.2 Fire Selectivity (Jacob's Index)

In order to examine the fire proneness of the environmental variables considered in this study (Slope aspect and inclination, Vegetation type) a resource selection index was calculated for each 25-year interval. Resource selection is based primarily on wildlife ecology (Manly et al., 2002), but its use has been extended to include fire selectivity (Bajocco & Ricotta, 2008; Barros & Pereira, 2014; Moreira et al., 2001, 2009; Moreno et al., 2011; M. C. S. Nunes et al., 2005; Oliveira et al., 2014a). The rationale behind fire selectivity is that fires burn selectively when the proportion of a class (e.g., type of vegetation) within a burned area is higher than the proportion of the available area to burn. The opposite applies when a specific class of variable is burned proportionally less than the available area (fire avoidance). In our work, we used Jacob's selectivity index (Jacobs, 1974) which is defined as:

$$D_i = \frac{r - p}{r + p - 2rp} \quad (1)$$

$r$  stands for the proportion of a resource class  $i$  used by fire, and  $p$  is the proportion of a resource class  $i$  available to fire. Jacobs' index values range between -1 and 1. Positive values indicate

fire preference; negative values indicate fire avoidance. The index was calculated for each class of the environmental factors (described in the subsequent sections) for each year. Similar to other studies (Barros & Pereira, 2014; M. C. S. Nunes et al., 2005), the available area for each fire to burn is defined as twice the amount of area burned by each fire. (Fig. 6).



**Figure 6: Illustration of burned area ( $r$ ) and available area ( $p$ ) to be used by a fire. The available area (the sum of the burned area + buffer zone) around each fire corresponds to twice the burned area.**

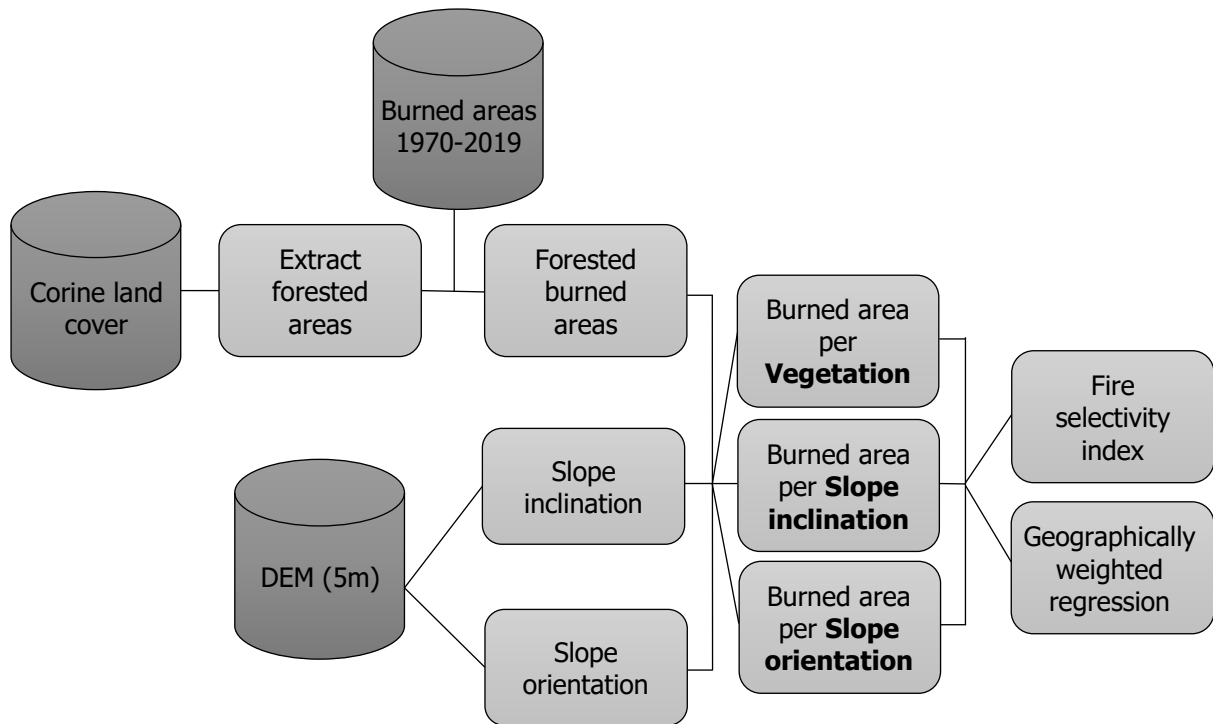
### 2.5.3 Geographically weighted regression

a Geographically weighted regression (GWR) was used to quantify the impact of the change in firefighting strategy on the relative importance of the environmental factors. GWR is applied in wide range of interdisciplinary fields including forest fires (Kolanek and Szymanowski, 2021; Koutsias et al., 2010; Martínez-Fernández et al., 2013; Nunes et al., 2016; Rodrigues et al., 2016). GWR is a local non-parametric regression method (Fotheringham et al., 2003) that allows the relationships between dependent and explanatory variables to vary over space. The basic form of a GWR model, provided by (Fotheringham et al., 1998, 2003) is defined as:

$$y_i = \beta_{i0} + \sum_{z=1}^j \beta_{iz} x_{iz} + \varepsilon_i \quad (2)$$



Where  $y_i$  is the dependent variable at location  $i$ ,  $\beta_{i0}$  is the intercept parameter at location  $i$ ,  $j$  is the number of explanatory variables,  $b_{iz}$  is the local regression coefficient for the  $z$ th explanatory variable at location  $i$ ,  $x_{iz}$  represents the  $z$ th explanatory variable at location  $i$  and  $\varepsilon_i$  denotes the random error at location  $i$ . Since GWR allows coefficients to be spatially heterogeneous, a sub-model for the location of each observation is created that considers only a subsample of the total observations, where observations in closer proximity have a greater effect in determining the local set of coefficients than observations located at further distances (Fotheringham et al. 1998). This neighbourhood is called a “kernel,” and the maximum distance from a regression point at a location  $i$  is defined as “bandwidth”. The bandwidth is an important parameter that can be defined in two different ways: i) fixed bandwidth, (fixed distance for each regression point) and ii) adaptive bandwidth (fixed number of nearest neighbours for each regression point). The first type of neighbourhood is more appropriate when data are regularly distributed across space whereas the second type is more appropriate for data that form spatial clusters. In the current work the adaptive bandwidth approach was utilized to fit the GWR model which was optimized based on the value of Akaike Information Criterion (Akaike, 1998). For each of the 3 environmental variables described above, a univariate GWR model was used to explore the relationship with the dependant variable (% of BA) for the two 25-year periods i)1970-1994 and ii)1995-2019.



*Figure 7: Flow chart depicting the processing steps and data used to relate BA to Vegetation type, Slope inclination and orientation.*

## 2.6 Results

Results presented below will first describe fire history for the entire time interval (1970-2019) and then analyse the spatio-temporal evolution of BA split according to the two 25-year periods. Finally, it will explore the relationship of BA to topography (Slope aspect and inclination) and Vegetation type. Factor-specific results will be discussed as they are presented in the following results sections while broader considerations will be explained in the Discussion section.

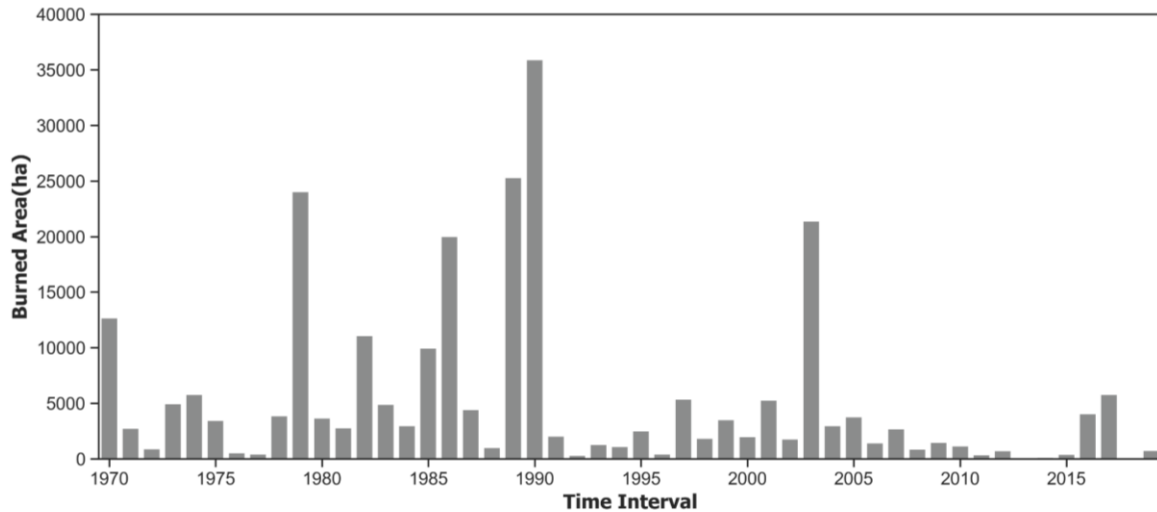
### 2.6.1 Fire history

In total, 3,382 fires burned 296,820 ha in 1970-2019. The mean and median areas of BA are 87.7 ha and 4.2 ha, respectively; these values reflect the typical positively skewed distribution of fire size where the vast majority of fires are small and a few fires, accounting for most of the burned area, are very large. The number of fires equal to or greater than 100 ha, 500 ha, and 1,000 ha is 378 (11.2%), 123 (3.6 %) and 65 (1.9 %), respectively. Of the total number of

fires, 2,424 (88.2 %) occurred in forested landscapes, and these burned an area of 263,645 ha (88.8 % of total BA).

Mean and median values for forested landscape fires are slightly greater than for all fires at 111.7 ha and 6.5 ha, respectively. The number of fires equal to or greater than 100 ha, 500 ha, and 1,000 ha is 314 (13.0 %), 106 (4.4 %), and 60 (2.5 %), respectively. As stated above, results presented below will deal exclusively with the forested BA that was occupied by one of the vegetation types mentioned in section 2.3.2 since the trends with respect to vegetation and topography for all fires and forested landscapes are nearly identical.

Annual forested BA varies significantly from year to year (Fig. 8) although there are clear differences between the first two decades (1970-1990) and the last three (1991-2019). The mean and median annual BA are 5156.4 ha and 2746.1 ha, respectively. Several big fires occurred in the 1980s followed by a sharp decrease in the early 1990s. Similar to the rest of southern Mediterranean Europe, most of the forested BA is related to a small number of large fires (Turco et al., 2016). Only 5 years (1979, 1986, 1989, 1990 and 2003) of the 50-year record account for almost half of the total forested BA (126,700 ha). The forested BA for each of these years surpasses 20,000 ha, attaining nearly 36,000 ha in 1989. Of the 5 years cited above, only 2003 is found in the second 25-year interval. As described by (Fox et al., 2015) for the Alpes-Maritimes, the decrease in BA corresponds to an improvement in fire-fighting strategy since the latter period had some of the hottest summers on record; the same explanation appears to hold for the neighboring departments studied here.

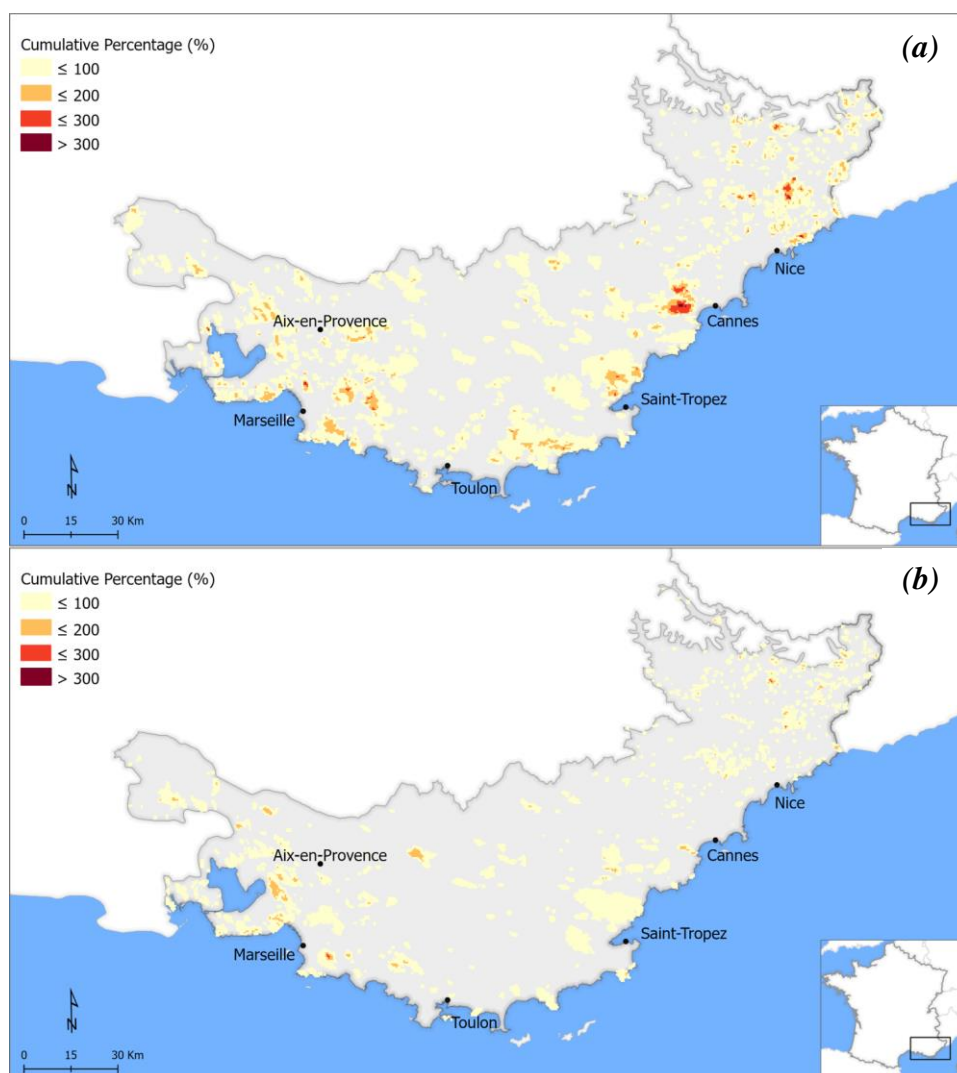


**Figure 8: History of annual forested burned area from 1970 to 2019.**

Figure 9 maps cumulative percentage area burned inside each 25 ha cell for 1970-1994 and 1995-2019, respectively. Generally, most fires occur in the WUI north of the large coastal cities since densely developed areas have too little vegetation to burn and relatively remote areas have too few ignition sources. Although we did not treat wind direction or speed, BA shapes in both periods tend to align themselves with known wind patterns in the region: they have a NW-SE orientation throughout most of the western and central sections (Bouches-du-Rhône and Var departments) but show little preferential orientation in the eastern department of Alpes-Maritimes where wind speeds are lower than the “Mistral” winds in the Rhône valley. There is a clear difference between the two periods with the second one having significantly fewer burned cells, which are also slightly more spatially dispersed. In addition, cumulative percentage values are noticeably lower with a small number of cells (302) exceeding 100 % and very few (9) reaching 200 %. All major hotspots disappear in the second interval apart from some located mainly in the western area of the study zone near Aix-Marseille.

The largest patches in both intervals are found in the central part of the study zone in the Var department which combines continuous forest cover and a lower population density that is distributed more evenly throughout the department. The two largest continuous BA clusters are found here, one north of Saint-Tropez and one east of Toulon. In the 1995–2019 time

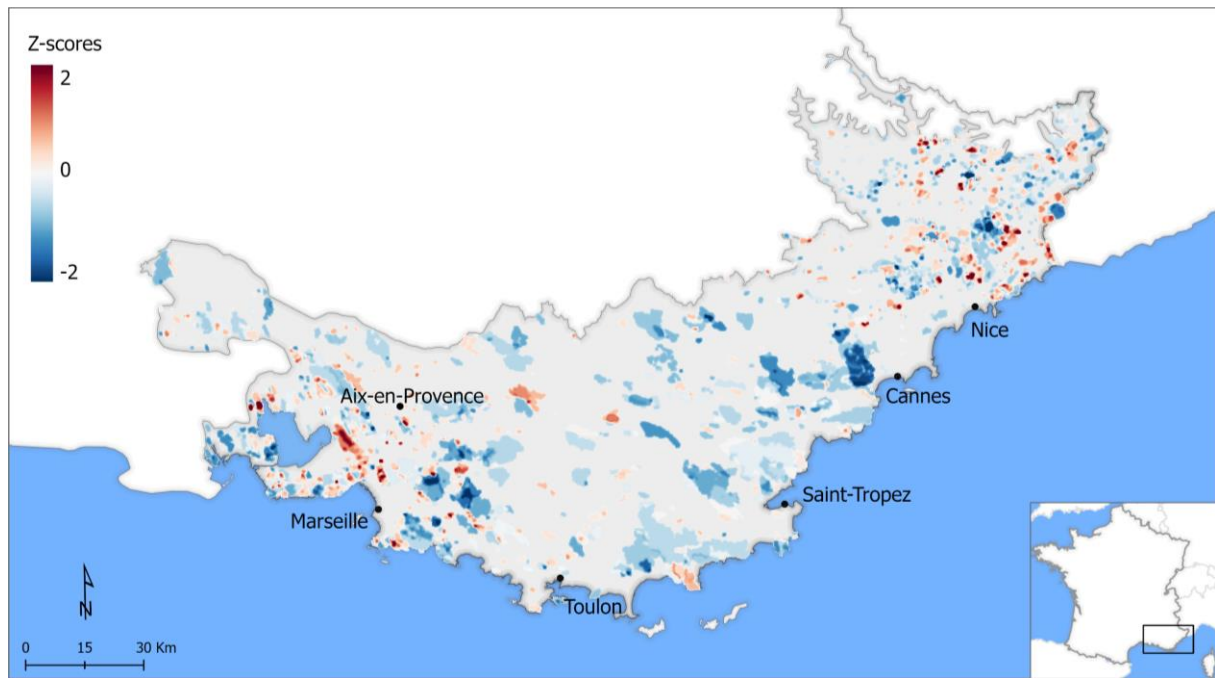
interval, the first cluster shrunk whereas the second one completely disappeared. In the western section of the study area (Bouches-du-Rhône), burned patches are located in constrained areas between densely built zones (Aix-en-Provence and Marseille) with several cells displaying high fire recurrence. In the eastern section of the study area (Alpes-Maritimes), where population is particularly dense along the coast (Cannes-Nice), BA cells are concentrated inland along the periphery of the coastal built-up area. A major hotspot with the highest cumulative percentage burned area is found just west of Cannes, and this patch almost disappears in the second period. In comparison to the rest of the study area, patches in the eastern department of the Alpes-Maritimes are smaller and more numerous with high to very high recurrence, even at higher altitudes.



*Figure 9: Cumulative percentage of forested burned areas in the 1970-1994 interval (a) and in the 1995-2019 interval (b) over a 500 x 500 m grid.*

### **2.6.2 Spatio-temporal analysis**

Results of the CMK method depict areas of increasing and decreasing trends in terms of mean annual BA over the study area (Fig. 10). Positive Z-scores (colored in red) correspond to areas with increasing trends and negative Z-scores (colored in blue) correspond to areas with decreasing trends. Overall, a general decreasing trend of BA throughout most of the study area can be observed, with approximately 60% of the cells corresponding to a negative value. The largest clusters of negative Z-scores are located predominately in the central areas of the region, north of Toulon, north of Saint-Tropez, and west of Cannes, with small negative patches north-east of Marseille and north of Nice. Positive Z-score clusters are more constrained in terms of size and are generally dispersed. Significant decreasing trends are relatively limited and can be spotted in areas such as east of Marseille, west of Cannes and north of Nice. Significant positive trends are detected in several locations (although limited in area) such as between Aix-en-Provence and Marseille and in the northeastern part (Alpes-Maritimes department) of the study area. Although contrasting negative-positive trends co-exist in close proximity near Marseille and Aix-en-Provence, the greatest speckled pattern is found in the Alpes-Maritimes department where fires are smaller and more randomly distributed.



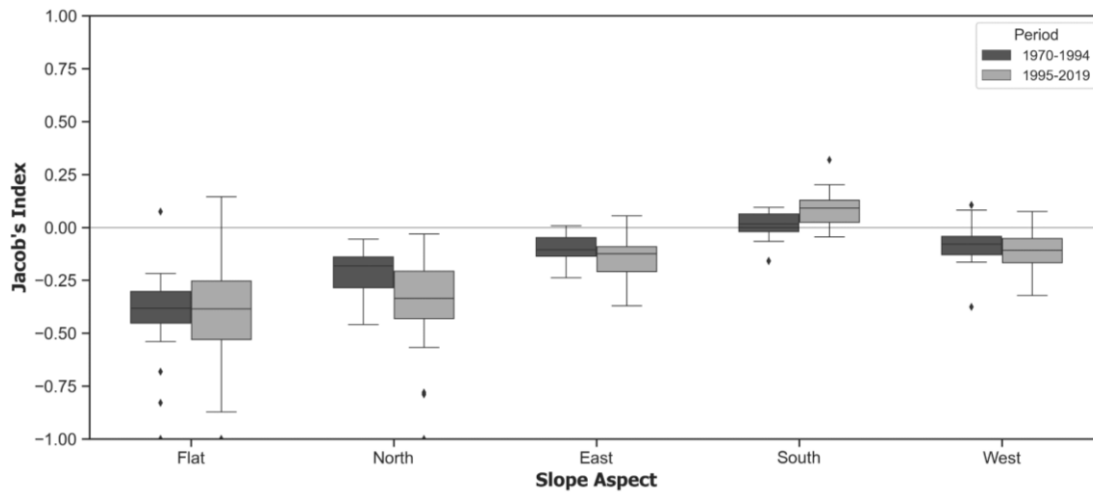
*Figure 10: Trends of mean annual burned area between 1970 to 2019 based on the Contextual Mann-Kendall method. Areas with positive Z-scores depict increasing trends of burned area, while negative Z-scores show decreasing trends.*

### 2.6.3 Fire selectivity and Topography

Topographic effects studied here include Slope aspect and inclination. Since some areas may have greater BA values simply because in a given topographic class is more frequent in the landscape, Jacob's selectivity index was calculated in order to identify potential classes of aspect and inclination that are preferred by fire between two periods: i) 1970-1994 and ii) 1995-2019.

Figure 11 shows fire preference (Jacobs' index  $>0$ ) and fire avoidance (Jacobs' index  $<0$ ) for the two 25-year periods under study. Between 1970-1994, S-facing slopes have a weak positive median value (0.02) while the others are all negative. Values become increasingly negative in the following order: W (-0.08), E (-0.12), N (-0.18) and flat (-0.38). In the second period (1995-2019), the median fire selectivity of S-facing slopes (0.1) increases and presents a clear difference with other trends which either remain the same (flat) or decrease. N-facing (-0.33)

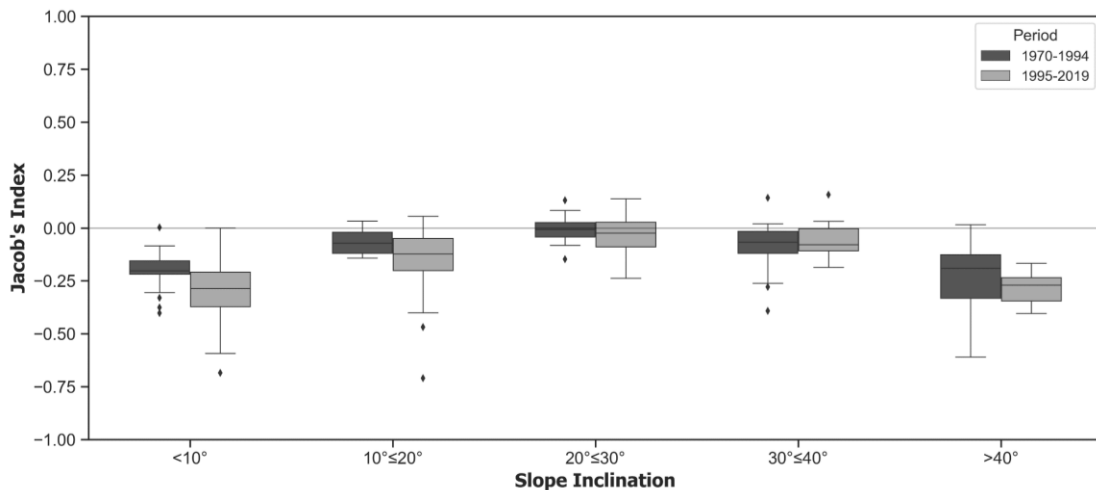
slopes appear to be even less prone to fire in the 1995-2019 interval, and flat surfaces continue to show the greatest aversion to fire.



**Figure 11: Boxplot representing the distribution of Jacobs' index (ranging from -1 to +1) for 1970-1994 (left) and 1995-2019 (right) according to Slope aspect. i) Median value (50<sup>th</sup> percentile): bar within the box, ii) first quartile (25<sup>th</sup> percentile): bottom part of the box, iii) third quartile (75<sup>th</sup> percentile): top part of the box. Whiskers represent observations outside the middle 50% and points represent outliers.**

As for aspect, figure 12 shows fire selectivity for each of the two periods based on Jacobs' selectivity index according to Slope inclination. Overall, fire is not selective with regards to inclination; in the first period, the gentlest ( $\leq 10^\circ$ ) and steepest ( $> 40^\circ$ ) inclination categories tend to be avoided by fire (values of -0.20 and -0.19, respectively). In the second period, median fire selectivity for gentlest slopes ( $\leq 10^\circ$ ) show slightly stronger avoidance, shifting from -0.2 to -0.29 while steepest ( $> 40^\circ$ ) slopes, located mainly in the eastern segment of the study area, exhibit a similar change, shifting from -0.19 to -0.27. Intermediate slope categories ( $10^\circ - 40^\circ$ ), which account for a high percentage of BA in the western (Bouches-du-Rhône) and central (Var) study area, do not exhibit any clear fire selectivity pattern.





**Figure 12:** Boxplot representing the distribution of Jacobs' index (ranging from -1 to +1) for 1970-1994 (left) and 1995-2019 (right) according to slope inclination. i) Median value (50<sup>th</sup> percentile): bar within the box, ii) first quartile (25<sup>th</sup> percentile): bottom part of the box, iii) third quartile (75<sup>th</sup> percentile): top part of the box. Whiskers represent observations outside the middle 50% and points represent outliers.

#### 2.6.4 Fire selectivity and Vegetation type

Forested and semi natural vegetation is distributed between 5 categories, of which Natural grasslands and Sclerophyllous vegetation have the lowest and the highest 50-year average covers, respectively, as the following values show: Broad-leaved forest (20.6 %), Coniferous forest (24.1 %), Mixed forest (19.2 %), Natural grasslands (11.2 %), and Sclerophyllous vegetation (24.9 %). Over the 50-year study period, Mixed and Broad-leaved forest maintain roughly the same area whereas Conifers present a slight but decreasing trend. Sclerophyllous vegetation expanded in the study area ( $\approx 6$  % increase), becoming the most common type in the last 3 decades. Finally, Natural grasslands is by far the least common type and shrunk slightly ( $\approx 3,5$  % decrease) over time.

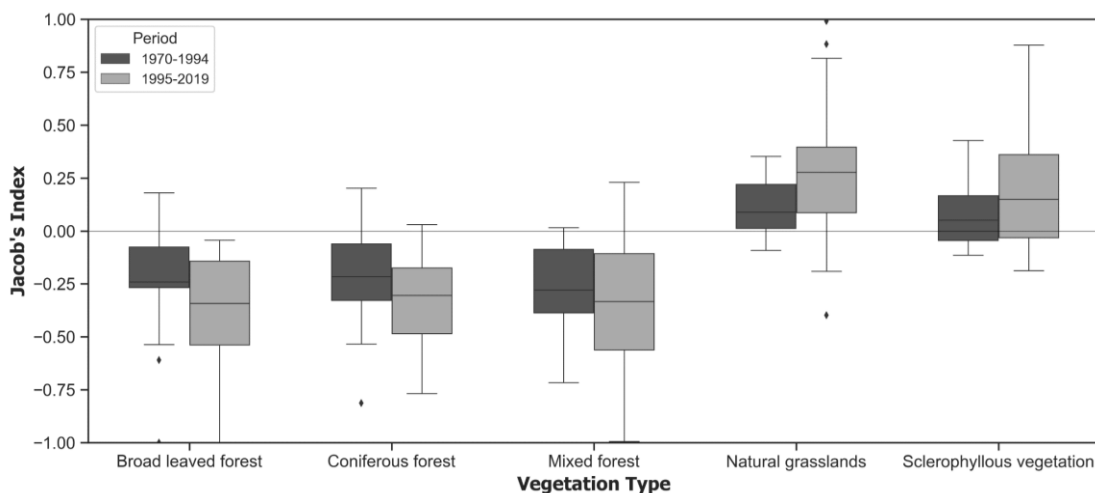
**Table 3:** Average and relative forested areas according to vegetation type between 1970 to 2019.

Type	Area (ha)	%
Broad-leaved forest	172,547	20.6
Coniferous forest	201,262	24.1
Mixed forest	160,973	19.2
Natural grassland	93,322	11.2

Sclerophyllous vegetation	208,057	24.9
Total	836,161	

Fire selectivity with regards to Vegetation type is presented in figure 13. In the first period, 3 types of vegetation show signs of fire avoidance: Mixed (-0.28), Broad-leaved (-0.24) and Coniferous (-0.21). Natural grasslands and Sclerophyllous vegetation display weak preference by fire with median values of 0.09 and 0.05, respectively.

Even though the order changes slightly in the second period, the effects of the fire suppression strategy on vegetation types are more evident than for the topographic factors. On the one hand, all three forest types are more clearly avoided by fire while on the other hand, Natural grasslands and Sclerophyllous vegetation show even stronger fire preference in the second period shifting from 0.08 to 0.28 and from 0.05 to 0.15, respectively.



**Figure 13: Boxplot representing the distribution of Jacobs' index (ranging from -1 to +1) for 1970-1994 (left) and 1995-2019 (right) according to vegetation type. i) Median value (50th percentile): bar within the box, ii) first quartile (25th percentile): bottom part of the box, iii) third quartile (75th percentile): top part of the box. Whiskers represent observations outside the middle 50% and points represent outliers.**

## 2.6.5 Geographically weighted regression

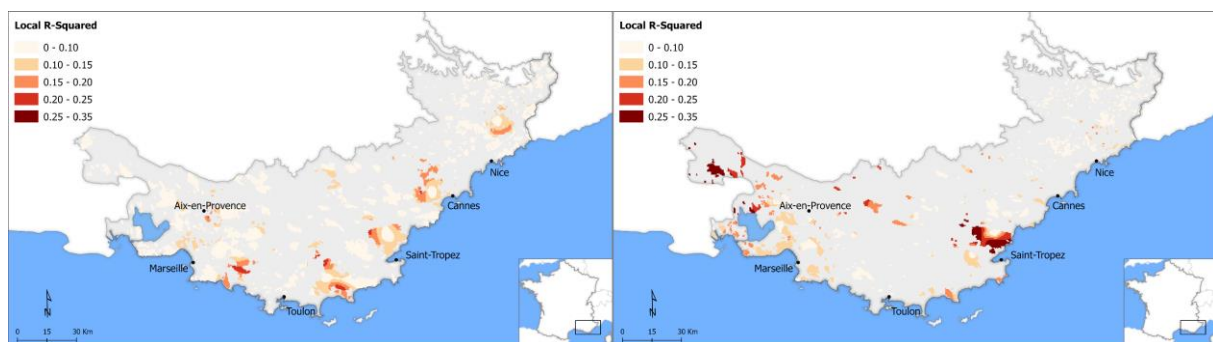
There is considerable spatio-temporal variability in the strength of the correlation between the BA and environmental variables throughout the study area. Coefficient of determination  $R^2$

values range spatially from 0.00 to 0.68 (Slope inclination) depending on the variable and time interval (Table 4). Explanatory power for all values tends to be weak, and topographic factors and Sclerophyllous vegetation shows the strongest correlations with BA. The remaining Vegetation types display a weak fit that is similar in both periods.

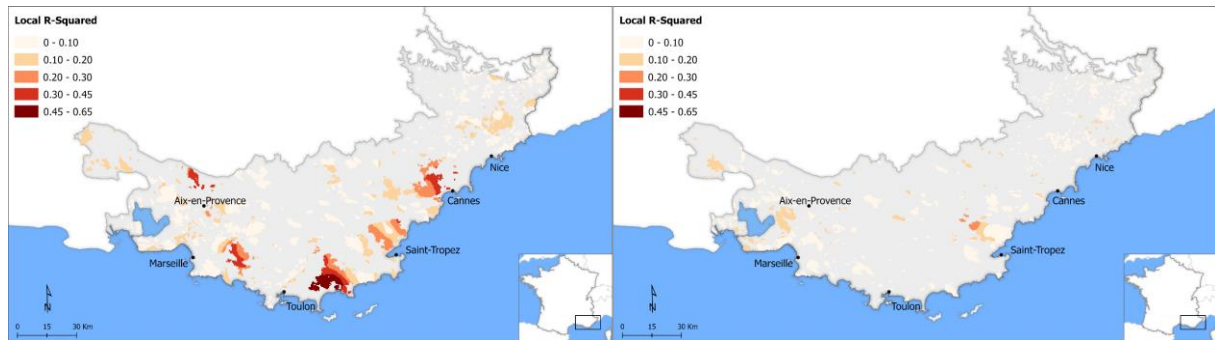
**Table 4: Descriptive statistics of local  $R^2$  per environmental factor for 1970-1994 (P1) and for 1995-2019 (P2).**

	Slope aspect		Slope inclination		Sclerophyllous vegetation		Natural grasslands		Coniferous forest		Broad leaved forest		Mixed forest		
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	
<b>Min</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
<b>Max</b>	0.24	0.36	0.68	0.25	0.48	0.47	0.19	0.21	0.20	0.21	0.11	0.23	0.26	0.25	
<b>Mean</b>	0.08	0.11	0.13	0.06	0.19	0.17	0.07	0.08	0.08	0.09	0.04	0.06	0.07	0.05	
<b>Median</b>	0.07	0.1	0.1	0.05	0.16	0.15	0.08	0.06	0.05	0.03	0.03	0.05	0.04	0.05	
<b>Std. Dev.</b>	0.08	0.12	0.08	0.04	0.12	0.11	0.03	0.05	0.03	0.03	0.02	0.05	0.05	0.04	

Figures 14 and 15 depict local  $R^2$  results of the application of GWR between percentage of BA and topographic factors. Overall, highest values are concentrated mainly in western and central parts (closer to the coastline) of the study area for both Slope aspect and inclination. The proportion of variance explained by aspect is slightly greater in the second period with several cells being in the highest class (0.25-0.35). Despite having a strong local fit in the first period, both distribution and variability changed drastically for Slope inclination in the second period.

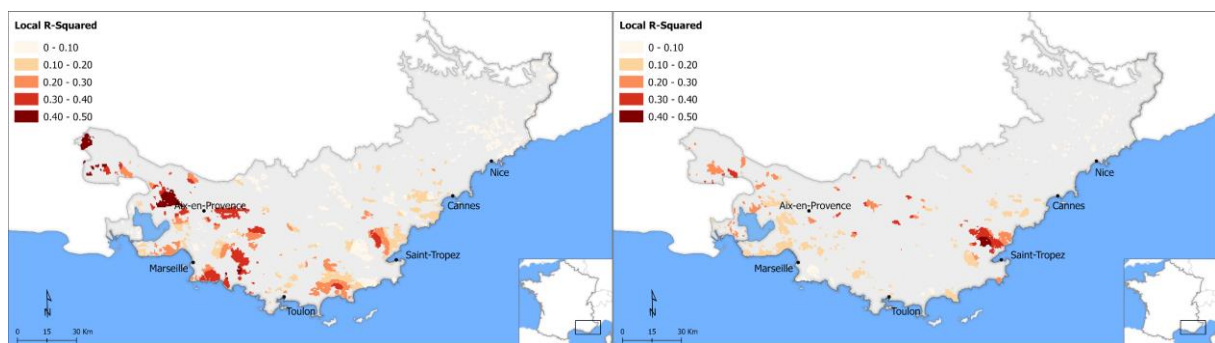


**Figure 14: Spatial distribution of local  $R^2$  between burned area and Slope aspect, for 1970-1994 (left) and 1995-2019 (right).**

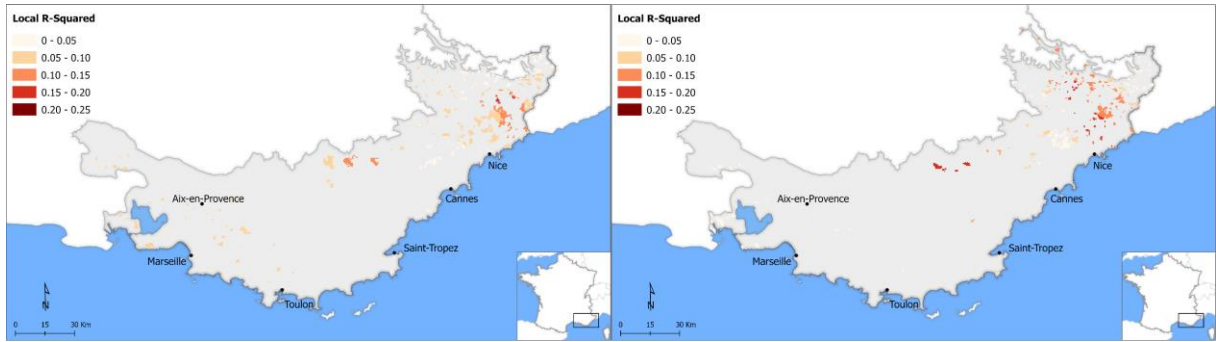


**Figure 15: Spatial distribution of local  $R^2$  between burned area and Slope inclination for 1970-1994 (left) and 1995-2019 (right).**

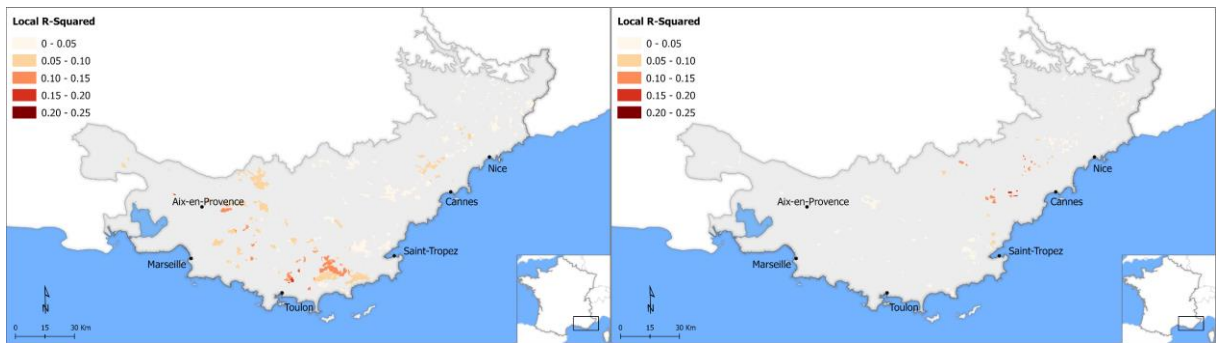
Figures 16 to 20 display local  $R^2$  results of the application of GWR between percentage of BA and percentage of each vegetation type. Similar to topographic variables, Sclerophyllous vegetation exhibits the same spatial pattern of high  $R^2$  values. A clear increase in local  $R^2$  can be observed when moving towards the western part of the region, which is more evident in the first period. Low fits are found for both periods in the higher altitude areas, located mainly in north-eastern segments of the area.  $R^2$  values for Natural grasslands are generally low and display slight differences both in terms of space and variance. Explanatory variables related to forest categories show very weak fit in the relationship with BA. In addition, the general clustering patterns are quite different between the two periods.



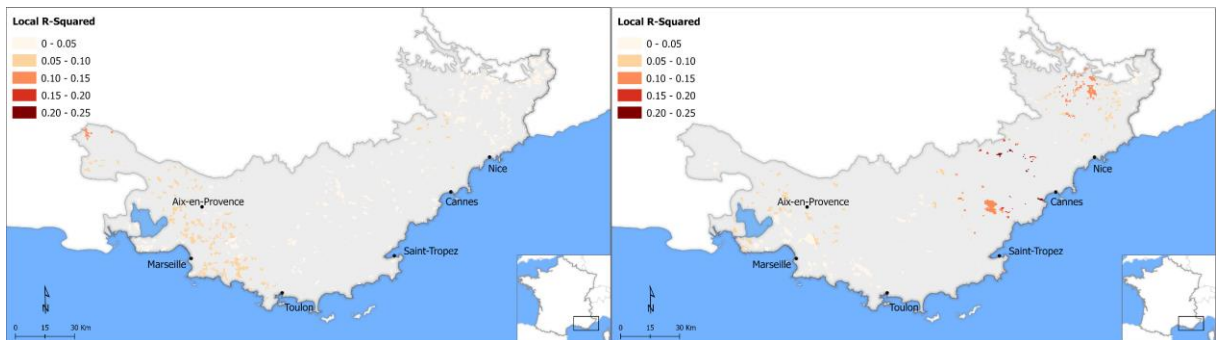
**Figure 16: Spatial distribution of local  $R^2$  between burned area and % cover of Sclerophyllous vegetation for 1970-1994 (left) and 1995-2019 (right).**



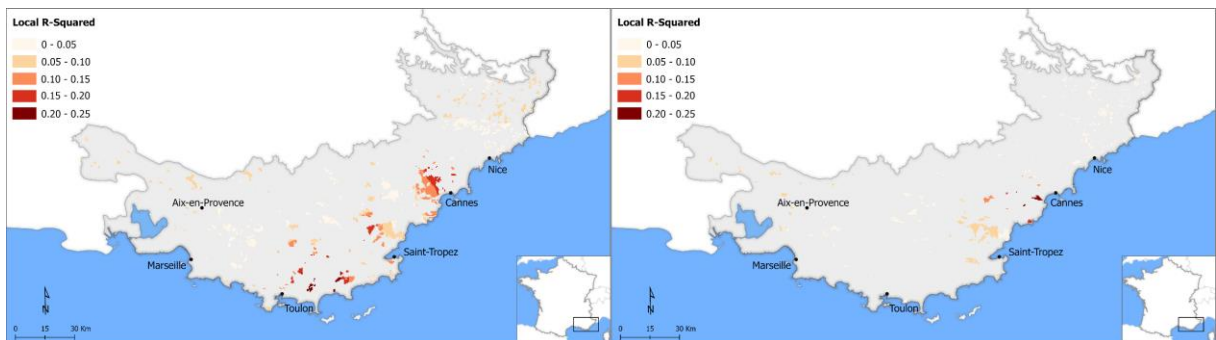
**Figure 17: Spatial distribution of local  $R^2$  between burned area and % cover of Natural grasslands for 1970-1994 (left) and 1995-2019 (right).**



**Figure 18: Spatial distribution of local  $R^2$  between burned area and % cover of Coniferous forest for 1970-1994 (left) and 1995-2019 (right).**



**Figure 19: Spatial distribution of local  $R^2$  between burned area and % cover of Broad leaved forest for 1970-1994 (left) and 1995-2019 (right).**



*Figure 20: Spatial distribution of local  $R^2$  between burned area and % cover of Mixed forest for 1970-1994 (left) and 1995-2019 (right).*

## **2.7 Discussion**

### **2.7.1 Fire history**

BA in south-eastern France has undergone substantial changes over the last 50 years. Annually, BA varies considerably but clear declining trends are observed in the second part of the temporal interval under study. Around half of the total BA (126,700 ha) was recorded in 5 years: 1979, 1986, 1989, 1990 and 2003. Due to particularly catastrophic fires in the 1980s, a new fire suppression policy (“Vulcain”) was initiated that came fully into effect in 1994 (Direction de la Sécurité Civile, 1994). This new strategy focused on aggressively suppressing fire ignitions under any weather conditions in order to avoid fire propagation to the extent where suppression would become both more difficult and more expensive. Although Fire Weather Index values were not calculated here for the 3 administrative departments, Fox et al. (2015) noted a general increase in summer temperatures between about 1980 and 2010, so the fire-fighting policy had a major impact on the decrease in total BA after 1994. Only 2003 stands out as a big fire year in the 1995-2019 interval, and although it was the hottest/driest year on record in the Alpes-Maritimes, it remained within the range of BA values of the big 1980s fires (Fox et al., 2015). Nonetheless, it raises doubts about the sustainability of rapid suppression in extreme conditions resources are spread thinly over a greater number of ignitions (Curt & Frejaville, 2018).

### **2.7.2 Spatio-temporal analysis**

The effect of the new firefighting strategy can also be viewed spatially: in general, fire patches are less large and are distributed over smaller geographic proximities with one another, and fire recurrence is lower. Spatio-temporal trends, however, vary from west to east according to

the specific population and environmental contexts of each department. In the western part of the study zone, around Aix-en-Provence and Marseille, hotspots, in the form of positive Z-scores, remain, and the new fire-fighting strategy had less effect since fires were already limited in size by vegetation continuity. Although limited in area, multiple clusters of positive trends are found in closer proximity to the built-up areas near Marseille and Aix-en-Provence in comparison to overall decreasing trends. Increased human activity, is known to affect fire ignition (Badia et al., 2011; Chas-Amil et al., 2013; Jiménez-Ruano et al., 2017; Lampin-Maillet et al., 2011) and in our context that can be potentially linked to the high arson activity found in the area (Curt et al., 2016). On the contrary, the central part of the study area, where most of the big fires occur, the new fire policy effectively limited fire propagation over the continuous vegetated cover that defines the region. This zone displays the largest clusters of negative Z-scores, decreasing BA with very few positive values and low fire recurrence. Ganteaume & Barbero (2019) provided evidence that large fires (>100 ha) declined sharply in the central segment of the study area after the introduction of the fire management policy and our results, using different methods, are coherent with their findings. Finally, in the eastern segment of the study area, frequent small, dispersed fire patches are found. Fire shapes are not elongated by wind direction compared to polygons in the western and central departments, and although negative fire occurrence trends dominate, particularly in the WUI band, there is a greater number of small positive patches compared to other zones.

### **2.7.3 Burned area and Topography**

S-facing slopes have the greatest BA, burn more frequently (Mouillot et al., 2003) and are more exposed to forest fires than other slopes due to both environmental factors (greater insolation and evapotranspiration) and WUI characteristics since S-facing slopes in southern France have more houses and therefore more potential ignition sources (Fox et al., 2018). S-facing (sum of SW, S, SE) slopes play an increasingly important role over time, and this could be linked to a

combination of hotter summers and an increasing number of human dwellings on these slopes as growth rates on S-facing slopes in the Alpes-Maritimes were 4-5 times greater than on N-facing slopes in 1990-2012.

Slope inclination favors fire propagation directly through more efficient radiative heat transfer (Rothermel, 1983) and increases the rate of spread and fire intensity (Capra et al., 2018; Csontos and Cseresnyés, 2015). In addition, slope inclination influences fire ignition and suppression indirectly through accessibility, solar radiation variations, fuel moisture, and fuel density which in turn influence flammability (Holden et al., 2009). In this study, Flat areas are most avoided by fire for several independent reasons: radiative heat transfer is less efficient on these slopes, more densely inhabited and more easily accessible with denser road networks, so lower fire preference probably depends as much or more on early suppression as on physical processes. The fire-avoidance of low slope inclinations strengthens over time, and this is coherent with more rapid suppression in this interval. BA in intermediate slope inclinations is not affected significantly by the change in firefighting strategy potentially due to factors that counter rapid suppression like more efficient radiative heat transfer, more difficult accessibility and presence of isolated or diffuse housing.

#### **2.7.4 Burned area and Vegetation type**

The role of vegetation in fire frequency and BA patches located in the Bouches-du-Rhône and Var departments was studied by Curt et al. (2013). Their case study reflects patterns observed here at a larger scale, namely that vegetation flammability is secondary to landscape organization. Large open patches of continuous fuel, as are found in the Var department, favor larger fires with longer return intervals than the small patchy wildland distribution in the Bouches-du-Rhône (Ganteaume & Barbero, 2019). Burned vegetation patterns observed here highlight the frequently cited role of Sclerophyllous vegetation (shrubland) (Ganteaume & Jappiot, 2013; Moreira et al., 2011; Oliveira et al., 2014; Tessler et al., 2016). Shrublands both



favor fire propagation in dry conditions (Baeza et al., 2002) and result from recurrent fires (Tessler et al., 2016). As Mermoz et al. (2005) suggested, the fire proneness of Sclerophyllous vegetation is connected to its ability to regenerate faster and generate quicker fuel accumulation; this also applies in our case since sclerophyllous vegetation covers the greatest area, greatest BA, greatest explained variance in the GWR analysis, and is one of two vegetation categories (with Natural grasslands) that have positive resource index values. These results are coherent with the findings of others working in Mediterranean environments where large fires tend to occur in landscapes with dense shrublands (Moreira et al., 2011; Ruffault & Mouillot, 2017). In a context where initial suppression is crucial to fire extinction, Sclerophyllous vegetation may resist early suppression better than other covers where initial propagation is perhaps slower. Moreover, firefighting assets appear to prioritize other types of vegetation during fire suppression since fire selectivity remains unchanged for bushlands, possibly due to the low cost of restoration (Oehler et al., 2012).

As other studies have concluded (Oliveira et al., 2014), Natural grasslands display a high fire susceptibility. Despite the change in the firefighting policy, grasslands are over-represented in BA in both time intervals, and this may be due to faster initial propagation or accessibility issues, as for example in certain mid to high-altitude areas over the eastern section of the study area, where burned clusters of this vegetation type are found. Sheep grazing is a common practice in high alpine pastures of the Alpes-Maritimes department, and Natural grassland fires may be due to bush clearing operations by shepherds which resulted in uncontrolled wildfires that affected much larger areas than originally intended. All three forest types (Broad leaved, Coniferous and Mixed) display a similar pattern characterized by fire avoidance, that is even more evident after the fire management policy change. This does not necessarily reflect a higher priority for suppression by firefighting assets over other vegetation types but may

indicated that fires in these vegetation types take more initial time to spread than in bushland, so they are suppressed before becoming large fires.

## **2.8 Conclusion**

In this study, results provide a coherent picture of the impact of a shift in firefighting strategy on fire occurrence and environmental characteristics. Burned area decreased sharply in SE of France after 1994 with the introduction of the new fire-fighting strategy. Rapid fire extinction was particularly effective in limiting big fires in the region. Large fire hotspots found mainly in the central parts disappear after the policy change, while new clusters of high fire recurrence appear in closer proximity to areas with increased human activity.

S-facing aspects have an increasingly bigger impact over time, and this may be linked to both environmental conditions and increased human presence on those slopes. Fire avoids low slope inclinations and even more so after the shift in fire suppression as flat areas are easier to access and more densely inhabited so lower fire preference is probably determined as much or more by early suppression as by physical processes (reduced radiative heat transfer).

Over half of the total BA in the last 50 years concerned sclerophyllous vegetation, thus confirming its strong association with high fire susceptibility and recurrence. Considering that sclerophyllous vegetation regenerates and expands faster than other vegetation types in the region, this may lead to an increase in fire risk in the future. Natural grasslands, even though they cover limited area and decline with time, are also preferred by fire which may be due to pastoral fires. On the contrary Broad leaved, Coniferous and Mixed forest are avoided by fire especially after the change in fire management policy.

Further ongoing exploitation of the fire GIS database in conjunction with WUI characteristics will likely further improve our understanding on the driving forces of BA and the impacts of fire-fighting strategies in the region.



# Chapter 3 : Arson fire cause prediction in Southeastern France

Reference to published article:

*Bountzouklis, C., Fox, D. M., & di Bernardino, E. (2022). Fire cause classification of undetermined fires in southeastern France. In D. X. Viegas & L. M. Ribeiro (Eds.), Advances in Forest Fire Research 2022 (pp. 1106– 1112). Imprensa da Universidade de Coimbra.*  
[https://doi.org/10.14195/978-989-26-2298-9\\_167](https://doi.org/10.14195/978-989-26-2298-9_167).

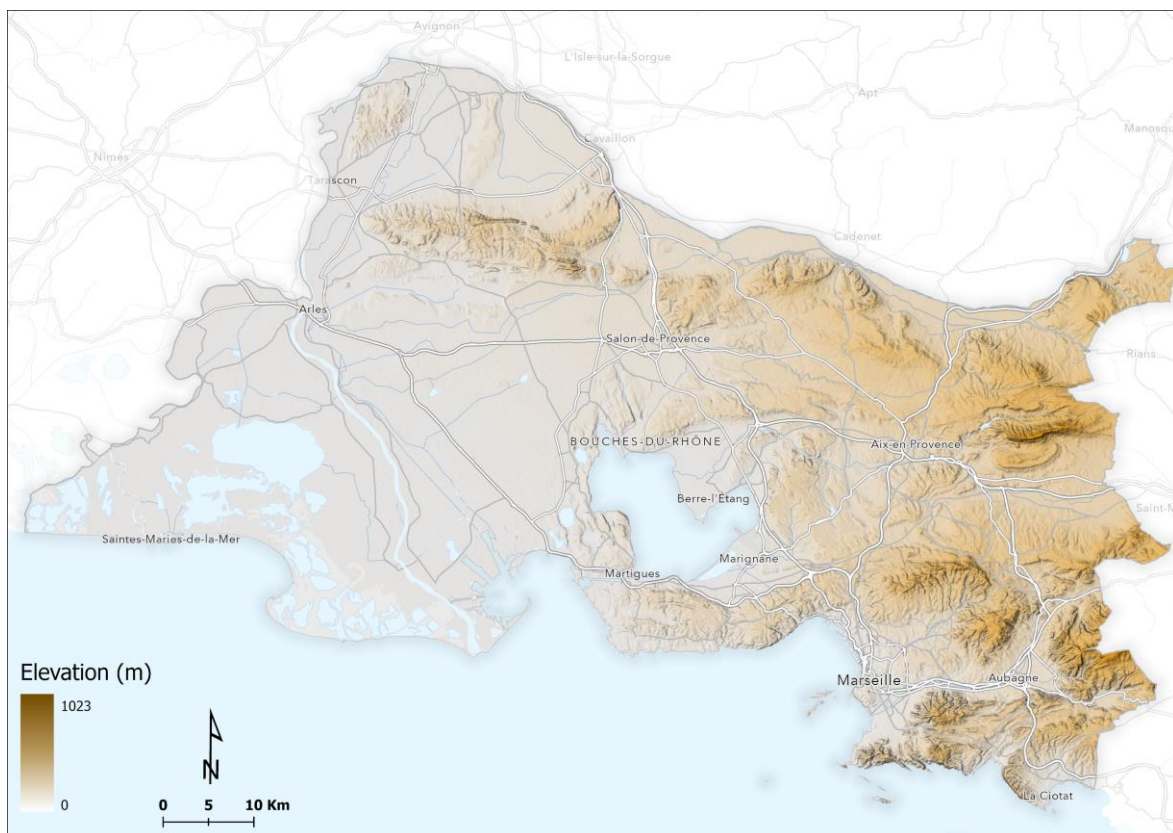
## 3.1 Introduction

Fire ignition patterns can vary significantly both temporally and spatially depending on the cause of ignition (Curt et al., 2016) and can be impacted by a plethora of environmental and anthropogenic drivers (Catry et al., 2009; Syphard et al., 2008; Syphard & Keeley, 2015). Some studies have demonstrated that arson fires can potentially be predicted both spatially and temporally (Gonzalez-Olabarria et al., 2012; Penman et al., 2013). In SE France, arson (particularly pyromania and conflict/interest) is the most frequent ignition cause for large fires (100>ha) (Ganteaume & Jappiot, 2013). Recording fire causes and studying their spatiotemporal patterns is important for establishing useful fire policies (Rodrigues et al., 2014) since a better understanding can enhance the efficacy of fire prevention measures (Oliveira et al., 2012). In France according to the national fire database (Prométhée) that contains records of fire ignition causes, approximately 70% of all fires between 1973 and 2020 were ignited by an unknown cause. The percentage of non-identified causes is high, and it is an additional constraint to the already limited research conducted on fire ignition causes. Thus, this section aims to examine whether a fire ignition point dataset coupled with machine learning methods can be used to identify the source (arson or non-arson) of unknown-caused fires and evaluate

the importance of several environmental and anthropogenic factors in determining the ignition source.

### 3.1.1 Study area

The study area covers the administrative department of “Bouches-du-Rhône”, which according to the official forest fire database in France (Prométhée), ranks second in terms of burned area and fire frequency in mainland France (Figure 21, Table 5). The department is characterized by gentle slopes and low to intermediate altitudes that increase when moving eastwards. Population density (388.8 people/km<sup>2</sup>) is higher in the eastern half of the department since that is where the second most populated city in France (Marseille) is found and because the westernmost parts are covered by wetlands and a national park. Therefore, the westernmost section has a low potential for fire ignition and propagation but increases when moving towards the eastern half of the department.



**Figure 21** *Departmental limits of Bouches-du-Rhône overlaid on a 25 m Digital Elevation Model.*

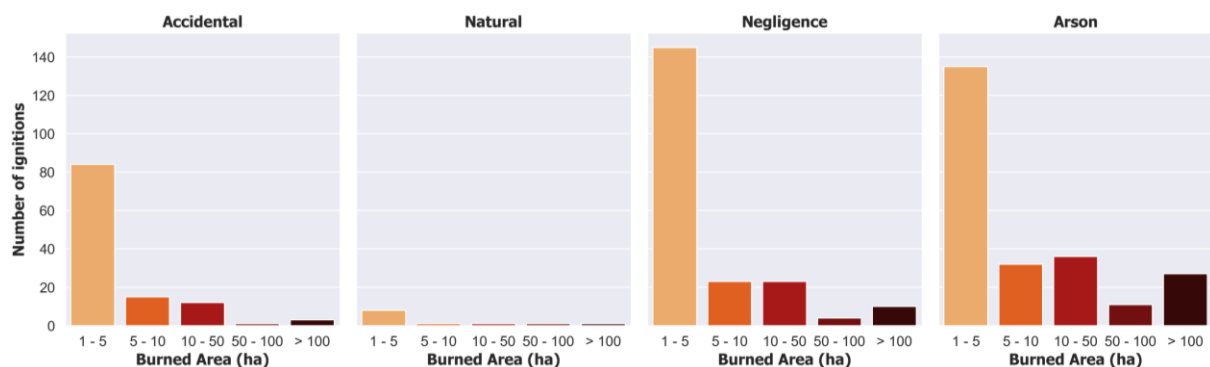
**Table 5 Environmental characteristics of Bouches-du-Rhône.**

Total area (km <sup>2</sup> )	3456
Forested area (km <sup>2</sup> )	1530
Mean slope (°)	8.8
Median slope (°)	5.7
Mean elevation (m)	142
Median elevation (m)	89

When considering only fires with a known cause, 63 % of the total burned area and approximately half (51 %) of all fire ignitions in the department are due to arson fires, according to Prométhée (Table 6). In addition, most of the large fires (>100 ha) in the study area are caused by arsonists (Figure 22). Even though negligence (professional & personal) is the second most frequent cause of fire ignition, it does not cause a proportionate volume of burned area.

**Table 6 Number of fires and volume of burned area per ignition cause from 1973 to 2020 in Bouches-du-Rhône.**

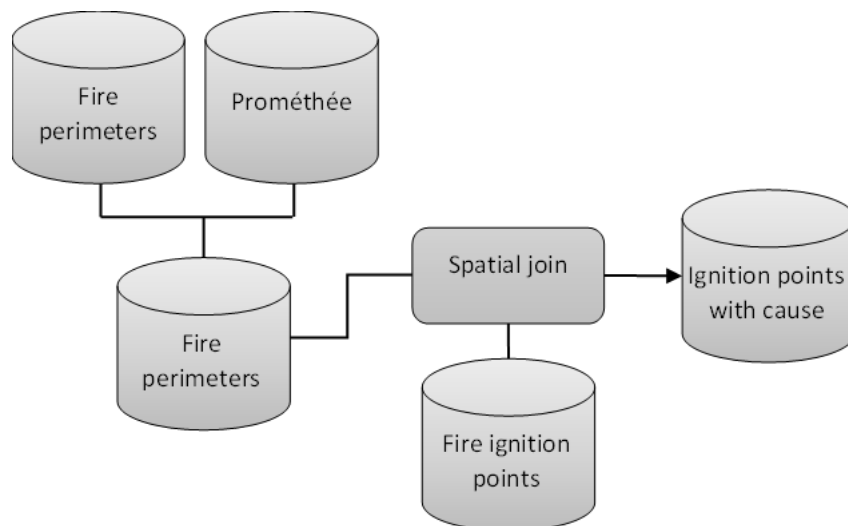
Fire Ignitions (#)	Percentage (%)	Burned Area (ha)	Percentage (%)	Cause
349	3.3	3,020	3.3	Accidental
161	1.5	335	0.4	Natural
1,034	9.7	5,223	5.7	Negligence
1,556	14.7	14,493	15.7	Arson
7,524	70.8	69,105	75	Unknown
10,624	-	92,176	-	Total



**Figure 22 Number of fires per volume category of burned area (> 1ha) and ignition cause from 1973 to 2020.**

### 3.2 Fire database

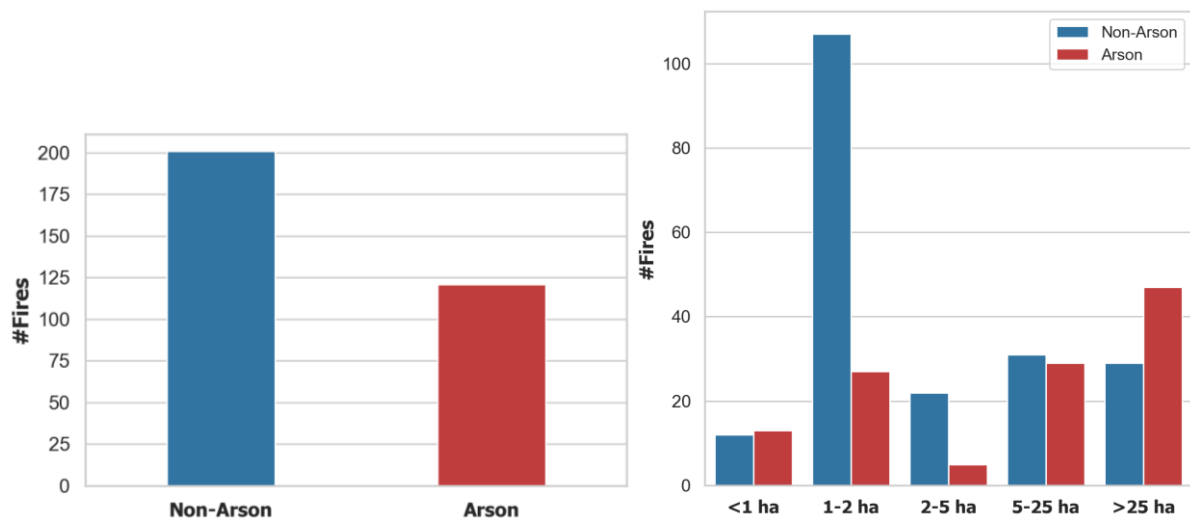
In France, the majority of forest fire research is based on Prométhée, the national database for forest fires. The specific fire database holds records of fires starting from 1973 and it includes information such as burned area, cause of ignition, date, and approximate location (within a 2x2 km grid) for each fire. In the current study, we used a geographic database that contains exact coordinates of fire ignitions that is provided by the National Forestry Office (Office National des Forêts, ONF), which to the best of our knowledge is the second time being utilized after (Ganteaume and Long-Fournel, 2015). The dataset consists of 3,234 fire ignition points ranging from 1960 to 2012, which however does not contain information on the cause of ignition. To enrich the ONF point database with the cause of ignition, two additional databases were used (Figure 23). Information from the Prométhée database was firstly merged with a polygon fire geodatabase that is described in (Bountzouklis et al., 2022b) and subsequently spatially joined with the point geodatabase resulting in a combined dataset that contains ignition coordinates, burn scars, and cause of ignition.



**Figure 23** Flow chart depicting the processing steps to generate the final dataset.

As earlier records on fire causes are considered less reliable (Ganteaume & Jappiot, 2013) only fires from 1996 to 2012 were considered, resulting in 323 fires (Figure 24). It was deemed best to classify causes into two major categories, arson and non-arson, due to the limited recorded

number of fires caused by accident, negligence and lightning strikes but also due to the significance of arson fires in the specific area.



**Figure 24** Number of fires per cause and burned area size.

### 3.3 Features

Multiple environmental and socioeconomic factors (Table 7 & Figure 25) that are known to be associated with forest fires, were acquired by a combination of European and national databases, in order to train a model that can identify the ignition cause of a fire. To account for any potential geometric errors of the ignition points and more importantly to include contextual geographic information, a circular buffer zone (500 m) was created around each fire ignition point to extract relative information.

*Table 7* List of environmental, anthropogenic, and spatiotemporal variables considered.

Type	Name	Description	Source
Land Cover	Artificial surfaces	Percentage of cover within each zone	Corine Land Cover - 2006
	Agriculture		
	Vegetation		
	Water/ Wetlands		
Topographic	Slope inclination	Mean inclination (°)	National Geographic Institute - 5 m spatial resolution
	Slope aspect	Percentage of cover within each zone (Flat, N, E, S, W)	
	Elevation	Mean elevation (m) above sea level	



	Topographic wetness index	Topographic-driven control on soil moisture	
<b>Vegetation type</b>	Coniferous	Percentage of cover within each zone	National Geographic Institute - 2006
	Coppice		
	Hardwood		
	Open Forest		
	Shrublands		
<b>Anthropogenic</b>	Population density	Individuals per sq. m.	National Institute of Statistics and Economic Studies - Sub-municipal level - 2006
	Gini index	Inequality index	
	Income	Mean taxable income (€)	
	Unemployment	Unemployment rate (%)	
	Primary road distance	Euclidean distance from fire location (m)	National Geographic Institute - 2008
	Secondary road distance		
	Power pylons distance		
	Railway distance		
	Wildland Urban Interface distance		
House density	Buildings per sq. m.		
<b>Spatiotemporal characteristics</b>	Season	Winter, spring, summer, autumn	Fire database
	Burned area size	<1 ha, 1-2 ha, 1-2 ha, 2-5 ha, 5-25 ha, >25 ha	
	Coordinates	XY coordinates of each fire	

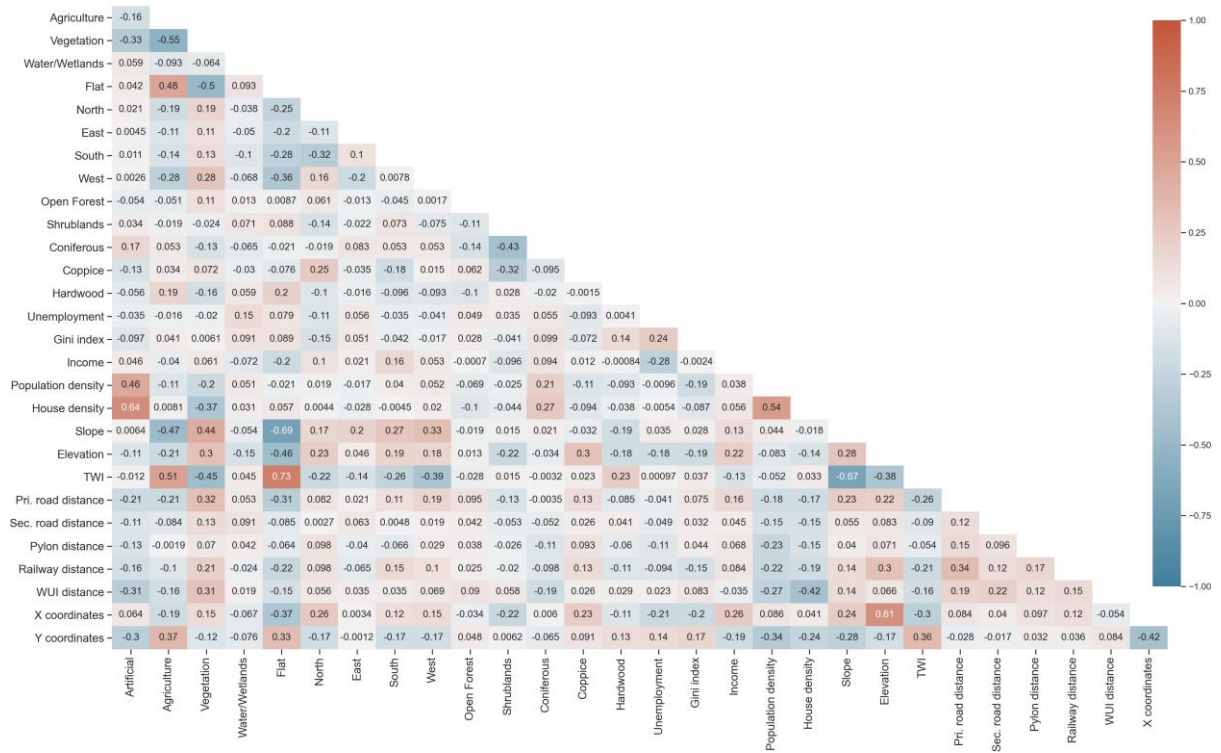


Figure 25 Kendall's Tau rank correlation coefficient heatmap of explanatory variables.

### 3.4 Methods

The analysis of the data is based on Random Forests (Breiman, 2001), a well-established machine learning algorithm in many disciplines but also in wildfire science (Jain et al., 2020). In order to train the model, 70% of the original dataset was utilized, while the remaining 30% was used for testing the accuracy in predicting the cause of a fire. Due to the limited number of observations, the process of splitting the data (using the same ratio) and executing the model was iterated (n=300) to have a more consistent perception of the accuracy of the model. Additionally, the processing chain included tuning the algorithm's hyperparameters as well as calculating the feature importance score for all variables.

To better comprehend and evaluate the list of the factors that can drive the classification output, the Gini impurity (importance) method was used. In scikit-learn, the importance of a node  $j$  in a single decision tree is computed using the following formula:

$$n_{ij} = w_j C_j - w_{left(h)} C_{left(j)} - w_{right(h)} C_{right(j)} \quad (3)$$

Where  $w_j$  is the weighted number of samples in node  $j$  as fraction of total weighted number of samples,  $C_j$  is the impurity in node  $j$  and  $\text{left}(j)$  and  $\text{right}(j)$  are its respective sub nodes. Successively, feature importance of feature  $i$  is calculated as:

$$f_i = \frac{\sum_{j:\text{node } j \text{ splits on feature } i} n_j}{\sum_{j \in \text{all nodes}} n_j} \quad (4)$$

Where  $f_i$  is the importance of a feature  $i$  and  $n_j$  the importance of a node  $j$ , which can be normalized to values ranging from 0 to 1 by dividing by the sum of all feature importance values:

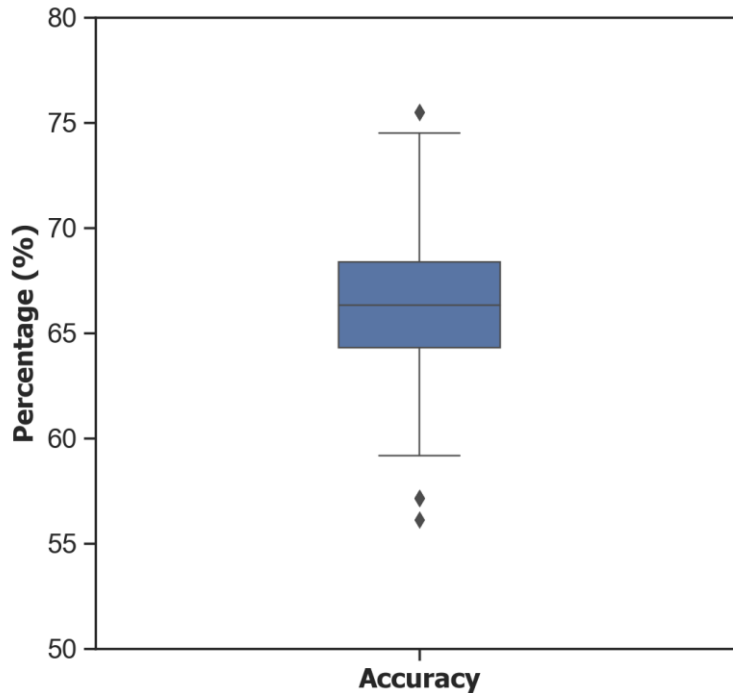
$$\text{normalized } f_i = \frac{f_i}{\sum_{j \in \text{all features}} f_j} \quad (5)$$

However, for RF the final feature importance values are computed based on the total number of trees in the model. The higher the value the higher the importance of a feature:

$$RF f_i = \frac{\sum_{j \in \text{all trees}} \text{normalized } f_{i,j}}{\text{Total number of trees}} \quad (6)$$

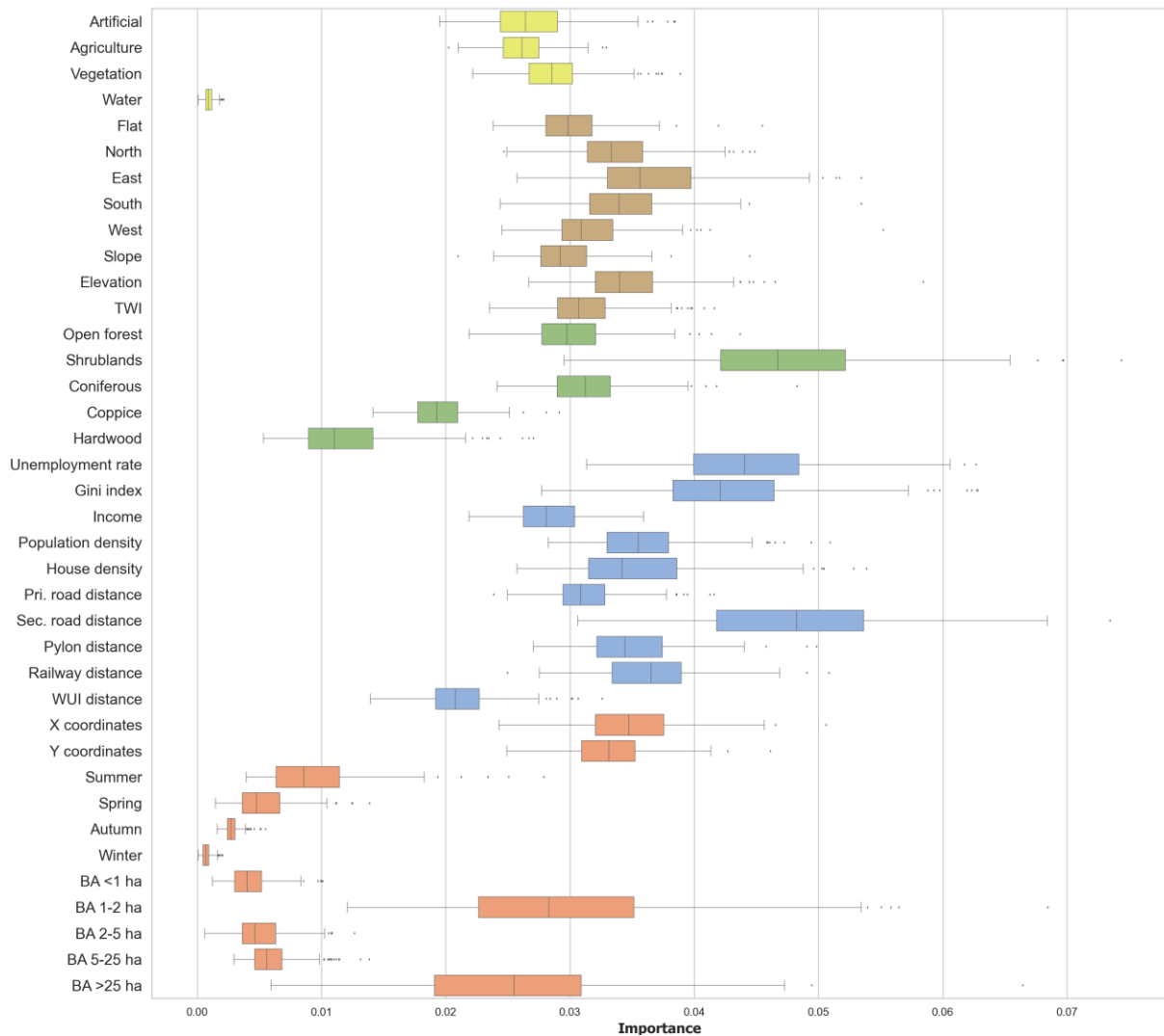
### 3.5 Results

The accuracy of the model is illustrated in the form of a boxplot (Figure 26) that represents accuracy scores from all iterations of the model. The accuracy of the model to classify unknown caused fires can vary substantially ranging from 56% to 76% (median value 67%), due to the small size of the dataset.



**Figure 26** *Boxplot representing Random Forest's classification accuracy for all iterations (n=300). (i) Bar within the box is the median value, (ii) bottom part of the box is the first quartile, and (iii) top part of the box is the third quartile. Whiskers represent observations outside the middle 50 % and points represent outliers.*

Figure 27 shows the importance values of each explanatory variable used in the model. Overall, anthropogenic features (in blue) appear to surpass in importance the rest. Moreover, topographic factors (in brown) seem to be more important than land cover and vegetation type while spatiotemporal variables except for XY coordinates hold the lowest importance. More specifically, Secondary road distance, Shrublands and Unemployment rate are the three variables displaying the highest importance. Finally, several variables mainly related to spatiotemporal characteristics and others such as certain vegetation types (Hardwood & Coppice) as well as Water can be excluded to decrease model complexity and execution time.



**Figure 27** Distribution of the variable importance values for all iterations (n=300).

### 3.6 Conclusion

The results of the study suggest that the source of unknown caused fires can be identified at an acceptable level of accuracy even with a limited number of fires. Anthropogenic drivers such as distance to secondary roads and unemployment rate, along with higher volumes of shrublands around ignition points are the most important features in determining the classification of unknown caused fires for the specific area.

Overall performance of such models would most likely greatly benefit from the exploitation of larger datasets as well as from the inclusion of weather-related variables. Finally, as location holds particular importance over certain fire causes, spatial extensions of machine learning

algorithms such as Geographic random forests (Georganos et al., 2021) and Geographically weighted neural networks (Hagenauer & Helbich, 2022) could provide significant enhancements over the original algorithms.



# Chapter 4 : Unknown wildfire ignition cause prediction using machine learning

Reference to published article:

*Bountzouklis, C., Fox, D. M., & di Bernardino, E. (2023). Predicting wildfire ignition causes in Southern France using eXplainable Artificial Intelligence (XAI) methods. Environmental Research Letters, 18 (4), 044038. <https://doi.org/10.1088/1748-9326/acc8ee>*

## 4.1 Introduction

In Europe, approximately 50% of all fires were caused by an unknown origin based on data reported from 19 European countries in the European Forest Fire Database (EFFIS) from 1999 to 2016 (De Rigo et al., 2017). In Mediterranean France, almost 70% of all fires between 1973 and 2020 were recorded without a cause of ignition according to the forest fire database for the Mediterranean area (Prométhée.com). Many experts in the field of fire management in Europe have acknowledged the importance of fires classified as having an undetermined origin (Tedim et al., 2022), since the lack of information regarding fire causes makes it difficult for fire managers to determine the most suitable course of action to prevent similar incidents from happening in the future. French fire experts, in particular, have identified fires of unknown origin as being of paramount importance (Tedim et al., 2022) among the various categories of fires: Natural, Accident, Negligence, Deliberate, and Rekindle of the harmonized classification scheme of fire causes in Europe (Camia et al., 2013). In southeastern France, Ganteaume & Guerra (2018) highlighted the fact that large areas are burned by fires of undetermined sources, and they argue for enhanced quality and quantity of investigations into fire ignition causes in order to improve the accuracy of fire databases. Fire ignition patterns can vary significantly both temporally and spatially depending on the cause of ignition (Curt et al., 2016) and can be



impacted by a plethora of environmental and anthropogenic drivers (Catry et al., 2009; Syphard et al., 2008; Syphard and Keeley, 2015). As such, documentation and research of fire causes and their spatiotemporal patterns are essential for establishing meaningful fire policies (Rodrigues et al., 2014) since a better understanding of these patterns can improve the efficacy of fire prevention strategies (Oliveira et al., 2012a). However, the absence of comprehensive fire-cause data hampers the utility of these databases.

Similar to other areas of study, the utilization of machine learning (ML) methodologies in wildfire science has seen a marked increase in popularity in recent years (Bot & Borges, 2022; Jain et al., 2020). Applications of ML in wildfire science include classifying and mapping fuel properties (Riley et al., 2014) and quantities (López-Serrano et al., 2016) as input variables in fire behavior models; fire and smoke detection (Zhao et al., 2018; Ba et al., 2019) for rapid firefighting response; active or final burned area and severity mapping based on remote sensing data for suppression planning and damage assessment (Pereira et al., 2017; Collins et al., 2018; Ban et al., 2020); forecasting of fire weather variables (Lagerquist et al., 2017; Bates et al., 2017); prediction of fire occurrence (Dutta et al., 2016; Vecín-Arias et al., 2016), fire spread/growth rates (Chetehouna et al., 2015) and burned area (Mayr et al., 2018; Hodges and Lattimer, 2019) for more effective dissemination of suppression assets; modeling and mapping of fire risk based on various environmental and anthropogenic variables, to identify potential drivers of fire ignition and/or burned area (Curt et al., 2016; Molina et al., 2019; Ghorbanzadeh et al., 2019).

While ML models have demonstrated great effectiveness at identifying complex patterns in large datasets, some are considered "black boxes" because it can be difficult to understand how the model arrives at its predictions or how certain patterns were identified (Loyola-Gonzalez, 2019). This lack of interpretability can be a barrier to adoption, as it may be difficult for stakeholders to trust such models without understanding the complete algorithm inference

pattern. In recent years, eXplainable Artificial Intelligence (XAI) / Interpretable Machine Learning (IML) has emerged as an approach that employs various techniques and strategies to enhance the interpretability, transparency, and explainability of ML models and their decision-making processes with the ultimate goal of fostering trust and accountability in the model's output. In the context of wildfire science, the application of XAI has been explored by only two recent studies to address wildfire occurrence and size (Al-Bashiti & Naser, 2022; Cilli et al., 2022).

Research conducted on fire ignition causes is fairly limited and poorly understood, but some studies have demonstrated that arson fires can potentially be predicted both spatially and temporally (Gonzalez-Olabarria et al., 2012; Penman et al., 2013). The objective of this study is to develop a ML-based model that can classify the ignition source of fires that have been recorded without a known cause in France. Furthermore, this study aims to evaluate the significance and the effect of various environmental and anthropogenic factors in determining the classification of different fire sources utilizing XAI methods.

#### **4.1.1 Study Area**

The study area comprises 15 administrative divisions (departments) in the south of France, with a total area of just over 80,000 km<sup>2</sup> (Table 8, Figure 27). The specific region is considered the most fire-prone in France and where most of the burned area is recorded, despite exhibiting decreasing trends in the last decades (Bountzouklis et al., 2022b). Environmental characteristics and landscapes vary significantly with both mountainous and coastal zones contained in the study area; the highest altitudes and steepest slopes are found in the northeastern parts where the French Alps are located (e.g., Hautes-Alpes, Alpes-de-Haute-Provence) whereas in the southern portions topography is low-lying and relatively flat (e.g., Bouches-du-Rhône, Hérault). Population density is influenced by topography: the highest concentrations are located in areas with low altitudes and gentle slopes, especially in the

southeastern Mediterranean coastal and near coastal zones (e.g., Bouches-du-Rhône, Alpes-Maritimes). The French Alps and the island of Corsica are largely covered by forests & semi-natural areas whereas the largest agricultural areas are concentrated mainly in the center of the study area.

**Table 8 Physical and human characteristics of the administrative divisions within the study area based on data from <sup>a</sup>French National Institute of Geographic (IGN), <sup>b</sup>Corine Land Cover (2018) & <sup>c</sup>National Institute of Statistics and Economic Studies (INSEE).**

<b>Administrative division</b>	<b>Area<sup>a</sup> (km<sup>2</sup>)</b>	<b>Mean elevation<sup>a</sup> (m)</b>	<b>Mean slope<sup>a</sup> (°)</b>	<b>Population density<sup>c</sup> (people/km<sup>2</sup>)</b>	<b>Artificial surfaces<sup>b</sup> (%)</b>	<b>Agricultural areas<sup>b</sup> (%)</b>	<b>Forests &amp; semi-natural areas<sup>b</sup> (%)</b>
<b>Alpes-de-Haute-Provence</b>	6,995	1,150	15.5	23.7	1.1	19.2	79.7
<b>Hautes-Alpes</b>	5,691	1,665	20.6	24.8	1.3	15.2	83.5
<b>Alpes-Maritimes</b>	4,294	1,110	20.3	257.0	8.9	4.5	86.6
<b>Corse-du-Sud</b>	4,019	536	15.3	40.4	2.2	11.1	86.7
<b>Haute-Corse</b>	4,707	592	16.2	39.8	1.9	12.4	85.6
<b>Ardèche</b>	5,566	596	11.1	59.4	2.7	28.1	69.3
<b>Aude</b>	6,344	359	7.4	59.9	2.8	48.4	48.7
<b>Bouches-du-Rhône</b>	5,091	142	3.7	405.1	15.7	42.2	42.1
<b>Drôme</b>	6,559	581	11.1	80.0	3.4	41.1	55.5
<b>Gard</b>	5,875	248	6.2	128.2	6.5	40.3	53.2
<b>Hérault</b>	6,230	264	6.5	195.5	7.2	41.5	51.3
<b>Lozère</b>	5,176	1,025	9.3	14.8	0.7	25.5	73.8
<b>Pyrénées-Orientales</b>	4,139	855	12.6	117.3	5.2	27.5	67.3
<b>Var</b>	6,032	364	7.8	181.6	9.2	20.9	69.9
<b>Vaucluse</b>	3,578	339	5.6	157.3	6.8	53.8	39.4

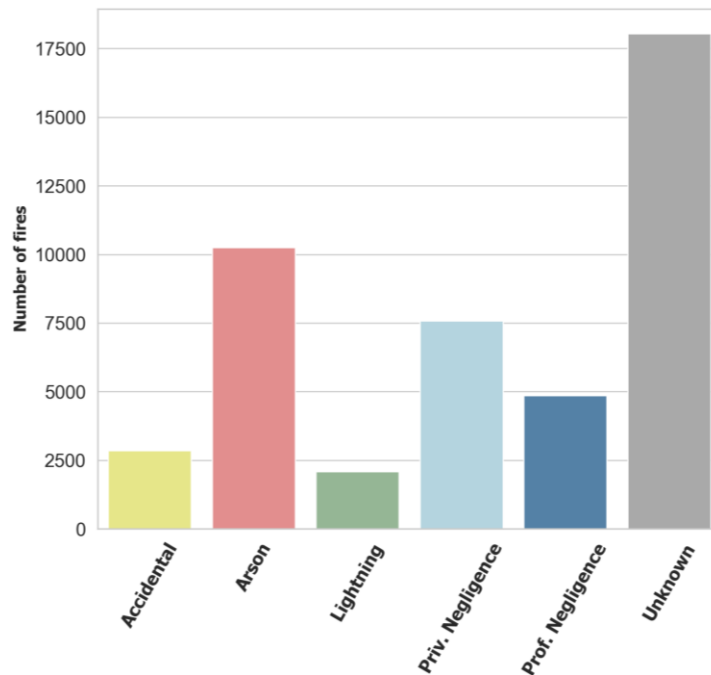


*Figure 28 Location map illustrating the administrative division limits of the study area.*

## 4.2 Fire database

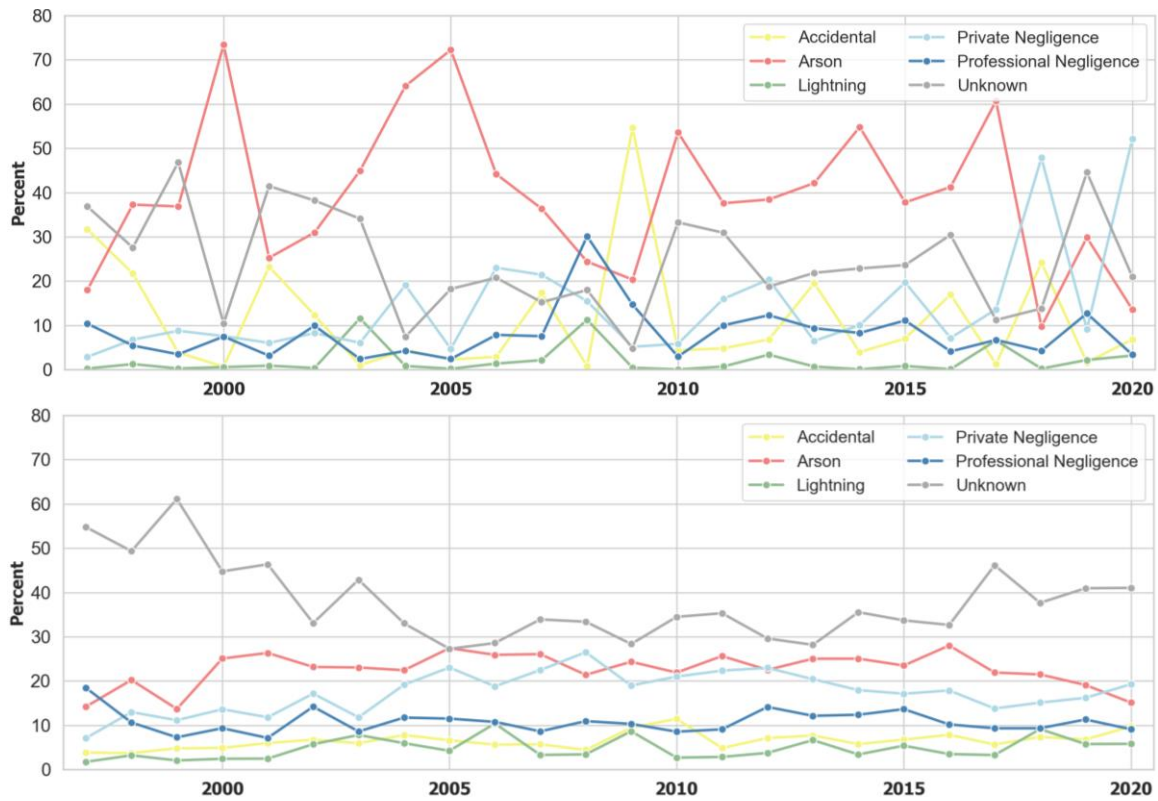
The current study was based on “Prométhée”, the official forest fire database for Mediterranean area in France. This database documents fires from 1973 onwards and contains information for each fire such as burned area, ignition source (known / unknown), time, date, and location within a 2x2 km grid. Similar to the harmonized European classification scheme on ignition causes (Camia et al., 2013), “Prométhée” includes 5 major fire ignition sources: i) accidental (e.g., power lines, vehicles), ii) arson (e.g., pyromania, conflict), iii) private negligence (e.g., cigarette butts, leisure), iv) professional negligence (e.g., industry, agriculture) and v) lightning. The total number of fires considered in our study is 48,038; these were recorded from 1997 to 2020. Fire records prior to 1997 were excluded from this study since classification on the origin of fires is considered less reliable (Ganteaume & Jappiot, 2013). The dataset comprised of records starting in 1997 is fairly balanced with regards to the number of fires of

known/unknown sources as approximately 60 % have a known cause of ignition. Within the known causes (n=27,620) frequency varies considerably; arson is the most frequent (38.4 %), followed by private negligence (26.7 %), professional negligence (17.2 %), accidental (10.1 %) and finally lightning (7.6 %) (Figure 29).



**Figure 29** Number of fires per cause in the study area (1997-2020).

After unknown causes, arson fires are both the most numerous and account for the greatest annual burned area most years (Figure 30a & 30b). This is followed by private negligence, which, even though is the second most frequent fire source, it does not cause a proportionate extent of burned area. Despite similar numbers of accidental and lightning fires, the annual percentage of area burned by accidental fires is often significantly greater than that burned by lightning fires and occasionally greater than the other causes. Lastly, although the percentage of burned areas by unknown origin fires is substantial most years, frequently second after Arson, it fluctuates widely from 5 % to 49% depending on the year.



**Figure 30** Percentage of a) burned area and b) fire ignitions according to cause and year.

### 4.3 Considered features

Table 9 describes the land cover, topographic, anthropogenic, and spatiotemporal variables that were used as features to predict the fire ignition source. The contextual geographic information of the selected factors was processed for each 2x2 km grid initially in ArcGIS Pro v2.9 and subsequently using python packages pandas (McKinney, 2010) and NumPy (Harris et al., 2020) to preprocess the data for the classification scheme (e.g., replace missing values, one-hot encoding, etc.) and finally for visualization purposes seaborn (Waskom, 2021).

**Table 9** List of features used to model the occurrence of forest fires.

Type	Name	Description	Source
Land Cover	Artificial surfaces	Percentage of cover within each grid	Corine Land Cover – 2006 (Raster - 100 m spatial resolution)
	Agriculture		
	Forest		
	Low vegetation (e.g., shrublands)		
	Water		

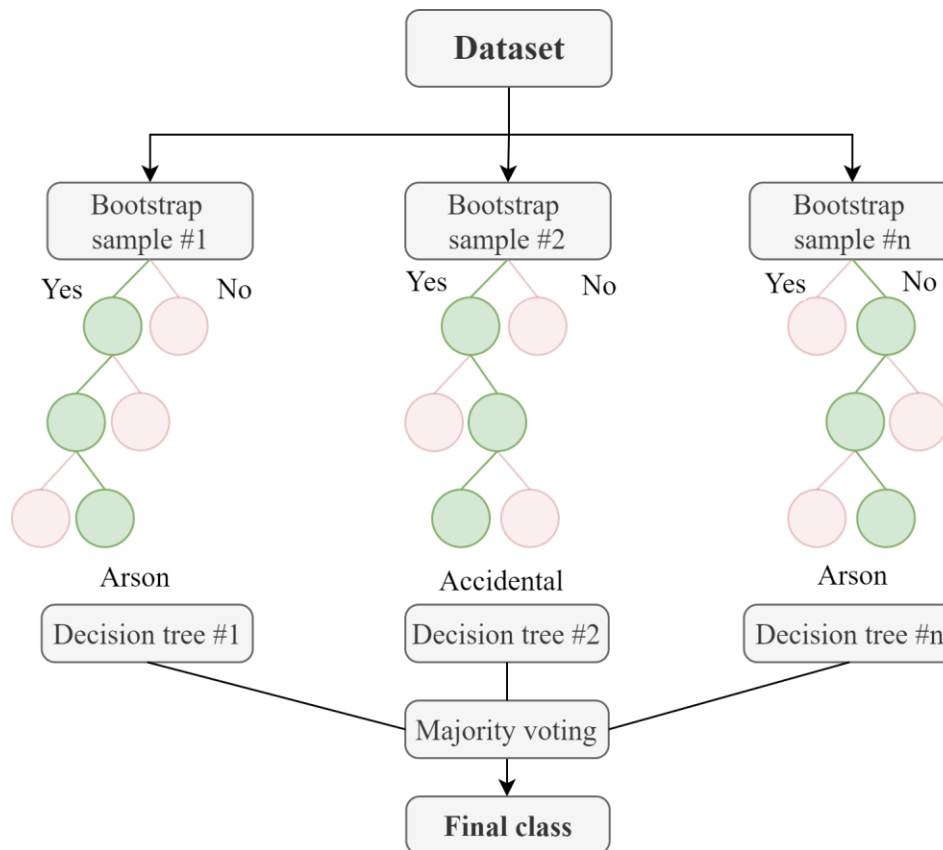
	Wetlands		
<b>Topographic</b>	Elevation	Mean elevation (m) above sea level	National Geographic Institute (Raster - 100 m spatial resolution)
	Slope aspect	Percentage of cover within each zone (North, East, South, West)	
	Topographic wetness index (TWI)	Topographic-driven control on soil moisture	
	Topographic ruggedness index (TRI)	Amount of elevation difference between adjacent cells of a DEM (S. J. Riley et al., 1999)	
<b>Anthropogenic</b>	Population density	Number of individuals/area	National Institute of Statistics and Economic Studies - Sub-municipal level - 2006 - (Tabular)
	Gini index	Inequality index	
	Income	Mean taxable income (€)	
	Unemployment	Unemployment rate (%)	
	Primary road density	Total line length/area	National Geographic Institute - 2008 (Vector)
	Secondary road density		
	Powerline density		
Railway density			
<b>Spatiotemporal</b>	Season	Winter (Dec -Feb), spring (Mar - May), summer (Jun - Aug), autumn (Sep -Nov)	Fire database "Prométhée"
	Burned area size	<0.1 ha, 0.1-0.5 ha, 0.6-1 ha, 2-5 ha, 1-25 ha, >25 ha	
	Coordinates	XY centroid coordinates	
	Time of the day	Morning (5:00 – 12:00), afternoon (12:00 – 17:00), evening (17:00 –	

		21:00), night (21:00 – 5:00)	
--	--	---------------------------------	--

#### 4.4 Methods

ML methodologies learn and adapt through the process of experience, where the size and quality of the input data play a critical role in determining the overall effectiveness of the model. Random forests (RF) (see e.g., Breiman, 2001) is a supervised ML algorithm used both for classification and regression that is well-established in many disciplines and has grown substantially in popularity in the field of wildfire science over the last decade (Jain et al., 2020). RF is based on decision trees (Breiman et al., 2017), where each decision tree is a series of If-Then-Else sequences with several branches connected by decision nodes and finally by leaf nodes that eventually determine a value or category such as the label of a classification task (Figure 31). Furthermore, a fundamental characteristic of RF is that a random subset of features is used at each node of each decision tree, resulting in several individually trained and uncorrelated decision trees, and these are finally merged into a larger ensemble model to limit overfitting and produce more accurate predictions.

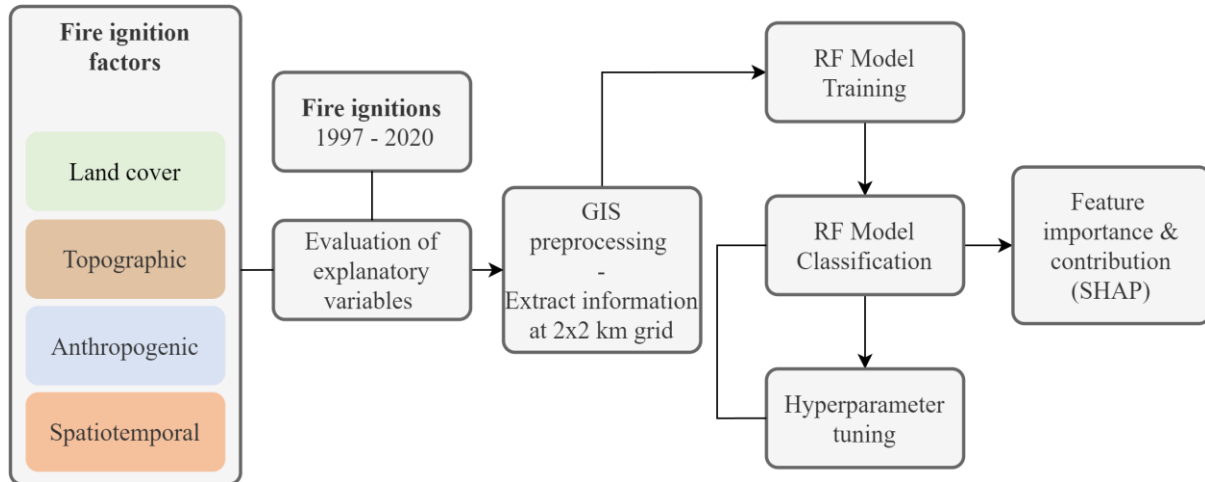




**Figure 31 Random forests diagram.**

The processing chain of RF (classification, accuracy score, confusion matrix, hyperparameter tuning, etc.) was carried out using the implementation of the algorithm in Python module Scikit-Learn (Pedregosa et al., 2011) (Figure 32). To address the unbalanced number of samples between classes, the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) was used, which is implemented under Python package scikit-learn imbalanced-learn (Lemaitre et al., 2016). SMOTE is a common method to produce synthetic data from a minority class (e.g., lightning ignitions) by randomly selecting one of the k-nearest-neighbors and using it to generate new, but randomly tweaked, similar samples. To train the classifier, 70% of the dataset was utilized, while the remaining 30% was used for testing the accuracy in predicting the cause of a fire. The synthetic samples created using SMOTE were utilized only during the training phase and not for the validation of the model. To finetune the algorithm hyperparameters such as number of trees, max number of features considered for splitting a node, max levels in each decision tree etc., scikit-learn Random Search Cross Validation

method was used; this allowed us to evaluate and narrow down a wide range of values for each hyperparameter. Subsequently, the Grid Search with Cross Validation method was used to examine different combinations of specific values for the selected hyperparameters.



**Figure 32** Flow chart depicting the processing chain used to classify fire causes.

To evaluate the model, accuracy, precision, recall and F1-score were used; these are commonly used as evaluation metrics for classifiers in the field of ML, which are calculated using the number of instances classified as true positives (TP), true negatives (TN), false positives (FP) or false negatives (FN). Accuracy is determined by dividing the number of correct predictions by the total number of instances. Precision specifies how many of the instances the classifier predicted as positive are actually positive while recall shows what fraction of the positive instances in the dataset were correctly identified by the classifier. F1-score serves as a comprehensive measure of both precision and recall. A F1-score of 1 denotes optimal performance, with both precision and recall being maximized. Conversely, a score of 0 represents the worst possible outcome, with both precision and recall being minimal. A score of 0.5, which is equivalent to random guessing, is suboptimal performance, whereas scores above 0.5 are generally considered to indicate good performance.

$$Accuracy = \frac{TP + TN}{Total\ number\ of\ instances} \quad (7)$$

$$Precision = TP / (TP + FP) \quad (8)$$

$$Recall = TP/(TP + FN) \quad (9)$$

$$F1score = 2 * (Precision * Recall)/(Precision + Recall) \quad (10)$$

To identify which features are driving the classification but also to comprehend the contribution of each one, the SHapley Additive exPlanations (SHAP) (Lundberg & Lee, 2017) method was utilized. SHAP is an approach based on game theory that is used to explain the ML model outputs by breaking down the prediction into contributions from each feature value. These contributions are combined and help us understand the overall importance of each feature value in the final prediction. SHAP values can be visualized using various plots, such as a summary plot, that allow us to display not only the strength of the impact a certain feature has but also the direction of the impact.

## 4.5 Results

As elaborated below, the results derived from the RF model are presented through classification metrics and a confusion matrix, subsequently followed by the description of which features drive the classification and how they influence it.

### 4.5.1 Fire ignition cause classification

The overall accuracy of the multiclass RF classification scheme reaches about 70% (69.8%). Detailed results per ignition cause are presented in Table 10 and evaluated using i) F1-score, ii) precision and iii) recall. Concerning the accidental class, the model displays the second highest F1-score (0.77) and a moderate discrepancy between precision (0.81) and recall (0.74). This indicates that the model is able to correctly identify most of the instances as accidental when it predicts that class, but it misses more instances that actually belong to that cause. Regarding the arson class, the model shows a lower F1-score of 0.64 and is not very accurate in terms of precision (0.60), meaning that it may predict some instances as arson that actually belong to a different class. However, the model performs better when it comes to

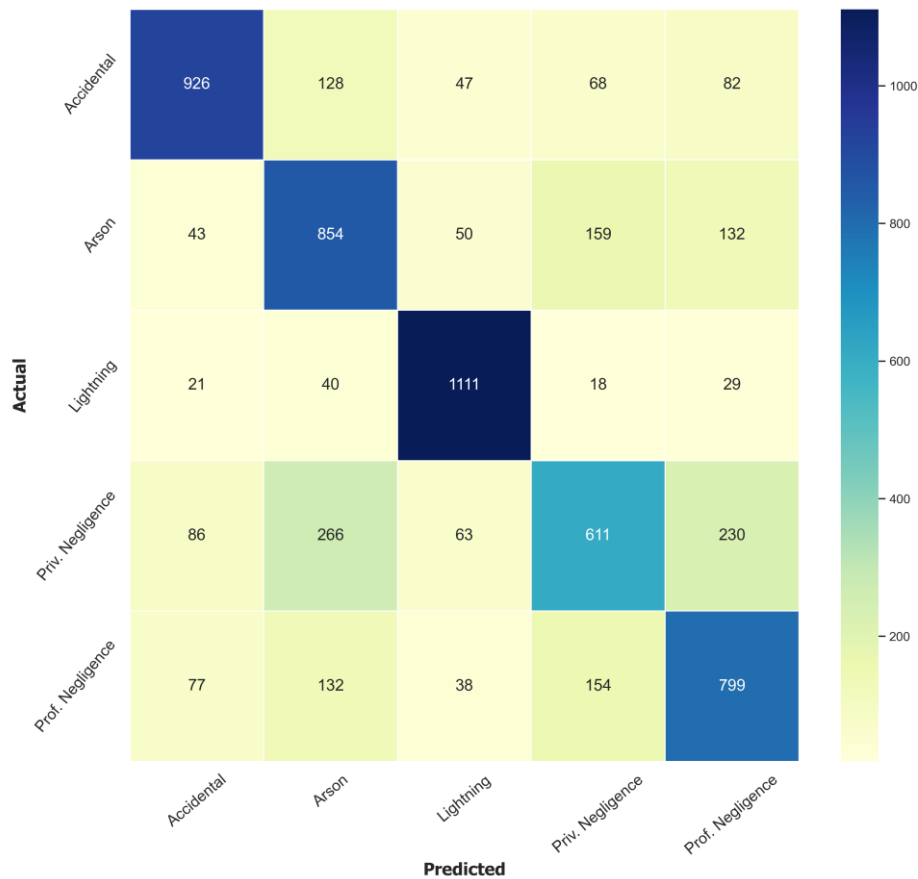
identifying most of the instances that belong to the arson class (recall score 0.69). The lightning class displays overall the best classification metrics (F1-score of 0.88). The precision score (0.85) is fairly lower than the recall score (0.91), suggesting that lightning fires are easier for the model to identify and are not confused with another class. On the contrary, the model performs the worst for the private negligence class, with an F1-score of 0.55. In this class, the precision score (0.59) is higher than the recall score (0.52), which suggests that the classifier has a higher rate of correctly identifying positive samples but is missing a higher proportion of the total number of positive samples. Finally, the professional negligence class exhibits relatively low but balanced scores between precision (0.67) and recall (0.63).

**Table 10 Classification metrics per ignition cause.**

<b>Ignition cause</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>Accuracy</b>
Accidental	0.81	0.74	0.77	69.8%
Arson	0.60	0.69	0.64	
Lightning	0.85	0.91	0.88	
Private negligence	0.59	0.52	0.55	
Professional negligence	0.64	0.68	0.65	

The confusion matrix (Figure 33) provides additional information with regards to the performance of the classification of ignition causes. More specifically, accidental fires are most frequently misclassified as arson ones. There is a high number (n=159) of arson fires that are wrongly classified as private negligence, and similarly, there are 266 private negligence fires that are misclassified as arson. This could mean that there are similarities between the causes of these fires, or that the model may not have enough information to accurately distinguish between these classes. As the most accurately predicted cause, lightning displays low misclassification numbers, which are distributed evenly among the other classes. In contrast, private negligence, that is a major negative contributor to the overall classification accuracy,

shares its errors primarily between professional negligence and arson classes. Finally, professional negligence fires are also often confused for either arson or private negligence fires.



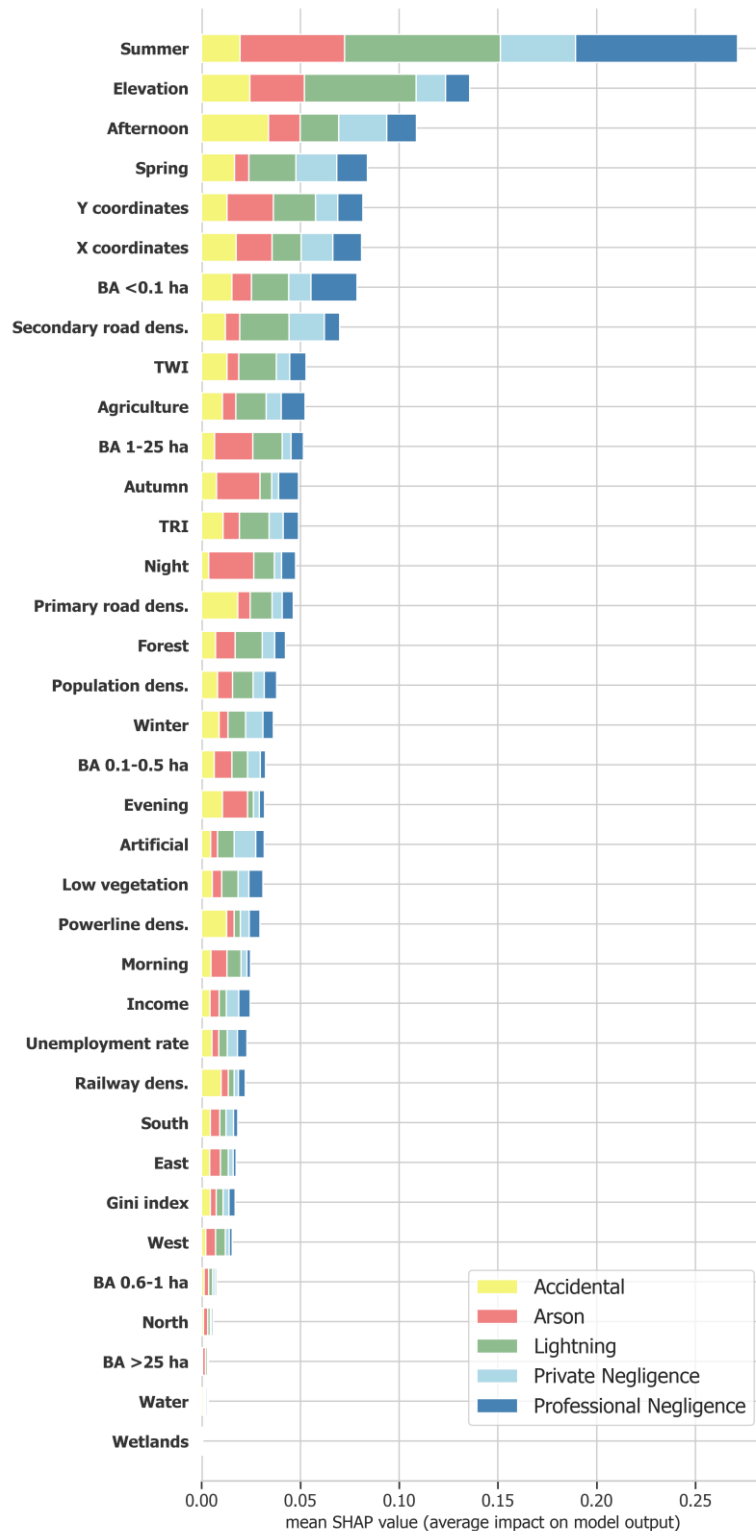
**Figure 33 Confusion matrix.**

#### 4.5.2 Feature importance and effect

Figure 34 illustrates the computed feature importance of the RF model for all classes, which is expressed through mean SHAP values that represent the average impact of a feature on the model output across all the instances in the dataset. Overall, feature importance values vary significantly both per feature type and ignition cause. The features summer, elevation and afternoon form a group that stands out significantly from the rest, followed by a second cluster with slightly less impactful features such as spring, geographic coordinates, BA <0.1 ha and secondary road density.

In the context of accidental fires, several features demonstrate comparable significance, with afternoon, elevation, summer, and primary road density being slightly more salient than other

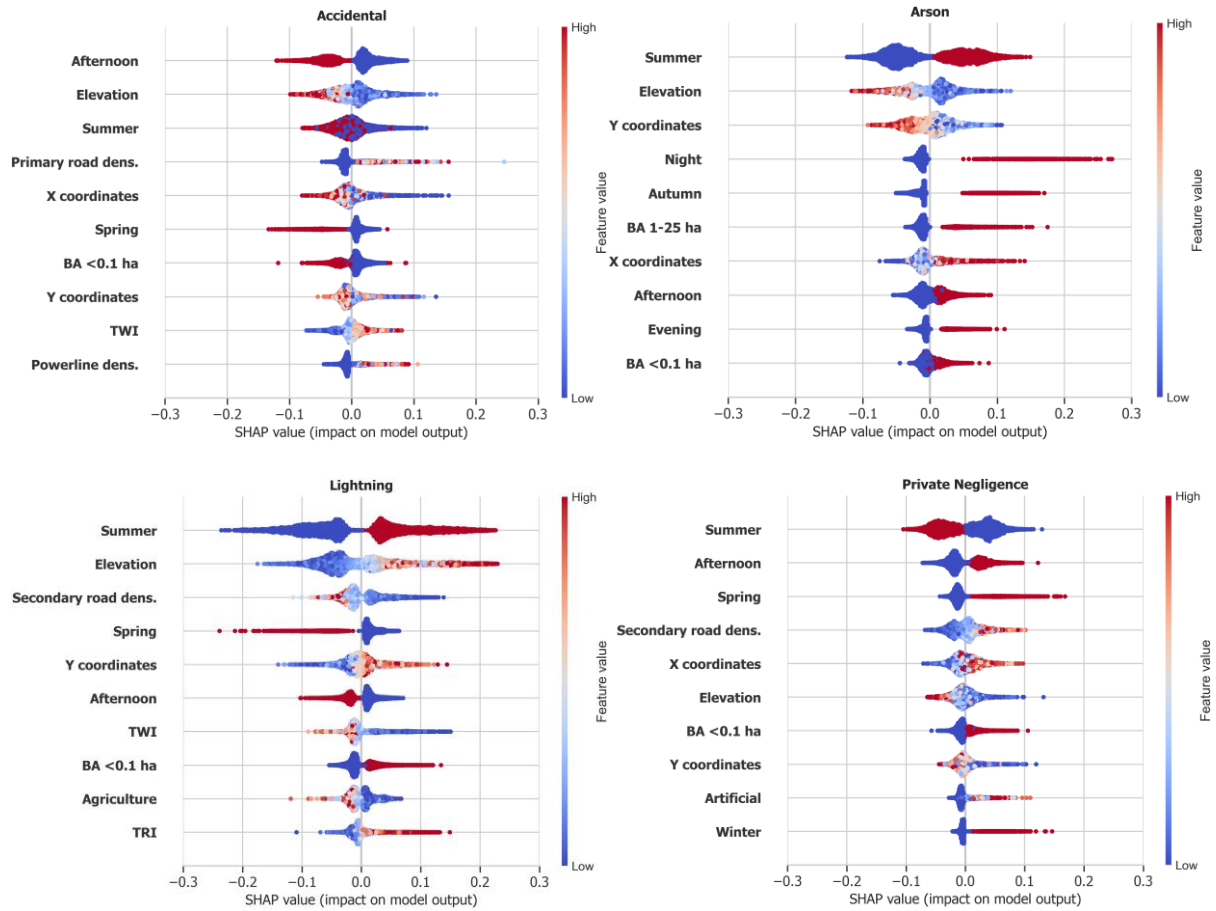
variables. Similarly, the relevance of features for arson fires is widely distributed, with spatiotemporal characteristics such as summer, night, and location being the most prominent factors. Regarding fires caused by lightning, summer and elevation are by far the most impactful variables followed by secondary road density. In the case of private negligence, summer exhibits the highest level of importance, although this distinction is not substantially greater than that of other variables, such as afternoon, spring, and secondary road density. Finally, with respect to professional negligence, summer represents the most influential factor by a significant margin, with only BA size (<0.1 ha) showing discernible differences from other variables.



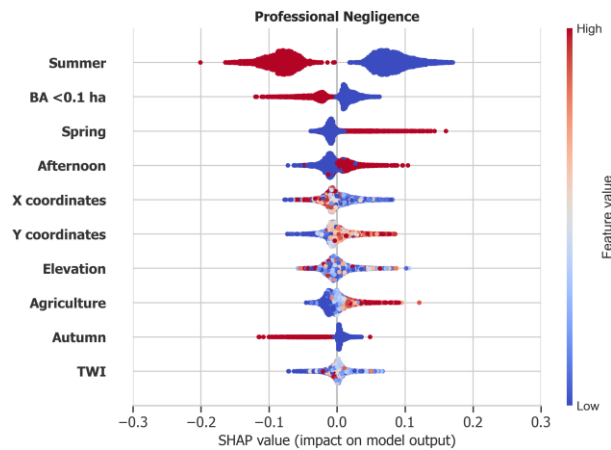
**Figure 34 Feature importance (mean SHAP values) for all classes.**

Figure 35 depicts the most influential (n=10) features for each class of the model in descending order. Furthermore, the impact of each feature on the ignition cause is also illustrated through the positive or negative SHAP values. These values indicate whether an instance is more or

less likely to belong to a particular class depending on the magnitude of the feature values. For example, instances with lower elevation values are more likely to be classified as an accidental or arson ignition, and less likely to be categorized as a lighting fire. Similarly, a fire that occurred during the summer is more likely to belong to the arson or lightning class, but less probable to be classified into the other categories.







**Figure 35 Importance and effect of the most influential variables (n=10) for each ignition cause. The color 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 center line) indicate a higher probability of belonging to a specific class, while negative values (left of the center line) indicate lower probability of belonging to a specific class.**

#### 4.6 Discussion

The performance of the RF classifier varies considerably between natural and human-induced fires. Lightning fires were classified with the highest accuracy since ignition dynamics for these fires are significantly different from human-caused fires. As reported by Curt et al., (2016), lightning fires tend to have small, burned areas, occur on steep, densely vegetated, mountainous slopes with low anthropogenic presence; seasonality also plays a significant role in the incidence of those fires (summer). This particular profile, which aligns with the interpretation of features effects through the SHAP values, enables the classifier to distinguish it from other causes more clearly.

In contrast to natural fires, human-caused ignitions are multi-faceted and more complex to model. Accidental fires are the least difficult human induced events to classify in our model, potentially attributed to the greater association of such fires with infrastructure elements such as powerlines and railways in contrast to other forms of anthropogenic causes. The most challenging cause to classify is private negligence, which is most often misclassified as arson and vice versa. Both arson and private negligence fires often occur in similar contexts,

specifically the wildland urban interface. The similarity in environmental contexts and conditions between these types of fires may make it difficult to distinguish between the two causes. However, this may also reflect a problem of reliability in the fire databases (Ganteaume and Guerra, 2018): in order to reduce the number of unknown caused fires the cause is either speculated or attributed without much physical evidence to support it (Camia et al., 2013). Professional negligence fires are also confused, but to a lesser extent, with private negligence. Both causes share common characteristics, as they tend to burn small/medium areas and occur mainly outside of the summer season (Curt et al., 2016) which is reflected in the significance and impact those features hold in the SHAP framework.

Socioeconomic data used in our model only pertains to a single year. While this approach may have its advantages, such as simplifying data collection and analysis, it can also potentially undervalue the importance of socioeconomic features by not capturing their temporal fluctuations, especially considering that most fires in France but also in the Euro-Mediterranean (95 %) region are caused by humans (Ganteaume et al., 2013; Ganteaume & Jappiot, 2013). Factors such as population density, unemployment rate, etc. represent dynamic phenomena that can change considerably over time in contrast to static variables such as topography or even to other dynamic variables as, for instance, land cover. The addition of geographic coordinates in our workflow not only partly tackles spatial non-stationarity, as the decision trees of the model in a way incorporate geographic space during their creation, but also enhances the results which is in accordance with other works that utilize ML algorithms for applications of spatial nature (Hengl et al., 2018). Spatial approaches of ML algorithms such as Geographic Random Forests (Georganos et al., 2021) and Geographically Weighted Neural Networks (Hagenauer and Helbich, 2022) would be advantageous for such applications considering the significance of spatial location and its strong links with different fire ignition causes.

As a first attempt, the current study utilized only the first-level causes (5 categories) available from the hierarchical structure of the “Promethee” fire database, which also includes second-level (15 categories) and third-level (31 categories) causes. Applying a similar procedure on selected sub level data could possibly improve functionality and understanding of ignition sources and their performance within the classification scheme. However, this would increase the complexity of the model and may negatively impact overall accuracy. Finally, the inclusion of fuel type characteristics and fire-weather variables can potentially strengthen and facilitate the distinction between different fire causes; for instance, arson fires burn larger areas (Ganteaume and Jappiot, 2013; Syphard and Keeley, 2015) and this may indicate these fires are set under more favorable weather conditions.

The practicality of this model is not intended for operational use or as a substitute method to conventional field investigation methods as it cannot provide physical evidence for the proper deduction of the cause of a forest fire. Instead, it is targeted as a method to analyze large-scale fire databases that contain a moderate percentage of unknown caused fires. The ideal balance would be neither too low, as insufficient data would result in a restricted training dataset, nor too high, as that would render the model less useful. Despite the limitations in identifying causes of unknown ignitions, the results can help to facilitate targeted prevention efforts (Oliveira et al., 2012b). Moreover, the benefits of harmonized classification systems, such as the one proposed by the European Commission (Camia et al., 2013) are emphasized. By utilizing such schemes, ML models can significantly benefit from increased harmonized data availability, provided that the data's reliability stays at an adequate level. This allows for the combination of historical national fire databases, leading to the development of larger databases with enhanced modeling potential. Other standardized georeferenced data initiatives in Europe (e.g., the European INSPIRE Directive, the European Geodata Infrastructure, etc.)

which aim to establish a common framework for the management and sharing of geospatial data across Europe are going in this direction.

ML algorithms have become increasingly popular in fire science (Jain et al., 2020). These algorithms can help identify complex relationships between various factors that contribute to fire occurrences. However, the success of these algorithms relies heavily on the availability of large, high-quality datasets. As fire science continues to advance, access to larger and more comprehensive datasets is becoming increasingly common. This includes georeferenced explanatory feature data which provides important contextual information that can be used to better understand the underlying causes of fires. As these datasets continue to grow in size and quality, ML algorithms will become even more powerful tools for analyzing fire occurrences and fire causes.

#### **4.7 Conclusion**

In this study we train and apply a model to classify fire ignition causes based on several environmental and anthropogenic features using an explainable artificial intelligence framework. The results suggest that the source of unknown caused fires can be identified at various levels of accuracy depending on the nature of the forest fire (e.g., F1-score lightning 0.87, accidental 0.74, arson 0.64). Spatiotemporal characteristics including geographic location, season, time of the day but also topographic factors like elevation are the most important features in determining the classification of unknown caused fires for the specific area and fire regime studied here. The role of spatial non-stationarity is highlighted through the importance it holds in our processing framework and should be treated by implementing models that utilize spatial approaches of machine learning algorithms, which are expected to have increased accuracy over the original ones. The increasing availability of large, high-quality datasets is an important factor driving the growth of ML algorithms in wildfire science

and will likely play a critical role in advancing our understanding of fire causes in the coming years.

# General Discussion

This thesis aimed to provide a deeper understanding of forest fire spatiotemporal patterns, ignition causes, and their associated factors in southern France using geospatial data analysis and ML techniques. This discussion chapter is divided into three sections: 1) Spatiotemporal changes of burned area in SE France in the context of climate change, 2) The role of ML in determining fire ignition causes: data challenges and Explainable AI and 3) Leveraging geospatial data and ML to advance wildfire research.

## **Part 1 : Spatiotemporal changes of burned area in SE France in the context of climate change**

The analysis of forest fire dynamics over a 50-year period (1970-2019) revealed substantial changes in the spatial distribution and patterns of BA after the implementation of the new fire suppression strategy in the early 1990s. Apart from the sharp decrease in total BA, the impact of the new firefighting strategy can be observed spatially as fire patches have generally become smaller, are distributed more closely to one another, and display lower fire recurrence. Spatiotemporal trends of BA differ regionally as in the western part of the study zone fire hotspots persist although limited in size. The new firefighting strategy has had a limited impact since fires were already restricted in size by vegetation continuity which can determine BA (Duane et al., 2015; Fernandes et al., 2016). Housing density is greatest in the Bouches du Rhône department, so shrublands and forest are fragmented by continuous and discontinuous urban areas and some agriculture. Despite being limited in size, multiple clusters of positive trends are found in closer proximity to the built-up areas. This observation supports the well-

established idea that human activity in urban areas can increase the risk of wildfire ignition (Badia et al., 2011; Chas-Amil et al., 2013; Jiménez-Ruano et al., 2017; Lampin-Maillet et al., 2011). In contrast, the central part of the study area, where most large fires occurred, saw the new fire policy effectively limit fire spread across the continuous vegetated cover characterizing the region. The findings of Ganteaume & Barbero (2019) align with these observations as the authors provide evidence that large fires (>100 ha) significantly declined after 1991. In the eastern segment of the study area, although negative fire trends are prevalent, particularly in the WUI zone, there is a higher number of small positive trend patches compared to other areas in the eastern portion of the study zone. Numerous small and dispersed fire patches are observed that appear less shaped/elongated by wind direction than in the western and central regions (Ruffault et al., 2018). Although wind speed or direction is not analyzed here, during the course of this thesis, an investigation into weather variables, including rainfall, wind, temperature, and relative humidity, was conducted. However, due to time constraints, it was not possible to integrate this part into the thesis. As a result, a separate paper is currently being prepared that will incorporate weather variables alongside the environmental and anthropogenic variables discussed in the preceding chapters.

The results revealed that the contribution of south-facing slopes to forest fires has increased over time, likely exacerbated by ongoing climate change, as south-facing aspects display greater insolation and evapotranspiration than other orientations. Other studies (e.g., Mouillot et al., 2003) have presented evidence that south-facing slopes S-facing slopes have the greatest BA and burn more frequently however, the findings of this thesis indicate that this trend is strengthening over time, a phenomenon that has not been previously documented by others. Finally, S-facing slopes in the study zone tend to have shallower soils than N-facing slopes (Fox, 2011) so soil water reserves are lowest, and they are therefore probably the first slopes to dry out and remain drier for longer than other aspects.

The burnt vegetation patterns observed in this study confirm the well-established importance of shrublands in forest fires, as reported in previous research (Ganteaume & Jappiot, 2013; Moreira et al., 2011; Oliveira et al., 2014; Tessler et al., 2016). Shrublands tend to facilitate fire propagation in dry conditions (Baeza et al., 2002) and both induce and result from recurrent fires (Tessler et al., 2016). As a result, the susceptibility of shrublands to fires may increase in the future due to hotter and drier conditions driven by climate change. That is especially relevant in the region since sclerophyllous vegetation covers the greatest area, represents around half of the total BA, regenerates and expands faster than other vegetation types.

Climate change is expected to alter fire patterns substantially, making the analysis of spatio-temporal changes increasingly crucial for predicting how these patterns will evolve in the context of changing climate. As global temperatures rise and extreme weather events, including droughts, become more frequent, we can expect fire frequency, intensity and duration to increase (Bowman et al., 2011; Keeley & Syphard, 2016). In addition, areas with historical fire susceptibility are expected to become more fire prone. This was evident during the summer of 2022, when large areas outside of the French Mediterranean zone were burned by wildfires. It is important to keep this in mind as AI models trained on historical wildfire data may lead to biased results, such as by over-predicting fire occurrences in areas with a known history of wildfires, potentially overlooking other areas that may have a greater risk. This underscores the importance of properly understanding and anticipating the spatial and temporal dynamics of wildfires to facilitate better management and mitigation strategies.

In order to effectively analyze and predict the spatiotemporal changes in fire patterns influenced by climate change, it is essential to have access to high-quality fire geodatabases that can facilitate the needs of such applications. Advancements in satellite technology and data processing techniques have also contributed to significant improvements in geospatial data quality. This includes improvements in spatial resolution, spectral resolution, and radiometric



resolution, allowing for more precise and accurate detection and mapping of burned areas, ignition points, and fire behavior (Chuvieco, 2020). For instance, moderate- to high-resolution satellite imagery, such as Landsat or Sentinel, is widely used to map burned areas and estimate the extent of wildfire events. These data are typically available at a spatial resolution of 10-30 m, which allows researchers to identify individual burn scars and estimate their size, shape, and location with high precision.

The fire geodatabase utilized in this thesis, provided by the National Forestry Office (ONF) covered a 50-year time span. Due to the long temporal extent of the database, the accuracy and the methods used to define burn scars significantly improved over time; with the advent of satellite imagery, burn scars became more detailed and accurate in shape compared to earlier polygons which had coarse shapes and lacked detail. Despite having high geometric accuracy, the specific database used in this thesis has not been used by any other researchers with a single exception (Ganteaume & Barbero 2019). The vast majority of forest fire research in France is based on the “Prométhée” database that contains information for each fire such as total burned area and ignition location within a much coarser 2x2 km grid. The particularly limited use of the ONF database can be attributed to the fact that the total number of recorded fires is significantly lower than in Prométhée and that, unlike Prométhée, it is not publicly accessible via a web portal. Although the total number of fires is substantially lower, the total area burned is almost identical. The difference in the number of fires recorded in the ONF database and the “Prométhée” database is possibly due to the fact that the ONF database provides the exact fire perimeter, which relies on the available technology of each period, while the “Prométhée” database requires only an approximate area of a fire. As a result, very small fires recorded in “Prométhée” may not be digitized in the ONF database. As earth observation satellites and drones continue to acquire data at higher temporal and spatial resolutions (Lippitt & Zhang, 2018), the difference between the two is expected to become less significant in the future,

eventually making the ONF database the preferred option. Nonetheless, in the context of mapping regional spatiotemporal changes in BA, the need for geometric accuracy outweighs the need to include every fire recorded in the region, and thus the ONF database is better suited for such applications.

As we continue to confront the challenges posed by climate change, the need for robust geodatabases that can support the analysis and forecasting of spatiotemporal wildfire dynamics becomes increasingly vital for effective wildfire management and mitigation efforts. By analyzing the spatiotemporal relationships between fire activity and environmental factors, researchers and policymakers will be better equipped to develop adaptive strategies that can effectively respond to the evolving risks posed by wildfires in a world increasingly affected by climate change. However, increasing complexity geospatial data analysis and advanced techniques such as ML, can widen the gap between fundamental science and practical land management. As such practices become more widespread, it is essential to develop effective communication strategies that bridge the gap between complex methodologies and practical applications in firefighting and land management. This may involve creating user-friendly tools and visualizations that clearly demonstrate the insights gained from ML models. Over time, as the firefighting and land management communities become more familiar with geospatial and ML techniques, the lessons learned from these advanced models will likely become more accessible and readily applicable. This will ultimately facilitate better decision-making, improved wildfire risk prediction, and more effective land management practices, benefiting both the scientific community and practitioners on the ground.

## **Part 2 : The role of ML in determining fire ignition causes: data challenges and Explainable AI**

The application of ML and geospatial data in predicting fire ignition causes was investigated using various spatial scales, fire databases, and feature evaluation techniques. In the first effort to model fire ignition source dynamics (Chapter 3), a dataset containing precise coordinates of fire ignitions was employed. The developed model was capable of distinguishing between arson and non-arson fires with limited accuracy. In addition, while the model provided information on the environmental and anthropogenic features influencing its predictions, it did not offer a clear explanation of how each feature impacted the classification of ignition sources. The primary drawback of this first attempt was that although the ignition points in the dataset had high spatial resolution, their limited quantity led to an unstable model performance. This was due to the short duration of the records and also to the fact that data were either not publicly accessible or the ignition coordinates were not recorded. Efforts made to obtain additional ignition points from the firefighting authorities were unsuccessful, possibly due to the lack of transparency in data sharing or changes in data collection practices over time. Regardless, it is crucial to address these limitations to enhance the predictive capabilities of ML and geospatial analysis in determining fire ignition causes.

Despite these significant limitations, this initial work on fire ignition causes laid the groundwork for creating a larger, more complex, and interpretable model. The second ML-based model, described in Chapter 4, utilized a vastly larger number of ignitions points but with poor spatial resolution, derived from the “Promethee” fire database. Having large and diverse datasets, but also the means to analyze them, is inherently essential in data centric approaches such as ML since they learn by finding patterns and relationships in the data

(Bishop & Nasrabadi, 2006) and help improve the accuracy and generalization ability of the models. The more data a ML model has access to, the better it can learn and make accurate predictions, and the better it can generalize new instances and reduce the risk of overfitting where the model memorizes the training data instead of learning to generalize to new data.

The performance of the classifier varied depending on the cause where lighting fires had the highest accuracy due to their distinct ignition dynamics and environmental contexts. In contrast to natural fires, human-caused ignitions are more complex and challenging to model, especially for private negligence fires which were often misclassified as arson and vice versa. The common environmental contexts and WUI characteristics between private negligence and arson fires made it difficult for the model to differentiate when using the chosen set of features. This issue might also point to a problem of reliability in fire databases (Ganteaume and Guerra, 2018), as causes are occasionally assumed without enough supporting physical evidence in an effort to decrease the number of fires with undetermined sources. (Camia et al., 2013).

Challenges related to both data quantity and quality can be addressed through data harmonization. Implementing harmonized classification systems, such as the one proposed by the European Commission (Camia et al., 2013), facilitates the merging of historical national fire databases, leading to the development of larger databases with enhanced modeling potential. Additionally, these systems provide clearer guidelines and definitions for each cause and contribute to more accurate and consistent data across different sources.

The research presented in this thesis tackled a significant challenge associated with the "black box" nature of ML-based models. By incorporating the XAI framework, the study not only enhanced the findings through the interpretation of feature effects using SHAP values, it also enabled comparisons with other studies, such as the study conducted by Curt et al. (2016). XAI is an emerging field in AI that seeks to enhance the transparency and accountability of complex ML models, and this is particularly relevant in the context of wildfire science. In fact, the

integration of such frameworks has only recently been adopted by researchers in their processing pipeline (Abdollahi & Pradhan, 2023; Al-Bashiti & Naser, 2022; Bountzouklis et al., 2023; Cilli et al., 2022; Khanmohammadi et al., 2023; Shmuel & Heifetz, 2023). Future research of ML applications in wildfire science should include such frameworks as by providing insights into the underlying factors that influence model predictions, these techniques can identify potential biases or errors in their outputs and improve the accuracy and reliability of wildfire risk assessments.

## **Part 3 : Leveraging geospatial data and ML to advance wildfire research**

The use of ML algorithms is becoming increasingly common in wildfire science (Bot & Borges, 2022; Jain et al., 2020). It is noteworthy that wildfire science relies on and benefits from a diverse set of relevant information sourced from numerous domains related to topography, vegetation, climate, land cover, soil, socio-economic characteristics, policies and firefighting resources as well as WUI characteristics. Given the significance of processing and analyzing large amounts of data from multiple sources in this context, it is evident that ML algorithms hold substantial value and potency for wildfire research. However, the increasing volume and complexity of geospatial data can pose challenges for researchers in terms of accessing, processing, and sharing this data effectively. This can be addressed through the use of cloud computing which, combined with ML, enables researchers to address critical issues in various fields, including natural hazards risk management (Yang et al., 2022). Recently, there has been a notable surge in the availability of cloud computing platforms specifically designed for geospatial analysis. For instance, Google Earth Engine (Gorelick et al., 2017), Microsoft-Planetary Computer (McFarland et al., 2022), and Amazon-Earth Amazon Web

Services have each launched cloud-based platforms that provide access to powerful geospatial data processing and analysis tools.

Moreover, the intersection between ML and geographic information science presents unique challenges and characteristics that must be addressed appropriately. Geospatial data are inherently distinct from non-spatial data due to their unique properties and can influence the performance of ML algorithms. These include spatial dependence, spatial heterogeneity and spatial scale which are often overlooked or insufficiently managed (Nikparvar & Thill, 2021). As described below, these are important to take into account when analyzing geospatial data or developing models of spatial processes. Spatial dependence refers to the fact that events or phenomena that occur in one geographic location can influence or be influenced by events or phenomena that occur in nearby locations. This fundamental concept is described through Tobler's first law of geography (Tobler, 1889), which states that "everything is related to everything else, but near things are more related than distant things". This is important to take it into account when analyzing geospatial data or developing models of spatial processes.

Spatial heterogeneity refers to the variation in human or physical characteristics or attributes within a given geographic area (Dutilleul & Legendre, 1993), such as climate, topography, land cover, and population density. Different areas within a geographic region may have different conditions, which can have implications for various processes that occur within those areas, such as wildfire BA. In other words, the assumption that distance between two points or units can infer a stationary spatial process may not always hold true due to the presence of non-stationary processes that arise from different geographic processes.

The scale of analysis refers to the geographic extent, and resolution of the data and consequently the geographic areas of the training and test datasets used in an analysis (Goodchild, 2011). When using different scales there are key issues that are associated with the use of areal units: i) the sensitivity of analytical results to the definition of geographic units

for which data are collected and ii) the sensitivity of contextual variables and analytical results to different delineations of contextual units (Openshaw, 1984). For instance, if the geographic units used for analysis are too large, the spatial heterogeneity within the units may not be accounted for, leading to inaccurate results. Conversely, if the units used are too small, they may not capture the full extent of spatial patterns, leading to overfitting and reduced generalizability. Additionally, contextual units, such as administrative boundaries, are used to aggregate data, and different aggregations can lead to different analytical results. Such observations have been made in the context of forest fire research as a given factor may vary in importance according to the scale of the study (Ganteaume & Long-Fournel, 2015; Laforteza et al., 2013; Moritz et al., 2005).

Considering the exponential increase of geospatial data, the rapid upsurge in popularity of AI and its subdomains, it is legitimate to acknowledge that there is an emerging scientific discipline, that of Geospatial Artificial Intelligence which combines knowledge and innovations in geography, spatial analysis, earth observation, AI methods (e.g., ML and DL), data mining, and high-performance computing to extract knowledge from geospatial big data and utilize them in natural hazards and risk management. Together, these topics represent some of the most promising future developments and are directly linked to wildfire science and management, and they hold great potential for further improving our ability to monitor, predict, and manage wildfire events in the future.





# General Conclusion

This Ph.D. thesis has contributed to a deeper understanding of forest fire spatiotemporal patterns, ignition causes, and their associated factors in southern France using geospatial data analysis and ML techniques.

The spatiotemporal dynamics and patterns of forest fires over a 50-year period (1970-2019) were analyzed, taking into consideration the rapid shift in fire suppression policy in the early 1990s. This analysis provided insights into the effectiveness of these strategies in reducing BA and changing the geographical distribution of fires. Many large fire hotspots have disappeared following the policy change, proving the effectiveness of fast fire suppression in limiting large fires in the region. However, new smaller hotspots emerged in closer proximity to areas with increased human activity. The examination of fire selectivity and the spatiotemporal relationship of BA with respect to environmental factors such as topography and vegetation type has changed over time. The findings emphasized the strong association of sclerophyllous vegetation and natural grasslands, with increased fire susceptibility and recurrence, and the growing impact of south-facing slopes over time.

Despite the scarcity of research on the direct influence of fire suppression strategies on BA's spatial distribution, geospatial data and analysis have been invaluable in quantifying spatiotemporal patterns and driving forces impacting BA distribution. This information is crucial for fire and civil protection agencies to allocate firefighting resources effectively and develop suitable prevention measures. Additional analysis of fire geodatabases, combined with WUI characteristics, will likely improve our comprehension of the factors affecting BA and the implications of firefighting tactics in the region.

Fire suppression, while crucial, is complemented by proactive measures like understanding and documenting forest fire ignition causes, which can be more cost-effective and sustainable in

the long term. This thesis illustrates the potential of predicting unknown ignition sources using anthropogenic and environmental features and highlights the benefits of integrating ML and XAI frameworks for comprehending and explaining the impact of different factors on each ignition cause. Predicting ignition sources yields varying accuracy levels, with natural fires showing the highest accuracy compared to human-caused fires like accidental and arson. As the first attempt to utilize an XAI framework for predicting fire ignition causes, this research establishes a new benchmark for similar studies in wildfire science and offers valuable insights for future research. In addition to the increasing availability of large, high-quality datasets, harmonized data collection is an essential factor in driving the growth of ML algorithms in wildfire science and will likely play a critical role in advancing our understanding of fire ignition causes in the coming years.

While the results of this research are promising, there is still room for improvement in the accuracy of the predictive models. Future studies may explore the use of spatial extensions of ML algorithms to enhance the performance of these models. Additionally, this research has focused on specific geographical areas and fire regimes, which may not be directly applicable to other regions or ecosystems. Future research could investigate the applicability of the developed methodologies and models to different spatial scales and fire regimes, broadening the understanding of wildfire dynamics in various contexts.

As our understanding of spatiotemporal forest fire dynamics grows and the availability of high-quality geospatial data increases, it is anticipated that the application of advanced geospatial analysis and ML techniques will continue to contribute significantly to the development of more effective wildfire management and prevention strategies. Despite the limitations, the findings of this work provide valuable insights for future research and contribute to the development of more effective wildfire management strategies that will ultimately benefit ecosystems and human welfare.



# References

- Abdollahi, A., & Pradhan, B. (2023). Explainable artificial intelligence (XAI) for interpreting the contributing factors feed into the wildfire susceptibility prediction model. *Science of The Total Environment*, 879, 163004. <https://doi.org/10.1016/j.scitotenv.2023.163004>
- Akaike, H. (1998). *Information Theory and an Extension of the Maximum Likelihood Principle* (pp. 199–213). [https://doi.org/10.1007/978-1-4612-1694-0\\_15](https://doi.org/10.1007/978-1-4612-1694-0_15)
- Al-Bashiti, M. K., & Naser, M. Z. (2022). Machine learning for wildfire classification: Exploring blackbox, eXplainable, symbolic, and SMOTE methods. *Natural Hazards Research*, 2(3), 154–165. <https://doi.org/10.1016/j.nhres.2022.08.001>
- Alexander, J. D., Seavy, N. E., Ralph, C. J., & Hogoboom, B. (2006). Vegetation and topographical correlates of fire severity from two fires in the Klamath-Siskiyou region of Oregon and California. *International Journal of Wildland Fire*, 15(2), 237. <https://doi.org/10.1071/WF05053>
- Andela, N., Morton, D. C., Giglio, L., Paugam, R., Chen, Y., Hantson, S., van der Werf, G. R., & Randerson, J. T. (2018). The Global Fire Atlas of individual fire size, duration, speed, and direction. *Earth System Science Data Discussions*, August 2018, 1–28. <https://doi.org/10.5194/essd-2018-89>
- Anderson, H. E. (1970). Forest fuel ignitibility. *Fire Technology*, 6(4), 312–319. <https://doi.org/10.1007/BF02588932>
- Artés, T., Oom, D., de Rigo, D., Durrant, T. H., Maianti, P., Libertà, G., & San-Miguel-Ayanz, J. (2019). A global wildfire dataset for the analysis of fire regimes and fire behaviour. *Scientific Data*, 6(1), 296. <https://doi.org/10.1038/s41597-019-0312-2>
- Ba, R., Chen, C., Yuan, J., Song, W., & Lo, S. (2019). SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention. *Remote Sensing*, 11(14), 1702. <https://doi.org/10.3390/rs11141702>
- Badia, A., Serra, P., & Modugno, S. (2011). Identifying dynamics of fire ignition probabilities in two representative Mediterranean wildland-urban interface areas. *Applied Geography*, 31(3), 930–940. <https://doi.org/10.1016/j.apgeog.2011.01.016>
- Baeza, M. J., De Luís, M., Raventós, J., & Escarré, A. (2002). Factors influencing fire behaviour in shrublands of different stand ages and the implications for using prescribed burning to reduce wildfire risk. *Journal of Environmental Management*, 65(2), 199–208. <https://doi.org/10.1006/jema.2002.0545>
- Bajocco, S., & Ricotta, C. (2008). Evidence of selective burning in Sardinia (Italy): Which land-cover classes do wildfires prefer? *Landscape Ecology*, 23(2), 241–248. <https://doi.org/10.1007/s10980-007-9176-5>
- Ban, Y., Zhang, P., Nascetti, A., Bevington, A. R., & Wulder, M. A. (2020). Near Real-Time Wildfire Progression Monitoring with Sentinel-1 SAR Time Series and Deep Learning. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-019-56967-x>
- Barros, A. M. G., & Pereira, J. M. C. (2014). Wildfire selectivity for land cover type: Does size matter? *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0084760>
- Bates, B. C., Dowdy, A. J., & Chandler, R. E. (2017). Classification of Australian Thunderstorms Using Multivariate Analyses of Large-Scale Atmospheric Variables. *Journal of Applied Meteorology and Climatology*, 56(7), 1921–1937. <https://doi.org/10.1175/JAMC-D-16-0271.1>
- Beer, T. (1991). The interaction of wind and fire. *Boundary-Layer Meteorology*, 54(3), 287–308. <https://doi.org/10.1007/BF00183958>

- Birch, D. S., Morgan, P., Kolden, C. A., Abatzoglou, J. T., Dillon, G. K., Hudak, A. T., & Smith, A. M. S. (2015). Vegetation, topography and daily weather influenced burn severity in central Idaho and western Montana forests. *Ecosphere*, 6(1). <https://doi.org/10.1890/ES14-00213.1>
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, Issue 4). Springer.
- Bond, W., & Keeley, J. (2005). Fire as a global ‘herbivore’: the ecology and evolution of flammable ecosystems. *Trends in Ecology & Evolution*, 20(7), 387–394. <https://doi.org/10.1016/j.tree.2005.04.025>
- Bot, K., & Borges, J. G. (2022). A Systematic Review of Applications of Machine Learning Techniques for Wildfire Management Decision Support. In *Inventions* (Vol. 7, Issue 1). MDPI. <https://doi.org/10.3390/inventions7010015>
- Bouillon, C., Coughlan, M., Fernandes, J. R., Amraoui, M., Chamusca, P., Madureira, H., Parente, J., & Pereira, M. G. (2020). The relation of landscape characteristics, human settlements, spatial planning, and fuel management with extreme wildfires. In *Extreme Wildfire Events and Disasters* (pp. 73–90). Elsevier. <https://doi.org/10.1016/b978-0-12-815721-3.00004-7>
- Bountzouklis, C., Fox, D. M., & Bernardino, E. Di. (2023). Predicting wildfire ignition causes in Southern France using eXplainable Artificial Intelligence (XAI) methods. *Environmental Research Letters*, 18(4), 044038. <https://doi.org/10.1088/1748-9326/acc8ee>
- Bountzouklis, C., Fox, D. M., & Di Bernardino, E. (2022a). Fire cause classification of undetermined fires in southeastern France. In D. X. Viegas & L. M. Ribeiro (Eds.), *Advances in Forest Fire Research 2022* (pp. 1106–1112). Imprensa da Universidade de Coimbra. [https://doi.org/10.14195/978-989-26-2298-9\\_167](https://doi.org/10.14195/978-989-26-2298-9_167)
- Bountzouklis, C., Fox, D. M., & Di Bernardino, E. (2022b). Environmental factors affecting wildfire-burned areas in southeastern France, 1970–2019. *Natural Hazards and Earth System Sciences*, 22(4), 1181–1200. <https://doi.org/10.5194/nhess-22-1181-2022>
- Bowman, D. M. J. S., Balch, J., Artaxo, P., Bond, W. J., Cochrane, M. A., D’Antonio, C. M., DeFries, R., Johnston, F. H., Keeley, J. E., Krawchuk, M. A., Kull, C. A., Mack, M., Moritz, M. A., Pyne, S., Roos, C. I., Scott, A. C., Sodhi, N. S., & Swetnam, T. W. (2011). The human dimension of fire regimes on Earth. *Journal of Biogeography*, 38(12), 2223–2236. <https://doi.org/10.1111/j.1365-2699.2011.02595.x>
- Bowman, D. M. J. S., Williamson, G. J., Abatzoglou, J. T., Kolden, C. A., Cochrane, M. A., & Smith, A. M. S. (2017). Human exposure and sensitivity to globally extreme wildfire events. *Nature Ecology and Evolution*, 1(3). <https://doi.org/10.1038/s41559-016-0058>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification And Regression Trees*. Routledge. <https://doi.org/10.1201/9781315139470>
- Broncano, M. J., & Retana, J. (2004). Topography and forest composition affecting the variability in fire severity and post-fire regeneration occurring after a large fire in the Mediterranean basin. *International Journal of Wildland Fire*, 13(2), 209–216. <https://doi.org/10.1071/WF03036>
- Calviño-Cancela, M., Chas-Amil, M. L., García-Martínez, E. D., & Touza, J. (2016). Wildfire risk associated with different vegetation types within and outside wildland-urban interfaces. *Forest Ecology and Management*, 372, 1–9. <https://doi.org/10.1016/j.foreco.2016.04.002>

- Camia, A., San-Miguel-Ayanz, J., & Durrant, T. (2013). Harmonized classification scheme of fire causes in the EU adopted for the European Fire Database of EFFIS. *Publications Office*. <https://doi.org/10.2788/86859>
- Capra, G. F., Tidu, S., Lovreglio, R., Certini, G., Salis, M., Bacciu, V., Ganga, A., & Filzmoser, P. (2018). The impact of wildland fires on calcareous Mediterranean pedosystems (Sardinia, Italy) – An integrated multiple approach. *Science of The Total Environment*, 624. <https://doi.org/10.1016/j.scitotenv.2017.12.099>
- Carmo, M., Moreira, F., Casimiro, P., & Vaz, P. (2011). Land use and topography influences on wildfire occurrence in northern Portugal. *Landscape and Urban Planning*, 100(1–2), 169–176. <https://doi.org/10.1016/j.landurbplan.2010.11.017>
- Carrega, P. (1991). A Meteorological Index of Forest Fire Hazard in Mediterranean France. *International Journal of Wildland Fire*, 1(2), 79. <https://doi.org/10.1071/WF9910079>
- Catarino, S., Romeiras, M. M., Figueira, R., Aubard, V., Silva, J. M. N., & Pereira, J. M. C. (2020). Spatial and temporal trends of burnt area in angola: Implications for natural vegetation and protected area management. *Diversity*, 12(8). <https://doi.org/10.3390/D12080307>
- Catry, F. X., Rego, F. C., Bação, F. L., & Moreira, F. (2009). Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire*, 18(8), 921–931. <https://doi.org/10.1071/WF07123>
- Chas-Amil, M. L., Touza, J., & García-Martínez, E. (2013). Forest fires in the wildland-urban interface: A spatial analysis of forest fragmentation and human impacts. *Applied Geography*, 43, 127–137. <https://doi.org/10.1016/j.apgeog.2013.06.010>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Chetehouna, K., Tabach, E. El, Bouazaoui, L., & Gascoin, N. (2015). Predicting the flame characteristics and rate of spread in fires propagating in a bed of Pinus pinaster using Artificial Neural Networks. *Process Safety and Environmental Protection*, 98, 50–56. <https://doi.org/10.1016/j.psep.2015.06.010>
- Chuvieco, E. (2020). *Fundamentals of Satellite Remote Sensing*. CRC Press. <https://doi.org/10.1201/9780429506482>
- Cilli, R., Elia, M., D’Este, M., Giannico, V., Amoroso, N., Lombardi, A., Pantaleo, E., Monaco, A., Sanesi, G., Tangaro, S., Bellotti, R., & Laforteza, R. (2022). Explainable artificial intelligence (XAI) detects wildfire occurrence in the Mediterranean countries of Southern Europe. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-20347-9>
- Collins, L., Griffioen, P., Newell, G., & Mellor, A. (2018). The utility of Random Forests for wildfire severity mapping. *Remote Sensing of Environment*, 216, 374–384. <https://doi.org/10.1016/j.rse.2018.07.005>
- Cornwell, W. K., Elvira, A., Kempen, L., Logtestijn, R. S. P., Aptroot, A., & Cornelissen, J. H. C. (2015). Flammability across the gymnosperm phylogeny: the importance of litter particle size. *New Phytologist*, 206(2), 672–681. <https://doi.org/10.1111/nph.13317>
- Csontos, P., & Cseresnyés, I. (2015). Fire-risk evaluation of austrian pine stands in Hungary - Effects of drought conditions and slope aspect on fire spread and fire behaviour. *Carpathian Journal of Earth and Environmental Sciences*, 10(3), 247–254.
- Curt, T., Borgniet, L., & Bouillon, C. (2013). Wildfire frequency varies with the size and shape of fuel types in southeastern France: Implications for environmental management. *Journal of Environmental Management*, 117, 150–161. <https://doi.org/10.1016/j.jenvman.2012.12.006>

- Curt, T., & Frejaville, T. (2018). Wildfire Policy in Mediterranean France: How Far is it Efficient and Sustainable? *Risk Analysis*, 38(3), 472–488. <https://doi.org/10.1111/risa.12855>
- Curt, T., Fréjaville, T., & Lahaye, S. (2016). Modelling the spatial patterns of ignition causes and fire regime features in southern France: Implications for fire prevention policy. *International Journal of Wildland Fire*, 25(7), 785–796. <https://doi.org/10.1071/WF15205>
- De Rigo, D., Liberta', G., Houston Durrant, T., Artés Vivancos, T., & San-Miguel-Ayanz, J. (2017). *Forest fire danger extremes in Europe under climate change: variability and uncertainty*. Publications Office of the European Union. <https://doi.org/10.2760/13180>
- Dickson, B. G., Prather, J. W., Xu, Y., Hampton, H. M., Aumack, E. N., & Sisk, T. D. (2006). Mapping the probability of large fire occurrence in northern Arizona, USA. *Landscape Ecology*, 21(5), 747–761. <https://doi.org/10.1007/s10980-005-5475-x>
- Dimitrakopoulos, A. P. (2001). A statistical classification of Mediterranean species based on their flammability components. *International Journal of Wildland Fire*, 10(2), 113. <https://doi.org/10.1071/WF01004>
- Dimitrakopoulos, A. P., Bemmerzouk, A. M., & Mitsopoulos, I. D. (2011). Evaluation of the Canadian fire weather index system in an eastern Mediterranean environment. *Meteorological Applications*, 18(1), 83–93. <https://doi.org/10.1002/met.214>
- Dimitrakopoulos, A. P., & Papaioannou, K. K. (2001). Flammability Assessment of Mediterranean Forest Fuels. *Fire Technology*, 37(2), 143–152. <https://doi.org/10.1023/A:1011641601076>
- Direction de la Sécurité Civile. (1994). Protection de la forêt contre l'incendie: Guide de stratégie générale. *Ministère de l'intérieur et de l'aménagement Du Territoire, (Colonel Battesti A.)*.
- Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*.
- Douglas, E. M., Vogel, R. M., & Kroll, C. N. (2000). Trends in floods and low flows in the United States: Impact of spatial correlation. *Journal of Hydrology*, 240(1–2), 90–105. [https://doi.org/10.1016/S0022-1694\(00\)00336-X](https://doi.org/10.1016/S0022-1694(00)00336-X)
- Dupuy, J., Fargeon, H., Martin-StPaul, N., Pimont, F., Ruffault, J., Guijarro, M., Hernando, C., Madrigal, J., & Fernandes, P. (2020). Climate change impact on future wildfire danger and activity in southern Europe: a review. *Annals of Forest Science*, 77(2), 35. <https://doi.org/10.1007/s13595-020-00933-5>
- Durbin, J., & Watson, G. S. (1950). *Testing for Serial Correlation in Least Squares Regression . I Author ( s ): J . Durbin and G . S . Watson Published by : Oxford University Press on behalf of Biometrika Trust Stable URL : http://www.jstor.org/stable/2332325 REFERENCES L. 37(1), 409–428.*
- Dutta, R., Das, A., & Aryal, J. (2016). Big data integration shows Australian bush-fire frequency is increasing significantly. *Royal Society Open Science*, 3(2), 150241. <https://doi.org/10.1098/rsos.150241>
- Elia, M., Giannico, V., Laforteza, R., & Sanesi, G. (2019). Modeling fire ignition patterns in Mediterranean urban interfaces. *Stochastic Environmental Research and Risk Assessment*, 33(1), 169–181. <https://doi.org/10.1007/s00477-018-1558-5>
- Eskandari, S., Pourghasemi, H. R., & Tiefenbacher, J. P. (2020). Relations of land cover, topography, and climate to fire occurrence in natural regions of Iran: Applying new data mining techniques for modeling and mapping fire danger. *Forest Ecology and Management*, 473, 118338. <https://doi.org/10.1016/j.foreco.2020.118338>

- Estes, B. L., Knapp, E. E., Skinner, C. N., Miller, J. D., & Preisler, H. K. (2017). Factors influencing fire severity under moderate burning conditions in the Klamath Mountains, northern California, USA. *Ecosphere*, 8(5). <https://doi.org/10.1002/ecs2.1794>
- Etlinger, M. G., & Beall, F. C. (2004). Development of a laboratory protocol for fire performance of landscape plants. *International Journal of Wildland Fire*, 13(4), 479. <https://doi.org/10.1071/WF04039>
- Evin, G., Curt, T., & Eckert, N. (2018). Has fire policy decreased the return period of the largest wildfire events in France? A Bayesian assessment based on extreme value theory. *Natural Hazards and Earth System Sciences*, 18(10), 2641–2651. <https://doi.org/10.5194/nhess-18-2641-2018>
- Fang, L., Yang, J., White, M., & Liu, Z. (2018). Predicting Potential Fire Severity Using Vegetation, Topography and Surface Moisture Availability in a Eurasian Boreal Forest Landscape. *Forests*, 9(3), 130. <https://doi.org/10.3390/f9030130>
- Fares, S., Bajocco, S., Salvati, L., Camarretta, N., Dupuy, J.-L., Xanthopoulos, G., Guijarro, M., Madrigal, J., Hernando, C., & Corona, P. (2017). Characterizing potential wildland fire fuel in live vegetation in the Mediterranean region. *Annals of Forest Science*, 74(1), 1. <https://doi.org/10.1007/s13595-016-0599-5>
- Fernandes, P. M., & Botelho, H. S. (2003). A review of prescribed burning effectiveness in fire hazard reduction. *International Journal of Wildland Fire*, 12(2), 117. <https://doi.org/10.1071/WF02042>
- Fernandes, P. M., & Cruz, M. G. (2012). Plant flammability experiments offer limited insight into vegetation–fire dynamics interactions. *New Phytologist*, 194(3), 606–609. <https://doi.org/10.1111/j.1469-8137.2012.04065.x>
- Fernandes, P. M., Monteiro-Henriques, T., Guiomar, N., Loureiro, C., & Barros, A. M. G. (2016). Bottom-Up Variables Govern Large-Fire Size in Portugal. *Ecosystems*, 19(8), 1362–1375. <https://doi.org/10.1007/s10021-016-0010-2>
- Finney, M. A., McAllister, S. S., Forthofer, J. M., & Grumstrup, T. P. (2021). *Wildland fire behaviour: dynamics, principles and processes*. CSIRO publishing.
- Fotheringham, A. S., Brunsdon, C., & Martin, C. (2003). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.
- Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30(11), 1905–1927. <https://doi.org/10.1068/a301905>
- Fox, D. M. (2011). Evaluation of the efficiency of some sediment trapping methods after a Mediterranean forest fire. *Journal of Environmental Management*, 92(2), 258–265. <https://doi.org/10.1016/j.jenvman.2009.10.006>
- Fox, D. M., Carrega, P., Ren, Y., Caillouet, P., Bouillon, C., & Robert, S. (2018). How wildfire risk is related to urban planning and Fire Weather Index in SE France (1990–2013). *Science of The Total Environment*, 621, 120–129. <https://doi.org/10.1016/J.SCITOTENV.2017.11.174>
- Fox, D. M., Martin, N., Carrega, P., Andrieu, J., Adnès, C., Emsellem, K., Ganga, O., Moebius, F., Tortorollo, N., & Fox, E. A. (2015). Increases in fire risk due to warmer summer temperatures and wildland urban interface changes do not necessarily lead to more fires. *Applied Geography*, 56. <https://doi.org/10.1016/j.apgeog.2014.10.001>
- Franquesa, M., Stehman, S. V., & Chuvieco, E. (2022). Assessment and characterization of sources of error impacting the accuracy of global burned area products. *Remote Sensing of Environment*, 280, 113214. <https://doi.org/10.1016/j.rse.2022.113214>
- Galizia, L. F., Curt, T., Barbero, R., & Rodrigues, M. (2020). Assessing the accuracy of remotely-sensed fire datasets across the Southwestern Mediterranean basin. *Natural Hazards and Earth System Sciences*, 1–19. <https://doi.org/10.5194/nhess-2020-230>



- Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1–37. <https://doi.org/10.1145/2523813>
- Ganteaume, A., & Barbero, R. (2019). Contrasting large fire activity in the French Mediterranean. *Natural Hazards and Earth System Sciences*, 19(5), 1055–1066. <https://doi.org/10.5194/nhess-19-1055-2019>
- Ganteaume, A., Camia, A., Jappiot, M., San-Miguel-Ayanz, J., Long-Fournel, M., & Lampin, C. (2013). A review of the main driving factors of forest fire ignition over Europe. In *Environmental Management* (Vol. 51, Issue 3, pp. 651–662). <https://doi.org/10.1007/s00267-012-9961-z>
- Ganteaume, A., & Guerra, F. (2018). Explaining the spatio-seasonal variation of fires by their causes: The case of southeastern France. *Applied Geography*, 90(March 2017), 69–81. <https://doi.org/10.1016/j.apgeog.2017.11.012>
- Ganteaume, A., & Jappiot, M. (2013). What causes large fires in Southern France. *Forest Ecology and Management*, 294, 76–85. <https://doi.org/10.1016/j.foreco.2012.06.055>
- Ganteaume, A., Jappiot, M., & Lampin, C. (2013). Assessing the flammability of surface fuels beneath ornamental vegetation in wildland–urban interfaces in Provence (south-eastern France). *International Journal of Wildland Fire*, 22(3), 333. <https://doi.org/10.1071/WF12006>
- Ganteaume, A., Lampin-Maillet, C., Guijarro, M., Hernando, C., Jappiot, M., Fonturbel, T., Pérez-Gorostiaga, P., & Vega, J. A. (2009). Spot fires: fuel bed flammability and capability of firebrands to ignite fuel beds. *International Journal of Wildland Fire*, 18(8), 951. <https://doi.org/10.1071/WF07111>
- Ganteaume, A., & Long-Fournel, M. (2015). Driving factors of fire density can spatially vary at the local scale in south-eastern France. *International Journal of Wildland Fire*, 24(5). <https://doi.org/10.1071/WF13209>
- Ganteaume, A., Marielle, J., Corinne, L.-M., Thomas, C., & Laurent, B. (2011). Effects of vegetation type and fire regime on flammability of undisturbed litter in Southeastern France. *Forest Ecology and Management*, 261(12), 2223–2231. <https://doi.org/10.1016/j.foreco.2010.09.046>
- Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuyse, S., Mboga, N., Wolff, E., & Kalogirou, S. (2021). Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 36(2), 121–136. <https://doi.org/10.1080/10106049.2019.1595177>
- Ghorbanzadeh, O., Kamran, K. V., Blaschke, T., Aryal, J., Naboureh, A., Einali, J., & Bian, J. (2019). Spatial prediction of wildfire susceptibility using field survey gps data and machine learning approaches. *Fire*, 2(3), 1–23. <https://doi.org/10.3390/fire2030043>
- Gill, A. M., & Moore, P. H. R. (1996). *Ignitibility of leaves of Australian plants* (Vol. 34). CSIRO Canberra.
- Gonzalez-Olabarria, J. R., Brotons, L., Gritten, D., Tudela, A., & Teres, J. A. (2012). Identifying location and causality of fire ignition hotspots in a Mediterranean region. *International Journal of Wildland Fire*, 21(7), 905. <https://doi.org/10.1071/WF11039>
- Goodchild, M. F. (2011). Scale in GIS: An overview. *Geomorphology*, 130(1–2), 5–9. <https://doi.org/10.1016/j.geomorph.2010.10.004>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>

- Hagenauer, J., & Helbich, M. (2022). A geographically weighted artificial neural network. *International Journal of Geographical Information Science*, 36(2), 215–235. <https://doi.org/10.1080/13658816.2021.1871618>
- Halofsky, J. S., Donato, D. C., Franklin, J. F., Halofsky, J. E., Peterson, D. L., & Harvey, B. J. (2018). The nature of the beast: examining climate adaptation options in forests with stand-replacing fire regimes. *Ecosphere*, 9(3), e02140. <https://doi.org/10.1002/ecs2.2140>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518. <https://doi.org/10.7717/peerj.5518>
- Hodges, J. L., & Lattimer, B. Y. (2019). Wildland Fire Spread Modeling Using Convolutional Neural Networks. *Fire Technology*, 55(6), 2115–2142. <https://doi.org/10.1007/s10694-019-00846-4>
- Holden, Z. A., Morgan, P., & Evans, J. S. (2009). A predictive model of burn severity based on 20-year satellite-inferred burn severity data in a large southwestern US wilderness area. *Forest Ecology and Management*, 258(11), 2399–2406. <https://doi.org/10.1016/j.foreco.2009.08.017>
- Holsinger, L., Parks, S. A., & Miller, C. (2016). Weather, fuels, and topography impede wildland fire spread in western US landscapes. *Forest Ecology and Management*, 380, 59–69. <https://doi.org/10.1016/j.foreco.2016.08.035>
- Jacobs, J. (1974). Quantitative Measurement of Food Selection. *Oecologia*, 14, 413–417. <https://doi.org/10.1385/1-59259-055-1:51>
- Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). *A review of machine learning applications in wildfire science and management*. <https://doi.org/10.1139/er-2020-0019>
- Jiménez-Ruano, A., Rodrigues Mimbreno, M., & De La Riva Fernández, J. (2017). Exploring spatial–temporal dynamics of fire regime features in mainland Spain. *Natural Hazards and Earth System Sciences*, 17(10), 1697–1711. <https://doi.org/10.5194/nhess-17-1697-2017>
- Keane, R. E. (2015). *Wildland Fuel Fundamentals and Applications*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-09015-3>
- Keeley, J. E., Bond, W. J., Bradstock, R. A., Pausas, J. G., & Rundel, P. W. (2011). *Fire in Mediterranean Ecosystems*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139033091>
- Keeley, J. E., & Syphard, A. D. (2016). Climate change and future fire regimes: Examples from California. In *Geosciences (Switzerland)* (Vol. 6, Issue 3). MDPI AG. <https://doi.org/10.3390/geosciences6030037>
- Khabarov, N., Krasovskii, A., Obersteiner, M., Swart, R., Dosio, A., San-Miguel-Ayanz, J., Durrant, T., Camia, A., & Migliavacca, M. (2016). Forest fires and adaptation options in Europe. *Regional Environmental Change*, 16(1), 21–30. <https://doi.org/10.1007/s10113-014-0621-0>
- Khanmohammadi, S., Arashpour, M., Golafshani, E. M., Cruz, M. G., & Rajabifard, A. (2023). An artificial intelligence framework for predicting fire spread sustainability in semiarid shrublands. *International Journal of Wildland Fire*. <https://doi.org/10.1071/WF22216>
- Kolanek, A., & Szymanowski, M. (2021). *Human Activity Affects Forest Fires : The Impact of Anthropogenic Factors on the Density of Forest Fires in Poland*. 1–21.

- Koutsias, N., Arianoutsou, M., Kallimanis, A. S., Mallinis, G., Halley, J. M., & Dimopoulos, P. (2012). Where did the fires burn in Peloponnisos, Greece the summer of 2007? Evidence for a synergy of fuel and weather. *Agricultural and Forest Meteorology*, *156*, 41–53. <https://doi.org/10.1016/j.agrformet.2011.12.006>
- Koutsias, N., Martínez-Fernández, J., & Allgöwer, B. (2010). Do factors causing wildfires vary in space? evidence from geographically weighted regression. *GIScience and Remote Sensing*, *47*(2), 221–240. <https://doi.org/10.2747/1548-1603.47.2.221>
- Krawczyk, B., Minku, L. L., Gama, J., Stefanowski, J., & Woźniak, M. (2017). Ensemble learning for data stream analysis: A survey. *Information Fusion*, *37*, 132–156. <https://doi.org/10.1016/j.inffus.2017.02.004>
- Lafortezza, R., Sanesi, G., & Chen, J. (2013). Large-scale effects of forest management in Mediterranean landscapes of Europe. *IForest - Biogeosciences and Forestry*, *6*(6). <https://doi.org/10.3832/ifor0960-006>
- Lagerquist, R., Flannigan, M. D., Wang, X., & Marshall, G. A. (2017). Automated prediction of extreme fire weather from synoptic patterns in northern Alberta, Canada. *Canadian Journal of Forest Research*, *47*(9), 1175–1183. <https://doi.org/10.1139/cjfr-2017-0063>
- Lampin-Maillet, C., Long-Fournel, M., Ganteaume, A., Jappiot, M., & Ferrier, J. P. (2011). Land cover analysis in wildland-urban interfaces according to wildfire risk: A case study in the South of France. *Forest Ecology and Management*, *261*(12), 2200–2213. <https://doi.org/10.1016/j.foreco.2010.11.022>
- Langford, Z., Kumar, J., & Hoffman, F. (2018). Wildfire Mapping in Interior Alaska Using Deep Neural Networks on Imbalanced Datasets. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 770–778. <https://doi.org/10.1109/ICDMW.2018.00116>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lekkas, E., Carydis, P., Lagouvardos, K., Mavroulis, S., & Diakakis, M. (2018). The July 2018 Attica (Central Greece) wildfires. *Newslett. Environ. Disaster Crises Manage. Strateg.*
- Lemaitre, G., Nogueira, F., & Aridas, C. K. (2016). *Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning*.
- Lippitt, C. D., & Zhang, S. (2018). The impact of small unmanned airborne platforms on passive optical remote sensing: a conceptual perspective. *International Journal of Remote Sensing*, *39*(15–16), 4852–4868. <https://doi.org/10.1080/01431161.2018.1490504>
- Lipton, Z. C. (2018). *The Mythos of Model Interpretability*.
- López-Serrano, P. M., López-Sánchez, C. A., Álvarez-González, J. G., & García-Gutiérrez, J. (2016). A Comparison of Machine Learning Techniques Applied to Landsat-5 TM Spectral Data for Biomass Estimation. *Canadian Journal of Remote Sensing*, *42*(6), 690–705. <https://doi.org/10.1080/07038992.2016.1217485>
- Loyola-Gonzalez, O. (2019). Black-box vs. White-Box: Understanding their advantages and weaknesses from a practical point of view. In *IEEE Access* (Vol. 7, pp. 154096–154113). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2019.2949286>
- Lundberg, S., & Lee, S.-I. (2017, May 22). A Unified Approach to Interpreting Model Predictions. *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*. <http://arxiv.org/abs/1705.07874>
- Mancini, L., Elia, M., Barbati, A., Salvati, L., Corona, P., Lafortezza, R., & Sanesi, G. (2018). Are Wildfires Knocking on the Built-Up Areas Door? *Forests*, *9*(5), 234. <https://doi.org/10.3390/f9050234>

- Manly, B., McDonald, L., Thomas, D., McDonald, T., & Erickson, W. (2002). Resource Selection by Animals Statistical Design and Analysis for Field Studies Second Edition. *Kluwer Academic Publishers*, 65(3), 25–28.
- Mann, H. B. (1945). Non-Parametric Test Against Trend. *Econometrica*, 13(3), 245–259.
- Marcinkevičs, R., & Vogt, J. E. (2023). Interpretable and explainable machine learning: A methods-centric overview with concrete examples. *WIREs Data Mining and Knowledge Discovery*. <https://doi.org/10.1002/widm.1493>
- Marques, S., Borges, J. G., Garcia-Gonzalo, J., Moreira, F., Carreiras, J. M. B., Oliveira, M. M., Cantarina, A., Botequim, B., & Pereira, J. M. C. (2011). Characterization of wildfires in Portugal. *European Journal of Forest Research*, 130(5), 775–784. <https://doi.org/10.1007/s10342-010-0470-4>
- Martin, J., & Hillen, T. (2016). The Spotting Distribution of Wildfires. *Applied Sciences*, 6(6), 177. <https://doi.org/10.3390/app6060177>
- Martínez-Fernández, J., Chuvieco, E., & Koutsias, N. (2013). Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. *Natural Hazards and Earth System Science*, 13(2), 311–327. <https://doi.org/10.5194/nhess-13-311-2013>
- Mayr, M. J., Vanselow, K. A., & Samimi, C. (2018). Fire regimes at the arid fringe: A 16-year remote sensing perspective (2000–2016) on the controls of fire activity in Namibia from spatial predictive models. *Ecological Indicators*, 91, 324–337. <https://doi.org/10.1016/j.ecolind.2018.04.022>
- McFarland, M., Emanuele, R., Morris, D., & Augspurger, T. (2022). *Microsoft Open Source microsoft/PlanetaryComputer: October 2022*. <https://doi.org/10.5281/ZENODO.7261897>
- McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>
- Mermoz, M., Kitzberger, T., & Veblen, T. T. (2005). Landscape influences on occurrence and spread of wildfires in Patagonian forests and shrublands. *Ecology*, 86(10). <https://doi.org/10.1890/04-1850>
- Mhaweji, M., Faour, G., & Adjizian-Gerard, J. (2015). Wildfire Likelihood's Elements: A Literature Review. *Challenges*, 6(2), 282–293. <https://doi.org/10.3390/challe6020282>
- Michelaki, C., Fyllas, N. M., Galanidis, A., Aloupi, M., Evangelou, E., Arianoutsou, M., & Dimitrakopoulos, P. G. (2020). Adaptive flammability syndromes in thermo-Mediterranean vegetation, captured by alternative resource-use strategies. *Science of The Total Environment*, 718, 137437. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.137437>
- Miller, J. D., Safford, H. D., Crimmins, M., & Thode, A. E. (2009). Quantitative Evidence for Increasing Forest Fire Severity in the Sierra Nevada and Southern Cascade Mountains, California and Nevada, USA. *Ecosystems*, 12(1). <https://doi.org/10.1007/s10021-008-9201-9>
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-hill.
- Molina, J. R., Lora, A., Prades, C., & Rodríguez y Silva, F. (2019). Roadside vegetation planning and conservation: New approach to prevent and mitigate wildfires based on fire ignition potential. *Forest Ecology and Management*, 444, 163–173. <https://doi.org/10.1016/j.foreco.2019.04.034>
- Molina, J. R., Martín, T., Rodríguez Y Silva, F., & Herrera, M. Á. (2017). The ignition index based on flammability of vegetation improves planning in the wildland-urban interface: A case study in Southern Spain. *Landscape and Urban Planning*, 158, 129–138. <https://doi.org/10.1016/j.landurbplan.2016.11.003>

- Molina-Terrén, D. M., Xanthopoulos, G., Diakakis, M., Ribeiro, L., Caballero, D., Delogu, G. M., Viegas, D. X., Silva, C. A., & Cardil, A. (2019). Analysis of forest fire fatalities in Southern Europe: Spain, Portugal, Greece and Sardinia (Italy). *International Journal of Wildland Fire*, 28(2), 85. <https://doi.org/10.1071/WF18004>
- Moreira, F., Rego, F. C., & Ferreira, P. G. (2001). Temporal (1958-1995) pattern of change in a cultural landscape of northwestern Portugal: Implications for fire occurrence. *Landscape Ecology*, 16(6), 557–567. <https://doi.org/10.1023/A:1013130528470>
- Moreira, F., Vaz, P., Catry, F., & Silva, J. S. (2009). Regional variations in wildfire susceptibility of land-cover types in Portugal: implications for landscape management to minimize fire hazard. *International Journal of Wildland Fire*, 18(5). <https://doi.org/10.1071/WF07098>
- Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., Barbati, A., Corona, P., Vaz, P., Xanthopoulos, G., Mouillot, F., & Bilgili, E. (2011). Landscape - wildfire interactions in southern Europe: Implications for landscape management. In *Journal of Environmental Management* (Vol. 92, Issue 10, pp. 2389–2402). <https://doi.org/10.1016/j.jenvman.2011.06.028>
- Moreno, J. M., Viedma, O., Zavala, G., & Luna, B. (2011). Landscape variables influencing forest fires in central Spain. *International Journal of Wildland Fire*, 20(5), 678–689. <https://doi.org/10.1071/WF10005>
- Moriondo, M., Good, P., Durao, R., Bindi, M., Giannakopoulos, C., & Corte-Real, J. (2006). Potential impact of climate change on fire risk in the Mediterranean area. *Climate Research*, 31, 85–95. <https://doi.org/10.3354/cr031085>
- Moritz, M. A., Morais, M. E., Summerell, L. A., Carlson, J. M., & Doyle, J. (2005). *Wildfires, complexity, and highly optimized tolerance* (Vol. 102, Issue 50). [www.pnas.org/cgi/doi/10.1073/pnas.0508985102](http://www.pnas.org/cgi/doi/10.1073/pnas.0508985102)
- Mouillot, F., Ratte, J. P., Joffre, R., Moreno, J. M., & Rambal, S. (2003). Some determinants of the spatio-temporal fire cycle in a mediterranean landscape (Corsica, France). *Landscape Ecology*, 18(7), 665–674. <https://doi.org/10.1023/B:LAND.0000004182.22525.a9>
- Mouillot, F., Schultz, M. G., Yue, C., Cadule, P., Tansey, K., Ciais, P., & Chuvieco, E. (2014). Ten years of global burned area products from spaceborne remote sensing—A review: Analysis of user needs and recommendations for future developments. *International Journal of Applied Earth Observation and Geoinformation*, 26, 64–79. <https://doi.org/10.1016/j.jag.2013.05.014>
- Mueller, S. E., Thode, A. E., Margolis, E. Q., Yocom, L. L., Young, J. D., & Iniguez, J. M. (2020). Climate relationships with increasing wildfire in the southwestern US from 1984 to 2015. *Forest Ecology and Management*, 460, 117861. <https://doi.org/10.1016/j.foreco.2019.117861>
- Neeti, N., & Eastman, J. R. (2011). A Contextual Mann-Kendall Approach for the Assessment of Trend Significance in Image Time Series. *Transactions in GIS*, 15(5), 599–611. <https://doi.org/10.1111/j.1467-9671.2011.01280.x>
- Nikparvar, B., & Thill, J. C. (2021). Machine learning of spatial data. In *ISPRS International Journal of Geo-Information* (Vol. 10, Issue 9). MDPI. <https://doi.org/10.3390/ijgi10090600>
- Nunes, A. N., Lourenço, L., & Meira, A. C. C. (2016). Exploring spatial patterns and drivers of forest fires in Portugal (1980–2014). *Science of the Total Environment*, 573, 1190–1202. <https://doi.org/10.1016/j.scitotenv.2016.03.121>
- Nunes, M. C. S., Vasconcelos, M. J., Pereira, J. M. C., Dasgupta, N., Alldredge, R. J., & Rego, F. C. (2005). Land Cover Type and Fire in Portugal: Do Fires Burn Land Cover

- Selectively? *Landscape Ecology*, 20(6), 661–673. <https://doi.org/10.1007/s10980-005-0070-8>
- Oehler, F., Oliveira, S., Barredo, J., Camia, A., San-Miguel-Ayanz, J., Pettenella, D., & Mavsar, R. (2012). *Assessing European wild fire vulnerability Assessing European wild fire vulnerability*. April.
- Oliveira, S., Moreira, F., Boca, R., San-Miguel-Ayanz, J., & Pereira, J. M. C. (2014a). Assessment of fire selectivity in relation to land cover and topography: A comparison between Southern European countries. *International Journal of Wildland Fire*, 23(5), 620–630. <https://doi.org/10.1071/WF12053>
- Oliveira, S., Moreira, F., Boca, R., San-Miguel-Ayanz, J., & Pereira, J. M. C. (2014b). Assessment of fire selectivity in relation to land cover and topography: A comparison between Southern European countries. *International Journal of Wildland Fire*, 23(5), 620–630. <https://doi.org/10.1071/WF12053>
- Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. C. (2012a). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*, 275, 117–129. <https://doi.org/10.1016/j.foreco.2012.03.003>
- Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. C. (2012b). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*, 275, 117–129. <https://doi.org/10.1016/j.foreco.2012.03.003>
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42–57. <https://doi.org/10.1016/j.rse.2014.02.015>
- Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography*.
- Otón, G., Pereira, J. M. C., Silva, J. M. N., & Chuvieco, E. (2021). Analysis of trends in the firecci global long term burned area product (1982–2018). *Fire*, 4(4). <https://doi.org/10.3390/fire4040074>
- Padilla, M., & Vega-García, C. (2011). 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. *International Journal of Wildland Fire*, 20(1), 46. <https://doi.org/10.1071/WF09139>
- Page, W. G., Jenkins, M. J., & Runyon, J. B. (2012). Mountain pine beetle attack alters the chemistry and flammability of lodgepole pine foliage. *Canadian Journal of Forest Research*, 42(8), 1631–1647. <https://doi.org/10.1139/x2012-094>
- Palheiro, P. M., Fernandes, P., & Cruz, M. G. (2006). A fire behaviour-based fire danger classification for maritime pine stands: Comparison of two approaches. *Forest Ecology and Management*, 234, S54. <https://doi.org/10.1016/j.foreco.2006.08.075>
- Parks, S. A., Holsinger, L. M., Panunto, M. H., Jolly, W. M., Dobrowski, S. Z., & Dillon, G. K. (2018). High-severity fire: evaluating its key drivers and mapping its probability across western US forests. *Environmental Research Letters*, 13(4). <https://doi.org/10.1088/1748-9326/aab791>
- 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 {P}ython. *Journal of Machine Learning Research*, 12, 2825–2830.
- Penman, T. D., Bradstock, R. A., & Price, O. (2013). Modelling the determinants of ignition in the Sydney Basin, Australia: Implications for future management. *International Journal of Wildland Fire*, 22(4), 469–478. <https://doi.org/10.1071/WF12027>

- Pereira, A., Pereira, J., Libonati, R., Oom, D., Setzer, A., Morelli, F., Machado-Silva, F., & de Carvalho, L. (2017). Burned Area Mapping in the Brazilian Savanna Using a One-Class Support Vector Machine Trained by Active Fires. *Remote Sensing*, 9(11), 1161. <https://doi.org/10.3390/rs9111161>
- Pereira, M. G., Aranha, J., & Amraoui, M. (2014). Land cover fire proneness in Europe. *Forest Systems*, 23(3), 598–610. <https://doi.org/10.5424/fs/2014233-06115>
- Pokorná, L., Kučerová, M., & Huth, R. (2018). Annual cycle of temperature trends in Europe, 1961–2000. *Global and Planetary Change*, 170. <https://doi.org/10.1016/j.gloplacha.2018.08.015>
- Pyne, S. J., Andrews, P. L., & Laven, R. D. (1996). *Introduction to wildland fire. Fire management in the United States*. (2nd ed.). John Wiley & Sons.
- Resco de Dios, V., Fischer, C., & Colinas, C. (2007). Climate Change Effects on Mediterranean Forests and Preventive Measures. *New Forests*, 33(1), 29–40. <https://doi.org/10.1007/s11056-006-9011-x>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should i trust you?” Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Ricotta, C., Bajocco, S., Guglietta, D., & Conedera, M. (2018). Assessing the Influence of Roads on Fire Ignition: Does Land Cover Matter? *Fire*, 1(2), 24. <https://doi.org/10.3390/fire1020024>
- Riley, K. L., Grenfell, I. C., Finney, M. A., & Crookston, N. L. (2014). Utilizing random forests imputation of forest plot data for landscape-level wildfire analyses. In D. X. Viegas (Ed.), *Advances in forest fire research* (pp. 583–590). Imprensa da Universidade de Coimbra. [https://doi.org/10.14195/978-989-26-0884-6\\_67](https://doi.org/10.14195/978-989-26-0884-6_67)
- Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Science*, 5, 23–27.
- Rodrigues, M., de la Riva, J., & Fotheringham, S. (2014). Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Applied Geography*, 48, 52–63. <https://doi.org/10.1016/j.apgeog.2014.01.011>
- Rodrigues, M., Jiménez, A., & de la Riva, J. (2016). Analysis of recent spatial–temporal evolution of human driving factors of wildfires in Spain. *Natural Hazards*, 84(3), 2049–2070. <https://doi.org/10.1007/s11069-016-2533-4>
- Rodrigues, M., Jiménez-Ruano, A., & de la Riva, J. (2020). Fire regime dynamics in mainland Spain. Part 1: Drivers of change. *Science of the Total Environment*, 721. <https://doi.org/10.1016/j.scitotenv.2019.135841>
- Rothermel, R. C. (1983). How to predict the spread and intensity of forest and range fires. *US Department of Agriculture, Forest Service, General Technical Report, INT-143*. <https://doi.org/10.2737/INT-GTR-143>
- Roy, H. G., Fox, D. M., & Emsellem, K. (2015). Spatial dynamics of land cover change in a Euro-Mediterranean catchment (1950–2008). *Journal of Land Use Science*, 10(3), 277–297. <https://doi.org/10.1080/1747423X.2014.898105>
- Ruffault, J., Curt, T., Martin-StPaul, N. K., Moron, V., & Trigo, R. M. (2018). Extreme wildfire events are linked to global-change-type droughts in the northern Mediterranean. *Natural Hazards and Earth System Sciences*, 18(3), 847–856. <https://doi.org/10.5194/nhess-18-847-2018>
- Ruffault, J., Moron, V., Trigo, R. M., & Curt, T. (2017). Daily synoptic conditions associated with large fire occurrence in Mediterranean France: evidence for a wind-driven fire

- regime. *International Journal of Climatology*, 37(1), 524–533. <https://doi.org/10.1002/joc.4680>
- Ruffault, J., & Mouillot, F. (2015). How a new fire-suppression policy can abruptly reshape the fire-weather relationship. *Ecosphere*, 6(10), 1–19. <https://doi.org/10.1890/ES15-00182.1>
- Ruffault, J., & Mouillot, F. (2017). Contribution of human and biophysical factors to the spatial distribution of forest fire ignitions and large wildfires in a French Mediterranean region. *International Journal of Wildland Fire*, 26(6). <https://doi.org/10.1071/WF16181i>
- San-Miguel-Ayanz, J., Durrant, T., Boca, R., Maianti, P., Liberta', G., Artes Vivancos, T., Jacome Felix Oom, D., Branco, A., De Rigo, D., Ferrari, D., Pfeiffer, H., Grecchi, R., Nuijten, D., & Leray, T. (2020). *Forest Fires in Europe, Middle East and North Africa 2019*. <https://doi.org/10.2760/468688>
- San-Miguel-Ayanz, J., Durrant, T., Boca, R., Maianti, P., Libertà, G., Artés Vivancos, T., Oom, D., Branco, A., De Rigo, D., Ferrari, D., Pfeiffer, H., Grecchi, R., Onida, M., & Löffler, P. (2022). *Forest Fires in Europe, Middle East and North Africa 2021*. Publications Office of the European Union. <https://doi.org/doi/10.2760/34094>
- San-Miguel-Ayanz, J., Durrant, T., Boca, R., Maianti, P., Libertà, G., Oom, D., Branco, A., De Rigo, D., Ferrari, D., Roglia, E., & Scionti, N. (2023). *Advance report on forest fires in Europe, Middle East and North Africa 2022*. Publications Office of the European Union. <https://doi.org/doi/10.2760/091540>
- San-Miguel-Ayanz, J., Moreno, J. M., & Camia, A. (2013). Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives. *Forest Ecology and Management*, 294, 11–22. <https://doi.org/10.1016/j.foreco.2012.10.050>
- San-Miguel-Ayanz, J., Schulte, E., Schmuck, G., & Camia, A. (2013). The European Forest Fire Information System in the context of environmental policies of the European Union. *Forest Policy and Economics*, 29, 19–25. <https://doi.org/10.1016/j.forpol.2011.08.012>
- Shmuel, A., & Heifetz, E. (2023). Developing novel machine-learning-based fire weather indices. *Machine Learning: Science and Technology*, 4(1), 015029. <https://doi.org/10.1088/2632-2153/acc008>
- Silva, J. M. N., Moreno, M. V., Le Page, Y., Oom, D., Bistinas, I., & Pereira, J. M. C. (2019). Spatiotemporal trends of area burnt in the Iberian Peninsula, 1975–2013. *Regional Environmental Change*, 19(2), 515–527. <https://doi.org/10.1007/s10113-018-1415-6>
- Simpson, K. J., Ripley, B. S., Christin, P., Belcher, C. M., Lehmann, C. E. R., Thomas, G. H., & Osborne, C. P. (2016). Determinants of flammability in savanna grass species. *Journal of Ecology*, 104(1), 138–148. <https://doi.org/10.1111/1365-2745.12503>
- Syphard, A. D., & Keeley, J. E. (2015). Location, timing and extent of wildfire vary by cause of ignition. *International Journal of Wildland Fire*, 24(1), 37–47. <https://doi.org/10.1071/WF14024>
- Syphard, A. D., Radeloff, V. C., Keeley, J. E., Hawbaker, T. J., Clayton, M. K., Stewart, S. I., & Hammer, R. B. (2007). HUMAN INFLUENCE ON CALIFORNIA FIRE REGIMES. *Ecological Applications*, 17(5), 1388–1402. <https://doi.org/10.1890/06-1128.1>
- Syphard, A. D., Radeloff, V. C., Keuler, N. S., Taylor, R. S., Hawbaker, T. J., Stewart, S. I., & Clayton, M. K. (2008). Predicting spatial patterns of fire on a southern California landscape. *International Journal of Wildland Fire*, 17(5), 602–613. <https://doi.org/10.1071/WF07087>
- Tedim, F., Leone, V., Lovreglio, R., Xanthopoulos, G., Chas-Amil, M.-L., Ganteaume, A., Efe, R., Royé, D., Fuerst-Bjeliš, B., Nikolov, N., Musa, S., Milenković, M., Correia, F., Conedera, M., & Boris Pezzatti, G. (2022). Forest Fire Causes and Motivations in the Southern and South-Eastern Europe through Experts' Perception and Applications to Current Policies. *Forests*, 13(4), 562. <https://doi.org/10.3390/f13040562>



- Tessler, N., Wittenberg, L., & Greenbaum, N. (2016). Vegetation cover and species richness after recurrent forest fires in the Eastern Mediterranean ecosystem of Mount Carmel, Israel. *Science of the Total Environment*, 572, 1395–1402. <https://doi.org/10.1016/j.scitotenv.2016.02.113>
- Tobler, A. W. R. (1889). Clark University. *Science*, ns-13(332), 462–465. <https://doi.org/10.1126/science.ns-13.332.462>
- Turco, M., Bedia, J., Di Liberto, F., Fiorucci, P., Von Hardenberg, J., Koutsias, N., Llasat, M. C., Xystrakis, F., & Provenzale, A. (2016). Decreasing fires in mediterranean Europe. *PLoS ONE*, 11(3). <https://doi.org/10.1371/journal.pone.0150663>
- Turco, M., Llasat, M.-C., von Hardenberg, J., & Provenzale, A. (2014). Climate change impacts on wildfires in a Mediterranean environment. *Climatic Change*, 125(3–4), 369–380. <https://doi.org/10.1007/s10584-014-1183-3>
- Urbieto, I. R., Franquesa, M., Viedma, O., & Moreno, J. M. (2019). Fire activity and burned forest lands decreased during the last three decades in Spain. *Annals of Forest Science*, 76(3). <https://doi.org/10.1007/s13595-019-0874-3>
- Van Wagner, C. E. (1987). *Development and structure of the Canadian forest fire weather index system* (Vol. 35). Canadian Forestry Service.
- Varela, V., Vlachogiannis, D., Sfetsos, A., Karozis, S., Politi, N., & Giroud, F. (2019). Projection of Forest Fire Danger due to Climate Change in the French Mediterranean Region. *Sustainability*, 11(16), 4284. <https://doi.org/10.3390/su11164284>
- Vasilakos, C., Kalabokidis, K., Hatzopoulos, J., & Matsinos, I. (2009). Identifying wildland fire ignition factors through sensitivity analysis of a neural network. *Natural Hazards*, 50(1), 125–143. <https://doi.org/10.1007/s11069-008-9326-3>
- Vecín-Arias, D., Castedo-Dorado, F., Ordóñez, C., & Rodríguez-Pérez, J. R. (2016). Biophysical and lightning characteristics drive lightning-induced fire occurrence in the central plateau of the Iberian Peninsula. *Agricultural and Forest Meteorology*, 225, 36–47. <https://doi.org/10.1016/j.agrformet.2016.05.003>
- Verdú, F., Salas, J., & Vega-García, C. (2012). A multivariate analysis of biophysical factors and forest fires in Spain, 1991 - 2005. *International Journal of Wildland Fire*, 21(5), 498. <https://doi.org/10.1071/WF11100>
- Viedma, O., Urbieto, I. R., & Moreno, J. M. (2018). Wildfires and the role of their drivers are changing over time in a large rural area of west-central Spain. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-36134-4>
- Viegas, D. X., Bovio, G., Ferreira, A., Nosenzo, A., & Sol, B. (1999). Comparative study of various methods of fire danger evaluation in southern Europe. *International Journal of Wildland Fire*, 9(4), 235. <https://doi.org/10.1071/WF00015>
- Vilar, L., Gómez, I., Martínez-Vega, J., Echavarría, P., Riaño, D., & Martín, M. P. (2016). Multitemporal modelling of socio-economic wildfire drivers in central Spain between the 1980s and the 2000s: Comparing generalized linear models to machine learning algorithms. *PLoS ONE*, 11(8). <https://doi.org/10.1371/journal.pone.0161344>
- Wang, X. L., & Swail, V. R. (2001). Changes of extreme Wave Heights in northern Hemisphere Oceans and related atmospheric circulation regimes. *Journal of Climate*, 14(10), 2204–2221. [https://doi.org/10.1175/1520-0442\(2001\)014<2204:COEWHI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<2204:COEWHI>2.0.CO;2)
- Waskom, M. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>
- White, R. H., & Zipperer, W. C. (2010). Testing and classification of individual plants for fire behaviour: plant selection for the wildland - urban interface. *International Journal of Wildland Fire*, 19(2), 213. <https://doi.org/10.1071/WF07128>

- Wu, Z., He, H. S., Fang, L., Liang, Y., & Parsons, R. A. (2018). Wind speed and relative humidity influence spatial patterns of burn severity in boreal forests of northeastern China. *Annals of Forest Science*, 75(3), 66. <https://doi.org/10.1007/s13595-018-0749-z>
- Yang, L., Driscoll, J., Sarigai, S., Wu, Q., Chen, H., & Lippitt, C. D. (2022). Google Earth Engine and Artificial Intelligence (AI): A Comprehensive Review. In *Remote Sensing* (Vol. 14, Issue 14). MDPI. <https://doi.org/10.3390/rs14143253>
- Zhao, Y., Ma, J., Li, X., & Zhang, J. (2018). Saliency Detection and Deep Learning-Based Wildfire Identification in UAV Imagery. *Sensors*, 18(3), 712. <https://doi.org/10.3390/s18030712>



# Appendix

Publications as part of this thesis:

## Articles

1) Bountzouklis, C., Fox, D. M., & Di Bernardino, E. (2022). Environmental factors affecting wildfire-burned areas in southeastern France, 1970–2019. *Natural Hazards and Earth System Sciences*, 22(4), 1181–1200. <https://doi.org/10.5194/nhess-22-1181-2022>.

2) Bountzouklis, C., Fox, D. M., & Di Bernardino, E. (2022). Fire cause classification of undetermined fires in southeastern France. In D. X. Viegas & L. M. Ribeiro (Eds.), *Advances in Forest Fire Research 2022* (pp. 1106– 1112). Imprensa da Universidade de Coimbra. [https://doi.org/10.14195/978-989-26-2298-9\\_167](https://doi.org/10.14195/978-989-26-2298-9_167).

3) Bountzouklis, C., Fox, D. M., & Di Bernardino, E. (2023). Predicting wildfire ignition causes in Southern France using eXplainable Artificial Intelligence (XAI) methods. *Environmental Research Letters*, 18(4), 044038. <https://doi.org/10.1088/1748-9326/acc8ee>.

## Scientific posters

3) Bountzouklis, C., Fox, D. M., & Di Bernardino, E. (2021). Topographic and Vegetation Controls on Wildfires in South-East France, 1970-2019. 3IA Côte d’Azur Annual Scientific Meeting, Nov 2021, Biot, France. (hal-03592370).

4) Bountzouklis, C., Fox, D. M., & Di Bernardino, E. (2022). Fire source classification of unknown caused fires based on Random Forests. 3IA Côte d’Azur Annual Scientific Meeting, Nov 2022, Biot, France. (hal-03592370).