A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

A Machine Learning-Based Analysis of Wildfire Dynamics in Portugal

Integrating Terrain, Policy, Demographics, and Climate Change

for Comprehensive Risk Assessment

FREDRIK JUNG

Work project carried out under the supervision of:

Sreyaa Guha

19/12/2023

Abstract:

This study investigates machine learning in wildfire risk management in Portugal. Incorporating environmental, demographic, socio-economic, and policy data, it examines wildfire dynamics and the shortcomings of conventional approaches. The research evaluates Portugal's National Wildland Fire Management Plan and the effect of climate change on wildfire risk. Findings indicate machine learning models enhance risk assessment accuracy and that robust climate action, aligned with the Paris Agreement, could mitigate wildfire severity. The study advocates for an integrated management strategy, blending technology, policy analysis, and environmental knowledge, and offers recommendations for better wildfire mitigation.

Keywords:

Machine Learning, Wildfire Risk Management, Climate Change, Integrated Policy Analysis, Portugal Wildfires, Predictive Modelling, Environmental Factors, Socio-Economic Influences

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Table of Figures

List of Tables

Table of Contents

A Machine Learning-Based Analysis of Wildfire Dynamics in Portugal

1 Introduction

Forests are a key resource for millions of individuals and show a substantial impact on employment, economic development, and biodiversity in many nations (Dieterle 2009). Nonetheless, forests are highly sensitive to changes in climate, factors such as rising temperatures and decreased rainfall can cause drought conditions, rendering forests more susceptible to fires. Portugal has recorded the highest number of wildfires among all European nations over the past decade. Wildfires consumed on average over 80,000 hectares of land annually between 2009 and 2021, equivalent to about eight times its capital Lisbon. The catastrophic wildfires that ravaged Central Portugal from June 17 to 24, 2017, claimed 66 lives, emphasizing the necessity of evaluating and forecasting forest fire hazards. This study investigates the nuanced interplay between environmental, socioeconomic, and policy-driven influences on wildfire dynamics within this Mediterranean nation.

A core element of understanding forest fires is identifying the areas that are susceptible of producing large and uncontrolled fires. We examine the composition of natural and artificial factors that cause structural danger using machine learning algorithms. This endeavor involves dissecting extensive historical data sets and environmental variables, combined with human activity patterns, to predict and understand the geographies most susceptible to wildfires in high resolution. By identifying these high-risk zones, we aim to reveal the underlying factors contributing to this susceptibility, enabling targeted interventions and informed policymaking.

Complementing this scientific approach is a critical analysis of how the evolving climate, as framed by the Paris Agreement's climate change mitigation targets, affects the frequency and intensity of wildfires. This research explores the specific climate changes and their correlation with wildfire patterns, with a focus on the Portuguese experience. It addresses the challenging task of navigating the goals of the Paris Agreement to achieve a better understanding of the impacts of climate change on forest fires in Portugal.

The socio-economic dynamics of Portugal also play a significant role in shaping wildfire susceptibility. A particular focus is placed on the demographic shifts in rural areas, where the migration of younger generations to urban centers has profound implications for the ecological balance and fire management capabilities of these regions. This study examines the relationship between rural depopulation and the escalation of wildfires, aiming to determine if and how this demographic transition contributes to increased wildfire risk through factors such as reduced local economic resilience and diminished firefighting resources.

Moreover, the research assesses the effectiveness of Portugal's National Plan for Integrated Wildland Fire Management, instituted following the catastrophic wildfires of 2017. This analysis serves as a pivotal research study to evaluate the success and limitations of national strategies in the aftermath of their implementation. A comparative analysis of wildfire damages before and after the plan's enactment aims to reveal the policy's regional impacts and efficacy, providing a foundation for future recommendations.

The structure of this thesis is designed to guide the reader through an in-depth investigation of these interconnected topics. The literature review examines the academic discourse on forest fire management, focusing on the interplay of environmental, political, and social factors, proceeding by describing the data sources employed by the thesis to explore the gaps highlighted in the literature review further. Each research question is thoroughly investigated in the individual parts, with extensive methodology and discussion sections which critically engages the formulated the research questions and synthesizes key findings. The final chapter concludes the overarching findings and puts forth recommendations for further improvements in wildfire managements.

In sum, this research sets the stage for a comprehensive exploration into the factors exacerbating wildfire risks in Portugal and the complex challenges involved in managing such natural disasters in a changing world. Through a blend of predictive analytics, climate policy analysis, socio-economic examination, and policy evaluation, this study strives to provide a holistic understanding of wildfire dynamics, the goal is to provide well-informed approaches to the reduction and control of these fires.

2 Literature Review

In this section we provide a comprehensive examination of the existing academic discourse on wildfire management, particularly within the context of Portugal. It delves into the complex interplay between environmental science, public policy, demographic shifts, and societal safety, highlighting the critical role of effective management strategies as climate change exacerbates wildfire intensity and frequency (Gómez-González, Ojeda, and Fernandes 2018; Turco et al. 2019; AGIF 2020; Pacheco and Claro 2021). This literature review focuses on outlining the current academic discourse on wildfire management, spotlighting key historical developments, theoretical models, empirical findings, the impact of climate change and the technological advancements in the field.

Human communities have had a long and conflict-ridden relationship with wildfires. Globally, wildfires have been viewed through different lenses: from natural disturbances essential for ecological balance to devastating events with dire societal implications (Brotons et al. 2013). The latter perspective has gained prominence with rising urbanization, making wildfire management a top priority (Catry et al. 2012; T. M. Oliveira et al. 2016; AGIF 2020; Ribeiro et al. 2020). In Portugal, the relationship with wildfires has been tumultuous. Historically, fires were used for agricultural purposes, shaping landscapes and ecosystems. However, the 20th century saw a transition. Portugal's vast stretches of eucalyptus and pine monocultures, often cited as fire-prone landscapes, emerged as commercial forestry gained traction (Carmo et al. 2011; Catry et al. 2012; Pereira, Aranha, and Amraoui 2014). The decades leading to the 21st century witnessed several catastrophic fire events, making it clear that Portugal's wildfire challenge was multifaceted, encompassing not just environmental but socio-economic dimensions (Sebastián-López et al. 2008; AGIF 2020). The turn of the century spurred national introspection, leading to a series of policies and strategies. Portugal’s National Forest Strategy of 2006 laid the groundwork, stressing the balance between sustainable forestry and fire risk reduction. However, the magnitude of wildfires in 2017 underscored the need for a more robust approach, culminating in the National Plan for Integrated Wildland Fire Management (Collins et al. 2013; Fernandes et al. 2014; AGIF 2020).

The study of wildfires, in essence, merges several disciplines. Ecologically, the disturbance theory propounds that wildfire, in their natural setting, are pivotal for ecosystem regeneration and nutrient cycling. From a policy perspective, the PAVE model (Prevention, Adaptation, Vulnerability reduction, and Emergency response) has been advocated by scholars like Fernandes et al. (2014), Moreira et al. (2020) or Pandey et al (2023) as a comprehensive approach to wildfire management. In dissecting the human-fire interface, socio-ecological systems theory provides a useful lens. It argues that human and environmental systems are intertwined, and the impact of wildfires often hinges on societal decisions about land use, resource allocation, and risk perception (Martínez, Vega-Garcia, and Chuvieco 2009; Ager, Vaillant, and Finney 2010; Carmo et al. 2011; Brotons et al. 2013; AGIF 2020). Notably, 96% of wildfires in the EU are human-caused (EFFIS 2023). Hereby, risk-hazard frameworks have been instrumental. These models assess not just the physical characteristics of fires but integrate variables like population density, infrastructure, and socio-economic factors to determine the "at-risk" areas, informing more targeted intervention strategies (Sebastián-López et al. 2008; Parente and Pereira 2016; Sayad, Mousannif, and Al Moatassime 2019).

The empirical landscape surrounding wildfire management has grown increasingly rich, illuminating the multifarious factors that impact both the incidence and outcomes of wildfires. Many studies such as the one by Hunter, Colavito, and Wright (2020), found that while climatic variables such as temperature and precipitation patterns have an impact on wildfire risk, anthropogenic factors including land management practices and socio-economic conditions are equally crucial. The post-2017 emphasis on integrated wildfire management was partially influenced by these findings, emphasizing a holistic approach (AGIF 2020).

A cross-national study by Moreira et al. (2011) compared Portugal's wildfire incidence with its Mediterranean neighbors, noting that despite similar climatic conditions, varying land management strategies and policies led to disparate outcomes. This underscores the argument that while environmental factors are consistent, human intervention can significantly modulate wildfire impacts.

2.1 Climate Change and Wildfire Risks

Building on the prior discussion, it's important to acknowledge that while climatic conditions are not the only contributors to wildfires, the effects of climate change on Portugal's landscape cannot be overlooked. The changes on the weather patterns in Portugal are varied and evident, notably in rising temperatures and aridity. Portugal follows global trends of warming trends, causing more hot and dry summers in contrast to milder winters (IPCC 2019). In every regional division of Portugal, the standard yearly air temperature has ascended over 0.3°C each decade since the 1970s. It is likely that this trend will continue throughout the rest of this century, with a predicted rise in extremely hot days and tropical nights compared to the period from 1971-2000 (IEA, 2021).

The temperature rise has been accompanied by a surge in evaporation rates, resulting in soil dryness and reduced water reserves. This alteration has led to longer and more severe droughts in Portugal, particularly in the southern areas (Portugal Government 2021). Studies indicate transformations in rain distribution, including reduced rainfall, especially during the spring and summer seasons (García-Ruiz et al. 2011). Portugal's average annual precipitation has decreased by about 25 mm per decade since 1970 and is projected to continue to decrease (IEA, 2021). Meanwhile, there is an increase in precipitation variability. This unpredictability in precipitation patterns could result in longer and more severe droughts and increased susceptibility to flooding.

The dual effect of increasing temperatures and changing precipitation patterns is acutely perceptible in connection with forest fires in Portugal. The higher temperatures result in higher evaporation rates, which dry out the vegetation and soil, increasing the susceptibility of the landscape to ignition and the spread of wildfires. Moreover, extended periods of high temperatures contribute to longer fire seasons, increasing the window of opportunity for potential wildfires (Sousa et al., 2015; Jolly et al., 2015). Decreased rainfall worsens these conditions, resulting in a more flammable environment favoring bigger and more severe wildfires. The correlation between reduced precipitation and heightened wildfire risk is well-supported in scientific literature (Romps et al., 2014; Westerling et al., 2006).

Recognizing the acute effects of rising temperatures and shifting precipitation patterns on Portugal's wildfire prevalence underscores the urgent need for global climate initiatives. It is within this context of heightened vulnerability and increased frequency of extreme weather conditions that the Paris Agreement emerges as a pivotal response. Formulated under the auspices of the United Nations Framework Convention on Climate Change, the agreement's ambitious targets to cap the global temperature rise are not merely aspirational but are grounded in scientific evidence linking temperature thresholds to escalating climate risks. Specifically, it aims to limit the increase in worldwide temperature to under 2 degrees Celsius, with additional attempts to further limit it to 1.5 degrees Celsius above the pre-industrial levels (IPCC 2019). These particular limits are based on empirical data indicating that a rise in temperature above this point would significantly intensify the hazards and repercussions of climate change, manifesting as increasingly frequent and severe bouts of extreme weather, elevated sea levels, and detrimental effects on ecosystems, human health, and economies (Delbeke et al. 2019).

Through international cooperation and commitment to these temperature goals, the agreement sets out to mitigate the very factors that exacerbate wildfire risks, thereby reinforcing the connection between global policy and local environmental realities.

2.2 Socio-Economic Factors and Demographic shifts

In the wake of establishing the Paris Agreement's critical role in addressing the environmental factors exacerbating wildfire risks, we shift our attention to another crucial element influencing these dynamics: socio-economic factors. Just as international accords seek to stabilize climate patterns, the socio-economic transformations within Portugal present a different but equally significant challenge.

The rural landscape of Portugal is undergoing a profound transformation, marked by a significant migration of the youth population towards urban epicenters. This demographic shift, chronicled by Almeida (2020) and Álvarez Lorente et al. (2020), extends beyond mere population movement, influencing the very fabric of rural economies and ecosystems. As rural youth relocate in search of enhanced prospects, the ensuing demographic vacuum raises critical questions about the future of these regions.

The intricate tapestry of rural Portugal has traditionally been woven with threads of agricultural practices passed down through generations. However, the shifting demographic patterns, as highlighted by The Portugal News (2019) and Project O.R. (2020), threaten to unravel this tapestry. Colonico et al. (2022) and Moreira et al. (2011) draw attention to the potential risks posed by the abandonment of these practices, suggesting an increased susceptibility of the landscape to wildfires. Yet, the literature often glosses over the youth's distinct role in this narrative, leaving a gap ripe for exploration.

As the youth demography recedes from rural settings, a shadow looms over the firefighting capabilities within these regions. Reports from the European Data Journalism Network (2023) shed light on the mounting challenges faced in managing and responding to wildfires in sparsely populated rural areas. Despite these insights, the literature remains limited on the nuanced impact of youth depopulation on the operational readiness and resources of firefighting units.

The socio-economic landscape serves as a backdrop against which the wildfire narrative unfolds. Research endeavors by Vilar et al. (2016) and Moreno et al. (2014) have ventured into this terrain, exploring the various socio-economic catalysts for wildfires. However, these studies often skirt around the specific influence of youth depopulation on wildfire incidence, leaving a conspicuous void in the discourse.

The dissonance between rural development initiatives and wildfire mitigation strategies is cast into sharp relief by Colonico et al. (2022). This discord underscores the pressing need for comprehensive policy frameworks that synergize rural development with wildfire prevention. By adopting an integrated approach that marries land management with socio-economic interventions, policymakers have the opportunity to mitigate the wildfire risks associated with demographic shifts.

2.3 Policy Interventions

As we navigate the effects of climate change as well as the socio-economic and demographic transformations in rural Portugal, it becomes apparent that a multifaceted approach is necessary for effective wildfire mitigation. The National Plan for Integrated Wildland Fire Management in Portugal embodies such a strategy, aiming to tackle wildfires from multiple angles. This plan, catalyzed by the aftermath of the catastrophic 2017 wildfires, signifies a critical juncture in policy intervention, as it transitions from traditional, reactive firefighting to a more holistic, proactive stance that incorporates land management, socio-economic initiatives, and demographic considerations into wildfire prevention and control. This policy intervention is not just a blueprint for action but an evolving mechanism that reflects the complexity of interlinking factors that exacerbate wildfire risks, including those stemming from the current demographic shifts (Gómez-González, Ojeda, and Fernandes 2018; AGIF 2020; Pandey et al. 2023).

The National Plan for Integrated Wildland Fire Management in Portugal is an intricate strategy that addresses the systemic risk of rural fires through a combination of governance, qualification, and information and communication systems. This approach aligns with both national and international frameworks for Disaster Risk Reduction. The plan is designed by the Agency for Integrated Rural Fire Management (*Agência para a Gestão Integrada de Fogos Rurais, I.P.* /AGIF), responsible for its strategic coordination and evaluation, as mandated by Decree-Law No. 12/2018 of 16 February 2018. The governance of this plan involves multi-level coordination (national, regional, and local), embracing a comprehensive approach to fire risk management. At the national level, the focus is on creating macro-policies and strategic guidelines that address fire risk through vegetation management and behavior change among landowners and rural land beneficiaries. This is supported by multi-annual budget programs for prevention and suppression efforts. At the regional and local levels, the plan promotes institutional consultation to ensure a bottom-up technical framework for managing fire risks effectively (AGIF 2023).

Moreover, the plan is not limited to a single sector but integrates various public policy domains like spatial planning, rural development, fiscal policy, nature conservation, energy strategy, education, communication, justice, security, and decentralization of competencies in local authorities. These sectors are collectively mobilized to address the complex problem of rural fires. The plan sets ambitious targets to safeguard Portugal against rural fires, aiming for a significant reduction in loss of lives, limiting large-scale fires (bigger than 500 ha) to less than 0.3% of total incidents, and controlling the cumulative burned area to below 660,000 hectares over a decade, subject to revision in 2023 (AGIF 2023). The 0.3% target for the proportion of big fires presents a challenge for scientific research as the term ‘wildfire’ was not clearly defined and potentially not in line with how data was recorded in the past.

To achieve these ambitious targets, the adage "prevention is better than cure" finds resonance in wildfire management literature. Preventive strategies range from controlled burns, vegetation management, to community awareness programs. Which is not only prone to be effective but economically viable (Pacheco and Claro 2021; Pinto, Sousa, and Valente 2022). Intriguingly, the role of time invested in preventive measures in Portugal, an aspect this thesis explores, has been relatively under-researched. However, preliminary findings suggest a potential inverse relationship between hours dedicated to preventive tasks and the severity of fire outcomes, a correlation that warrants deeper investigation (AGIF 2023; Simon, Crowley, and Franco 2022). Early assessments of the Portuguese National Plan (AGIF 2023) further suggested promising trends post the policy’s implementation, with a reduction in both the number of fires and the total area burned. However, a scientific evaluation is still pending. The lack of public access to detailed measures and spending plans within Portugal's National Plan for Integrated Wildland Fire Management presents a challenge for scientific research. The transparency of such information is crucial to scientifically assess, critique, and contribute to the improvement of public policies. Additionally, the plan's target of limiting large-scale fires to less than 0.3% of total incidents is hindered by the lack of a clear definition of what constitutes a wildfire. This ambiguity complicates the calculation of large fires in proportion to all fire incidents. A clear, standardized definition of wildfires is essential for consistent data collection and analysis, allowing for accurate tracking against stated goals and the effectiveness of the policy measures implemented.

Another layer of complexity is added when examining policy efficacy across regions. Just as environmental and socio-economic factors differ across Portugal, the success of policy interventions can also be heterogeneous. Beyond just the natural predisposition, local governance and resource allocation also play pivotal roles. Various studies (Paton and Tedim 2014; Paveglio, Abrams, and Ellison 2016; Skulska et al. 2020) postulated that regions with robust local community engagement and resource allocation towards preventive measures witnessed fewer devastating wildfires. This lends credence to the argument for decentralized and region-specific strategies in wildfire management. Collins et al. (2013) and Mateus and Fernandes (2014) documented varying degrees of policy impact across regions, hinting at the necessity for region-specific tailoring of interventions. Intricacies of regional disparities in wildfire outcomes within Portugal have drawn academic attention. Nunes (2012) and Carmo et al. (2011) delved into how varying topographies, land use patterns, and even local customs and traditions lead to different wildfire frequencies and intensities across districts. Northern mountainous regions, with their dense vegetation, have inherently higher fire risks than the more urbanized coastal regions.

2.4 Technological Advancements in Wildfire Prediction and Management

In the intricate web of factors influencing wildfire management, from socio-economic and demographic shifts to climate change and policy adaptation, we find that advancements in technology play an increasingly important role. As Portugal confronts the multifaceted challenges of wildfire prediction and management, it is the integration of technology that promises a new frontier in enhancing regional strategies. This fusion of tech-centric solutions with localized knowledge forms a synergistic force, enabling more precise forecasting and effective mitigation efforts tailored to the unique topographies and societal structures of the Portuguese landscape. As we turn to examine the strides made in technological innovation, it becomes clear that such advancements are not just complementary but integral to the evolution of wildfire management practices (Alkhatib et al. 2023; Scott et al. 2013).

2.5 Machine Learning Algorithms for Wildfire Prediction

Recent advancements in machine learning algorithms have significantly enhanced the accuracy of predicting wildfire risks, surpassing the capabilities of traditional approaches. These modern algorithms are adept at processing image data, such as satellite imagery, and can uncover complex relationships between risk factors, leading to improved model performance (Alkhatib et al. 2023). Machine learning techniques consider numerous factors such as weather conditions, vegetation, terrain, and human activity, and can manage large datasets efficiently.

Among classical machine learning models, Logistic Regression, Support Vector Machines, and Random Forests have been notably effective. Logistic Regression, a traditional classification algorithm, is well-suited for binary forest fire occurrence prediction as a linear classifier. However, it is sensitive to outliers and struggles with non-linear problems and high-dimensional datasets. To mitigate this, Ermitão et al. (2023) employed Principal Component Analysis for mapping susceptibility in Portugal, although Principal Component Analysis itself lacks validation for classification tasks. Support Vector Machines is proficient in binary and multi-class classification problems and, through kernel functions, can handle non-linear challenges found in feature-rich Geographical Information Systems (GIS) datasets. However, it requires careful data preprocessing due to its sensitivity to noise and parameter selection. Random Forests, particularly noted for its high classification accuracy, was used by Gibson et al. (2020) to achieve a 98% accuracy rate in predicting fires in eastern Australia. Other successful applications have been reported in Vietnam, Serbia, North Korea, and Spain (Tien Bui et al. 2016; Gigović et al. 2019; Jin and Lee 2022; Vilar et al. 2010).

Furthermore, ensemble methods such as Gradient Boosting and Extreme Gradient Boosting (XGBoost) have shown promise in wildfire prediction due to their capacity to handle non-linear relationships and interaction effects among variables. XGBoost, in particular, enhances prediction by combining weak learners sequentially (Chen & Guestrin, 2016). Deep learning techniques like convolutional neural networks (CNN) and recurrent neural networks (RNN) have also been explored for their potential in wildfire prediction. Convolutional neural networks, effective in image and spatial data analysis, are useful for interpreting satellite imagery and topographic data, while recurrent neural networks are adept at analyzing temporal sequences, which can include historical weather patterns and previous climatic events (Krizhevsky, Sutskever, & Hinton, 2012; Hochreiter & Schmidhuber, 1997).

Despite these advancements, the accuracy of machine learning models depends heavily on data quality and completeness. The dynamic nature of wildfires and climate patterns poses challenges, such as model overfitting or underperformance in unforeseen scenarios. Ongoing model evaluation and updating with the latest climate data are crucial to sustain the reliability of these predictive models.

2.6 Why We Should Care About Mapping

The advancements in machine learning technologies are also affecting ancient technologies, in particular maps. Mapping has been one of the most important human inventions and allows us to understand the world. As capabilities in data collection and distribution has exploded in modern times new uses has emerged, especially when understanding complex phenomena such as forest fires. With this change in information availability we can develop robust statistical models and reduce the reliance on expert systems (González, Kolehmainen, and Pukkala 2007; Jaiswal et al. 2002) to better understand these problems.

Mapping is an indispensable tool in both short- and long-term perspectives. Spatial Decision Support System (SDSS) tools are used by firefighters and forest sappers. They are vital for providing incident commanders with immediate, location-specific data, enabling quick and informed decision-making during emergencies. The real-time spatial information is essential in dynamic situations such as large-scale forest fire emergencies, where quick, data-driven decisions could prevent irreversible losses. Importantly maps can be used to observe changes over time, while simultaneously inspecting the individual factors that interplay in the cause and effects of fires. Mapping is being used by all levels of society to ensure preventative measures are effectively being employed. The US Forest Service, responsible for the management of land in national forests and supervision of responses to wildland fires within the US, relies on mapping to make policies, offer incentives for performance, and facilitate decision-making processes (Schultz, Thompson, and McCaffrey 2019). After the devastating forest fire season of 2017 in mainland Portugal, the government declared by law that mapping of the structural forest fire danger was necessary (Diário da República, 2017) as an addition to the AGIF. The first map was created in 2020 by the Institute for Nature Conservation and Forests (*Instituto da Conservação da Natureza e das Florestas /* ICNF) with the assistance of a private consulting firm Pal Consulting (S. Oliveira, Gonçalves, and Zêzere 2021).

The mapping process used the assumption that fire ignitions were non-random and could be linked to three main factors, namely topography (degrees of slope, and elevation), forest land cover (COS) and historically burned areas. Their method involved using a Bayesian-bivariate equation to classify each 25^2 m area of Portugal into 5 categories of structural fire danger. It is generally understood that the development of more advanced models improves prediction accuracies (Cruz et al. 2018) and the failure of linear regression to capture the relationships between multiple non-linear discriminants (S. Oliveira et al. 2012). This indicates that with more advanced models mapping of the susceptible regions could be improved.

2.7 Implications for current research

Amidst the backdrop of Portugal's changing landscape, the literature review reveals a pressing need for innovative research approaches in wildfire management (Gómez-González, Ojeda, and Fernandes 2018; IPCC 2019).

Building on recent advancements, the predictive prowess of machine learning models offers a promising frontier that merits further exploration (Alkhatib et al. 2023; Gibson et al. 2020). These technological tools could be adeptly tailored to the nuances of Portugal's geography and climate, thereby advancing the precision of wildfire risk assessments.

The escalation of wildfire risks due to climate change further commands a focused research agenda that evaluates the alignment of Portugal's policies with the Paris Agreement's ambitious targets (Sousa et al. 2015; Romps et al. 2014). Such studies offer invaluable insights into the effectiveness of national strategies against the backdrop of a warming globe.

Demographic shifts, especially the rural exodus of the younger population, present a pivotal area of study. Research spearheaded by Almeida (2020) and Álvarez Lorente et al. (2020) has initiated a discourse on the implications of such shifts for land management and wildfire preparedness. Future work should delve deeper into the ramifications of these changes, potentially reshaping Portugal's wildfire defense mechanisms. The socio-economic landscape of Portugal is inherently tied to wildfire management. While initial studies by Vilar et al. (2016) and Moreno et al. (2014) have helped to illuminate economic factors influencing wildfire incidence, there is an apparent lacuna concerning the direct impact of youth depopulation. Unraveling this intricate relationship will provide clarity on the socio-economic catalysts of wildfires.

Finally, policy interventions, as epitomized by Portugal's National Plan for Integrated Wildland Fire Management, necessitate rigorous evaluation to gauge their efficacy across diverse regions and communities (AGIF 2020; Pandey et al. 2023). Research here should not only assess the effectiveness but also the adaptability of these policies, taking cues from Paton and Tedim (2014) and Skulska et al. (2020), to inform a decentralized, context-specific approach to wildfire management.

This literature review has carved out a roadmap for this research, underpinned by a critical synthesis of the current literature, to navigate the multi-dimensional challenges of wildfire management in Portugal. The aim is to foster a research ecosystem that is both reflective and forward-looking, capable of informing robust, integrative strategies for a resilient Portugal.

3 Data sources

In this section, a comprehensive overview of the various data sources utilized in the study is provided, encapsulating a wide range of aspects crucial to understanding wildfire dynamics in Portugal. These sources include data on fires, weather conditions, policy impacts, demographic trends, topographic characteristics, road distances, and socio-economic insights, each offering unique insights and contributing to the holistic analysis conducted. This chapter serves as a pivotal foundation for the research, detailing the origins, nature, and relevance of each dataset, thereby establishing the groundwork for the subsequent analysis and findings presented in this study.

3.1 Fires

The foundation of the thesis is the data retrieved on the forest fires in Portugal. This data was obtained from the ICNF. The organization manages the sustainable development of natural resources, including wildlife, forests, and protected areas across Portugal. Their main responsibility lies in the protection of the country’s national and natural parks, including those protected by the Natura 2000 network (“The Natura 2000 Protected Areas Network — Eea.Europa.Eu,” n.d.) which make up the most valuable natural areas of the European Union. Since Portugal is one of the EU countries most prone to forest fires, a significant portion of conservation efforts are focused on effective forest management to prevent and combat wildfires.

Their mission includes registering instances of forest fires in Portugal, which is made available through the Forest Fire Information Management System (*Sistema de Gestão de Informação dos Incêndios Florestais* / SGIFR). The system has been operational since 2001. The SGIFR collates information on fires in Portugal between 1980 and 2023 and is continuously updated to reflect the current state of the country. The database collects information from all fires regardless of size and had at the point of retrieval 803,696 fires registered in mainland Portugal. The mean size of each fire in the collected data was 6,15 hectares, with extremes as high as

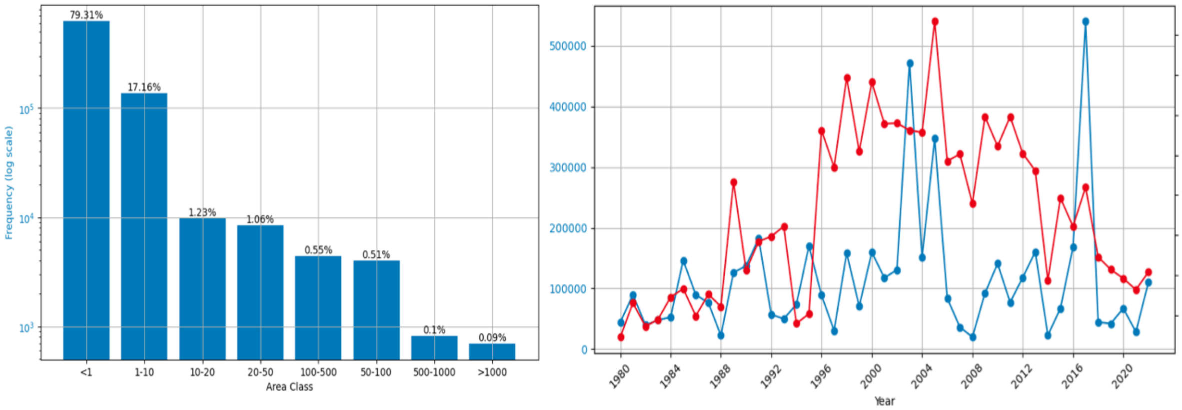


Figure 1: Fire size and frequency log scale, temporal distribution. Data source: ICNF

53,618 hectares and low as 0,1 hectares. We observe from the distribution of fires in Figure 1 that a disproportionate spread is present in the data, as both median and mode is equal to 0,1 hectares, indicating a skew in observed samples. Causes of fires have been registered but with several limitations. Slightly above 50 percent of the registered fires lack causes altogether, and 60 percent of the registered causes are unknown.

Collection methods incorporate satellite, GPS measurements, and external monitoring from physical watchtowers, as well as public and agency reporting. The database has undergone development and entries after the establishment of the SGIFR offer more detail.

The forest fire database categorizes its data into multiple primary sections, documenting the temporal and geographical specifics of every fire. This includes the initial sighting time, response times, and precise location crucial data such as coordinates, region, district, and parish. The database also records information on the fire itself, such as its duration, overall size, as well as explicit descriptions of the area and vegetation impacted. Furthermore, the database incorporates computed variables for assessing fire weather conditions using the Canadian fire weather index (FWI) system. Finally, the database classifies fire causes into general categories including accidental, intentional, natural, or unknown, and provides more detailed descriptions whenever possible. To create a comprehensive database of all available data, certain data-processing was necessary. The data from each source file underwent normalization into a standard format, with elimination of extraneous variables. All point geography was adjusted to *EPSG:3763 ETRS89/TM06 Portugal* and linked to a unique identifier.

3.2 Weather

The weather dataset was obtained from the OpenMeteo Historical Weather API, which uses ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Hersbach et al. 2020). The dataset is comprehensive, ranging from 1980 to 2022, and provides a wide range of climate variables relevant to the study of conditions driving wildfires. The ERA5 dataset is the fifth iteration of ECMWF reanalysis and provides a comprehensive time series of atmospheric, land-based, and marine parameters, with records dating back to 1940.

This study focused on three primary variables, namely temperature, precipitation, and wind, which play a crucial role in igniting and spreading wildfires. The role of wind is especially crucial as it directly affects the speed and direction of wildfire propagation. The high resolution, both in space and in time, of the ERA5 data offers a comprehensive view of these climatic variables, enhancing the understanding of their impact on wildfire behavior.

3.3 Policy

The policy dataset used in this study, sourced from Statistics Portugal (*Instituto Nacional de Estatística* / INE), provides a detailed and granular perspective on various factors influencing wildfire management from 1989 to 2021. The Nomenclature of Territorial Units for Statistics (NUTS) is a geocode standard for referencing the subdivisions of countries for statistical purposes. This dataset's granularity, mostly on a yearly basis and down to the NUTS Level 3, offers a unique insight into regional differences and specificities. In this study, we are working with NUTS regions, which allows for a more detailed and localized analysis of policy impacts and wildfire trends. The dataset encompasses 50 variables, broadly categorized into several key areas: resources and manpower of fire brigades, hours spent on forest fire prevention, involvement of non-governmental organizations in environmental preservation, forest management and land use data, and economic indicators relevant to fire management. These variables include specifics such as the number of fire brigades and firemen, investments and expenditures by fire brigades, the proportion of special protection areas, forest species distribution, and economic data related to municipalities and forest products.

The policy dataset from INE is further examined through a series of graphical representations to discern trends over time. The subsequent figures provide an empirical backdrop to the narrative of wildfire policy and resource allocation in Portugal.

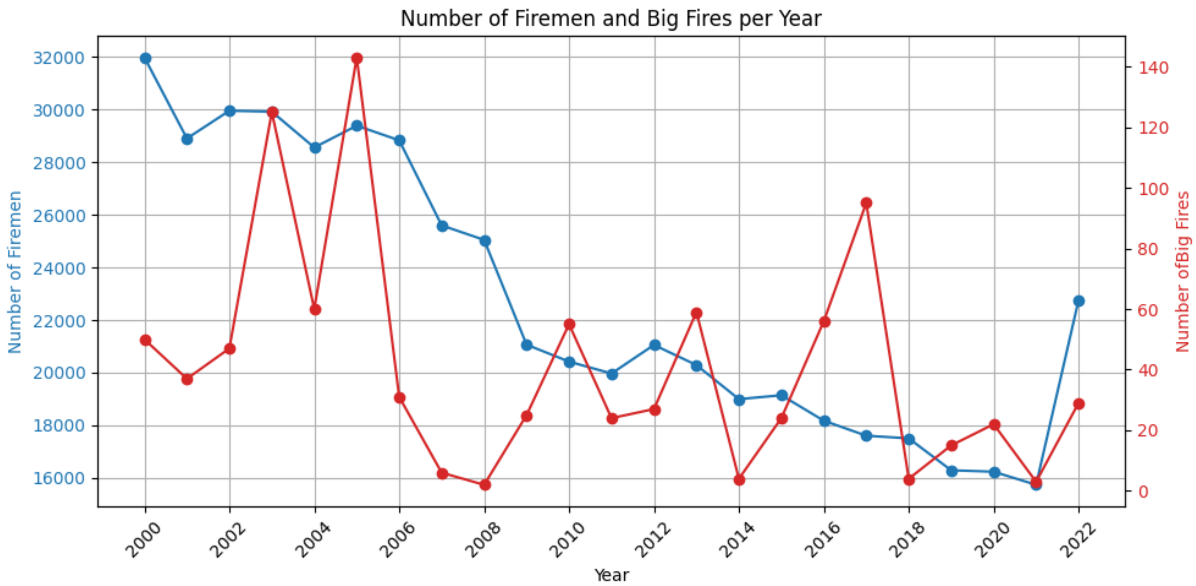


Figure 2: Number of firemen and big fires from the years 2000-2022. Data source: INE

Figure 2 portrays the number of firemen from the year 2000 through 2022. A general decline in the number of firemen is observable from the early 2000s until 2021. The year 2022 marks an exceptional upsurge, which could indicate a policy response or an increase in recruitment initiatives.

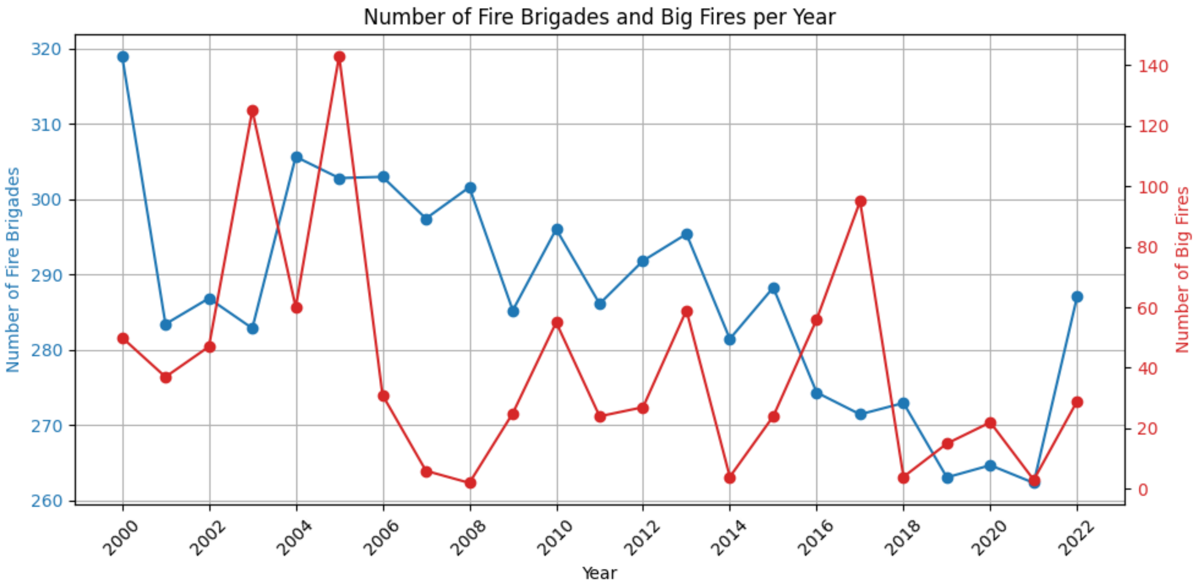


Figure 3 Number of fire brigades and big fires from the year 2000-2022. Data source: INE

Figure 3 reveals the fluctuation in the number of fire brigades over the same period. The initial years exhibit a sharp decrease, followed by a period of relative stability with intermittent peaks and troughs. The count of fire brigades then slowly declines until 2021 with a notable upsurge in 2022.

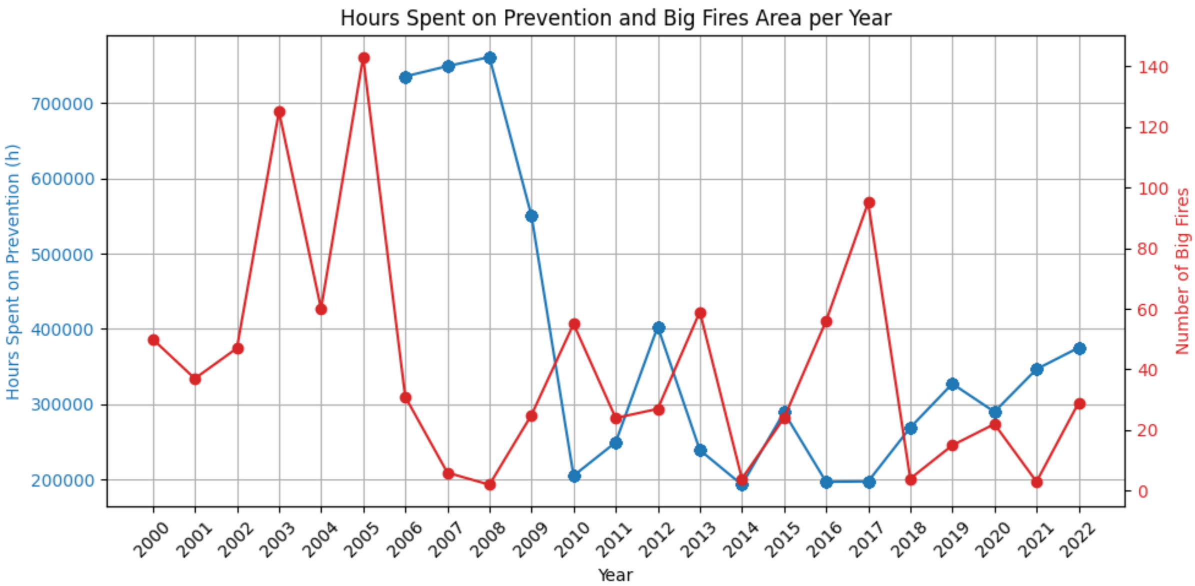


Figure 4: Hours dedicated to fire prevention annually by the Department of Nature and Environment, and number of big fires. Data source: INE

Figure 4 showcases the hours dedicated to fire prevention annually by the Department of Nature and Environment, in contrast to the number of large fires. Notably, the highest recorded prevention efforts around 2005 and 2006 preceded by a sharp decline.

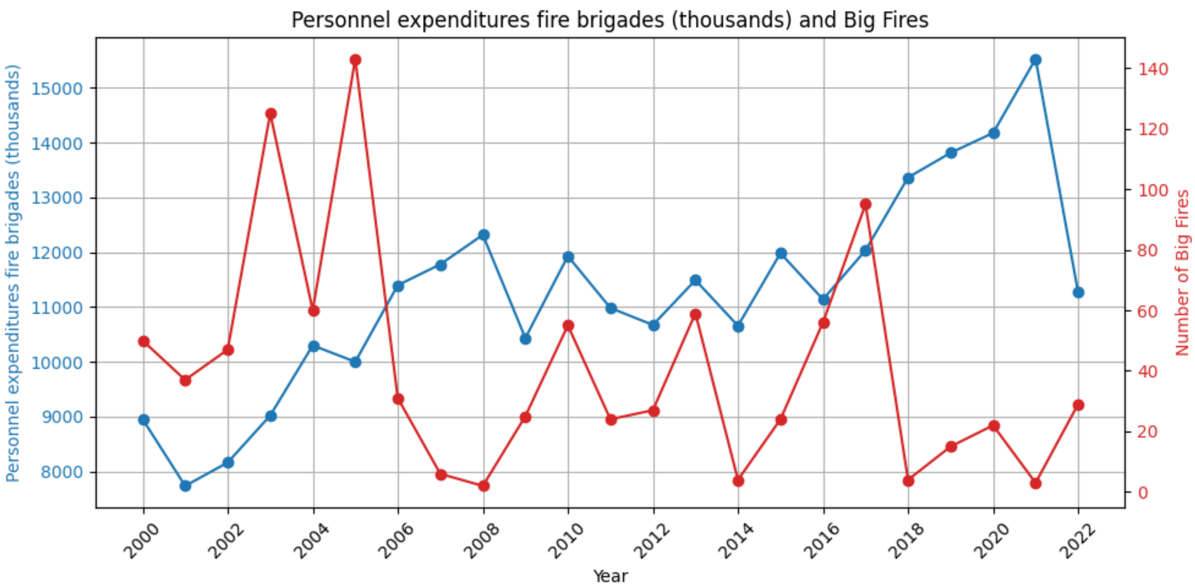


Figure 5: Personnel expenditures fire brigades (thousand) and number of big fires. Data source: INE

Figure 5 examines the personnel expenditures of fire brigades in thousands, alongside the number of large fires each year. While expenditure on personnel has generally increased since

2000, the frequency of large fires does not mirror this trend. This incongruity points to the need for a nuanced analysis of how financial resources are allocated and their direct or indirect effects on wildfire incidence and management efficacy.

These visual analyses provide a clear indication of the temporal trends in human and financial resources dedicated to wildfire management and their initial, apparent relationship with wildfire occurrences. Nevertheless, it is imperative to apply statistical methods to further investigate these relationships and to understand the underlying causative factors.

For a detailed overview of these variables and their metadata, refer to the variables' table in the appendix.

3.4 Economics

The study further incorporates key economic variables to provide a broader context for economic trends in sectors closely related to forest management and the broader national economy. Three critical economic indicators were retrieved from reliable financial data sources, namely Yahoo Finance API and Bloomberg Terminal.

The first variable, *Close_wood*, represents the daily closing price of the ETF tracking companies in the forestry and timber sector, identified by the ticker symbol WOOD. This data, sourced daily from Yahoo Finance API from 2008 to 2022, offers insights into the economic performance of the forestry sector, which is directly relevant to wildfire management and policy analysis. The second variable, *Close_cor*, is the daily closing price for CORTICEIRA AMORIM, the world's largest cork and cork product producer, trading under the ticker symbol COR. Also obtained daily from Yahoo Finance API, this data spans from 1995 to 2022, providing a long-term perspective on the economic trends in a key industry directly affected by wildfire dynamics.

Lastly, the *GDP* variable, sourced annually from the Bloomberg Terminal, reflects the most recent quarter-over-quarter percentage change in Portugal's Gross Domestic Product (GDP).

Available from 1988 to 2022, this variable provides a macroeconomic view of Portugal's economy, offering a backdrop to understand the broader economic implications of wildfires.

3.5 Demographics

The demographic data, sourced from the National Statistics Institute (2021), provides granular insights into the age composition of the population. The demographic variables in the dataset include *Population Density*, *Total Population*, resident population by age groups (e.g. *Population 20 – 24 years*), *Region Unemployment Rate*, and the number of campers per region clustered by nationality (e.g. *Number of Campers France per Region*). These variables were derived from census data provided by the INE.

The *Total Population*, the population by age groups, and the *Population Density* are retrieved on a NUTS level 3, where the *Population Density* is calculated as the number of individuals per square kilometer in each district.

The *Region Unemployment Rate* and the number of campers is retrieved on a regional level, equivalent to the granularity of NUTS level 2.

The study of demographic data in this thesis sheds light on the significant trends in rural depopulation over the past two decades, specifically from 2001 to 2021. This decline is particularly pronounced in districts classified as rural, defined by a population density of less than 100. A detailed analysis of six rural districts, namely Portalegre, Guarda, Castelo Branco, Beja, Evora and Braganca, reveals a consistent pattern of population decline. In Portalegre, for example, the total population will fall from 25,957 in 2001 to 22,250 in 2021, a decrease of 14.28%. Similar trends are observed in the other districts, with Guarda and Castelo Branco experiencing a decline of 8.39% and 5.95%, respectively.

Table 1: Rural Depopulation Trend (2001 - 2021), Data Source: INE

District	**Total Population (2001)**	**Total Population (2021)**	**Absolute Change**	**Percentage Change**
Portalegre	25,957.0	22,250.0	-3,707.0	-14.28%
Guarda	43,811.0	40,137.0	-3,674.0	-8.39%
Castelo Branco	55,793.0	52,472.0	-3,321.0	-5.95%
Beja	35,766.0	33,644.0	-2,122.0	-5.93%
Evora	56,552.0	53,881.0	-2,671.0	-4.72%

Braganca	34,797.0	34,737.0	-60.0	-0.17%

This demographic change is not limited to the general population but is also evident in the youth segment of the population. The youth population in these rural areas has decreased significantly over the same 20-year period. In Portalegre, the youth population fell by 39.46%, from 5,149 to 3,117. The other districts followed suit, with Guarda, Castelo Branco and Beja recording reductions of 34.21%, 33.47% and 31.55%, respectively.

Table 2: Rural Youth Depopulation Trend (2001 - 2021), Data Source: INE

District	**Youth Population (2001)**	**Youth Population (2021)**	**Absolute Change**	**Percentage Change**
Portalegre	5,149.0	3,117.0	-2,032.0	-39.46%
Guarda	9,191.0	6,047.0	-3,144.0	-34.21%
Castelo Branco	10,764.0	7,161.0	-3,603.0	-33.47%
Beja	12,046.0	8,245.0	-3,801.0	-31.55%
Evora	7,435.0	5,158.0	-2,277.0	-30.63%
Braganca	7,400.0	5,385.0	-2,015.0	-27.23%

In addition, the evolution of rural youth depopulation is further clarified by examining the youth population from 2009 to 2021, which are presented in a graph. This graphical analysis complements the tabular data, providing a visual narrative of the declining trend in rural youth population over the specified period. This comprehensive demographic examination provides a critical foundation for understanding the broader implications of rural depopulation, particularly in the context of wildfire severity and frequency, which is central to this thesis.

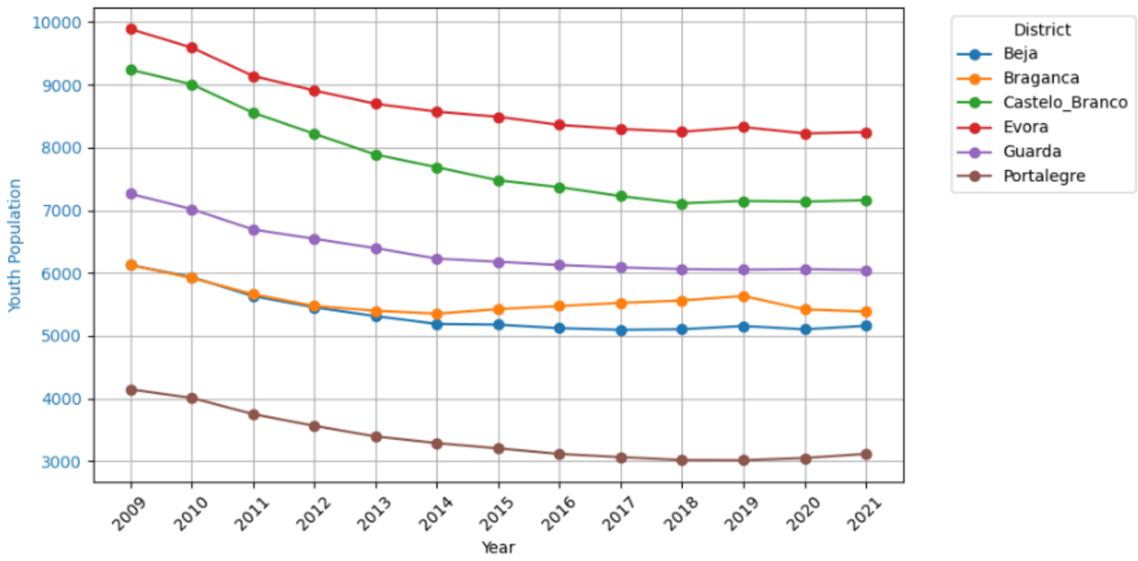


Figure 6: Evolution of Rural Youth Population (2009 - 2021), Data Source: INE

3.6 Topographic

The topographic data included in the dataset was retrieved from the Universidade de Porto. This data includes the *Elevation*, *Aspect*, and *Slope* for each exact location of the ignited wildfires from the dataset. The topographic data was initially retrieved in Geotiff format which were subsequently processed using the Geodata software Quantum-GIS (QGIS) to map each data point geographically to the fire data. This data represents an instance obtained in 2018, recorded in a resolution of 100 meters. From these variables, the elevation as well as the slope present the most significant impact on the severity of an ignited wildfire. Elevation affects local climate, which in turn alters vegetation types and moisture levels, key components in wildfire behavior. Slope plays an important role in the speed and direction of fire spread. Fires tend to move faster uphill due to the preheating of vegetation, with steeper slopes exacerbating this effect.

3.7 Socio-Economic

As socio-economic data, the variable *land_use* was gathered. Land use refers to the information that describes how various regions of land are utilized by humans. This data categorizes land based on various uses such as residential, commercial, agricultural, industrial, and recreational. The data was retrieved from the Copernicus Land Monitoring Service (CLMS) as TIF file. The original information was obtained by using GIS in a resolution of 100 meters and was procured in 2018.

3.8 Road distance

The *road_distance_km* variable indicates the distance in kilometers between any given point in Portugal and the closest major roadway. This information was obtained from the WorldPop (“WorldPop :: Geospatial Covariate Data Layers” n.d.) project, which focuses on comprehensive geospatial data sets. The specific data collection for Portugal was procured on November 2, 2018, in Geotiff format—a raster image format frequently employed in geographic information and satellite imagery handling. The dataset has a resolution of 3 arc-seconds, which is equivalent to approximately 100 meters at the equator. This resolution is suitable for detailed regional analysis. The dataset's projection uses the WGS84 system, a standard in global positioning systems, ensuring compatibility with various mapping and spatial analysis tools. The raster values indicate the distance from the center of each cell to the closest major road intersection, as mapped by the OpenStreetMap project in 2016.

3.9 Data Merging

In constructing a comprehensive and coherent dataset for wildfire risk analysis, the study used a dual-method approach to effectively merge the collected variables. For certain variables, particularly all weather-related data, the merging was performed directly using the exact geographic coordinates (latitude and longitude) of each fire event. This method provided a high degree of accuracy by directly associating each fire event with its immediate climatic conditions.

In contrast, other variables were integrated using a two-step approach, primarily when the data were not directly linked to specific fire events. First, the exact coordinates of each fire event were mapped to the NUTS regions polygons, which are divided into three levels for subdividing countries, with level 3 being the most detailed. This approach was particularly applied to policy data, which was merged based on NUTS level 3 areas, providing an indirect but informed link to fire incidents.

The levels of aggregation varied for different variables, reflecting the nature of the data and its relevance to the study. For example, data on the number of fire brigades, firefighters, firefighting services, and specific firefighting activities were merged at the NUTS level 3, closely aligning them with individual fire incidents. Economic variables, which capture broader regional economic data, were integrated at the NUTS level 1. Similarly, demographic variables largely followed the NUTS level 3 of integration, with exceptions such as camper data and unemployment rates, which were merged at the NUTS level 2 due to their broader scope.

Satellite-derived data, including essential topographic information like elevation, slope, aspect, road distance, and land use, were methodically merged for each fire incident. The Geodata software QGIS was utilized to extract satellite data coordinates to accomplish this. Specifically, these coordinates were matched with the precise latitude and longitude of each fire incident, ensuring a direct and precise integration of topographical and land use data.

The systematic merging of data played a crucial role in facilitating this study's capacity to deliver a sophisticated analysis of the diverse factors that affect wildfire risks. By comparing a variety of data sets with geographic information regarding the sites or relevant administrative regions affected by fires, the study could conduct an all-encompassing evaluation of the interplay among environmental, policy, economic, demographic, and topographical factors in relation to wildfire incidents. The accuracy and reliability of the dataset formed the foundation of the research, facilitating a thorough investigation of the intricate dynamics involved in wildfire incidents.

4 Introduction to Individual Parts

This part of the thesis provides a detailed exploration of wildfire research and management through four distinct but interconnected subsections, each addressing a unique aspect of this complex field. The first sub-section explores the use of machine learning algorithms to identify regions in mainland Portugal that are highly susceptible to wildfire. It focuses on the structural hazards posed by specific locations, land cover and human intervention, and aims to reduce the societal costs associated with such fires by integrating historical data, environmental variables, and human activity patterns.

The second subsection explores the impact of climate change, specifically the implications of the Paris Agreement, on the intensity of wildfires in Portugal. Using machine learning, this section analyses and predicts how specific climate changes affect wildfire patterns, providing insights that are crucial for policymaking and environmental management.

The third part presents an empirical evaluation of Portugal's National Plan for Integrated Wildland Fire Management, launched in 2018. This part critically assesses the effectiveness of this national strategy in mitigating wildfire damages, in line with the overarching theme of data-driven insights in environmental management and exemplifies the application of analytical techniques in evaluating and informing public policy decisions.

The final sub-section explores the relationship between rural youth depopulation and the escalation of forest fires in Portugal. It examines how demographic change, in particular the migration of younger generations to urban centers and the abandonment of traditional agricultural practices, affects the frequency and severity of forest fires. This study seeks to understand the complex interplay between demographic trends and environmental challenges.

Taken together, these sections provide a comprehensive overview that highlights the importance of integrating different data sources, analytical techniques, and theoretical frameworks. Together, they offer a multifaceted perspective on understanding and managing wildfire, highlighting the evolving challenges posed by climate change and socio-economic dynamics.

5 Predicting Wildfire Risks in Portugal: A Machine Learning Approach under the Paris Agreement Climate Targets

5.1 Introduction

This research study analyses the impact of climate change on wildfire intensity in Portugal, in the context of the Paris Climate Agreement. It explores the use of machine learning algorithms to analyze the relationship between climate change and forest fire patterns, with the aim of providing actionable insights for policy makers and resource managers. The research looks at climate trends in Portugal, including temperature increases and precipitation shifts, and their subsequent impact on wildfires. Using datasets from different Shared Socioeconomic Pathways (SSPs) and applying advanced machine learning techniques, the study seeks to understand how specific climate variables and international climate policies influence wildfire dynamics. The methodology involves data collection, cleaning and analysis to identify the most appropriate predictive models. This is followed by hypothesis testing using machine learning and regression analysis to assess the relationships between climate variables and wildfire risk. The thesis not only contributes to the academic discourse, but also has practical implications for wildfire and climate change mitigation strategies.

5.2 Literature Review and Gap Analysis

Positioning itself within the global discourse on climate action, the study primarily examines the implications of the Paris Climate Agreement and explores the potential of machine learning algorithms for insightful predictive analysis. A major global commitment to tackling climate change was made in 2015 with the Paris Climate Agreement, a landmark pact under the United Nations Framework Convention on Climate Change. Its ambitious goal of limiting global warming to below 2°C is aimed at mitigating the direct consequences of climate change, including but not limited to extreme weather events and sea level rise (IPCC, 2019). Portugal's experience with climate change reflects these global trends. The country has experienced notable increases in temperature and shifts in precipitation patterns, changes that have significantly altered its environmental and ecological dynamics. These climatic changes have led to increased vulnerability to forest fires, a phenomenon that has increased in both frequency and severity in recent years (Hungary Climate Resilience Policy Indicator - Analysis - IEA, 2021). There is a well-established link in the scientific literature between climate change and the escalation of forest fires. Studies have consistently shown that the combination of higher temperatures and changes in precipitation patterns, such as those observed in Portugal, create conditions that are more conducive to the outbreak and spread of wildfires (Sousa et al., 2015; Jolly et al., 2015). It highlights the importance of a comprehensive understanding of how these climate factors interact to affect fire patterns, especially under the threat of ongoing global climate change. In addressing this need, the use of machine learning algorithms is emerging as a promising and innovative approach. These advanced computational methods such as random forests and neural networks, are recognized for their ability to analyze complex datasets and uncover hidden patterns. These tools are particularly adept at predicting risks and trends in environmental phenomena, including wildfires (Hastie et al., 2009; Biau & Scornet, 2016; Chen & Guestrin, 2016). By focusing on Portugal, a region that has been significantly affected by wildfires, this study aims to fill a gap in current research. It seeks to move beyond a generalized understanding of the implications of global warming on wildfires to a more nuanced exploration of how specific climatic variables play a role in shaping wildfire dynamics.

5.3 Hypothesis Development

To investigate the complex interactions between climate change, policy interventions and wildfire risks in Portugal, this thesis presents a structured, hypothesis-driven approach. The research is based on a primary hypothesis, with two additional hypotheses serving as supporting investigations to provide a holistic understanding of wildfire dynamics under changing climatic and policy conditions.

H1 (Main Hypothesis): *Adherence to the Paris Climate Agreement's objectives causes a reduction or stabilization in the intensity of wildfires in Portugal, as evidenced by decreased* *burned areas*. This hypothesis draws on the findings of Gómez-González, Ojeda, and Fernandes (2018) and Turco et al. (2019), which examine the impacts of climate change mitigation policies on environmental outcomes, including wildfire frequency and severity.

H2a (supporting hypothesis): *An increase in temperature, as a result of ongoing climate change, directly causes an increase in the area burned by wildfires in Portugal*. This relationship is supported by the studies of Sousa et al. (2015) and Jolly et al. (2015), which provide evidence of the correlation between rising temperatures and increased wildfire incidence, particularly in Mediterranean climates.

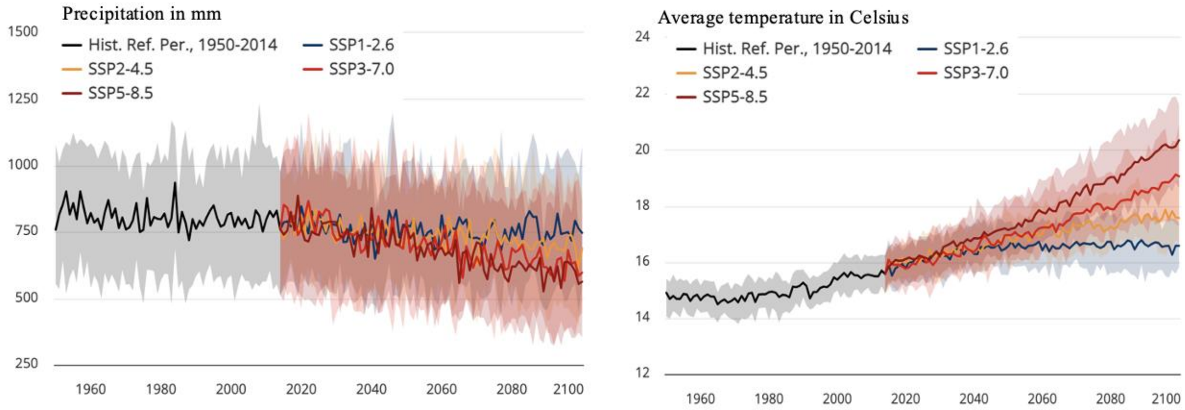
H2b (supporting hypothesis): *A decrease and change in precipitation patterns, consistent with climate change projections, leads to an increase in the area burned by wildfires in Portugal.* This hypothesis aligns with the findings of Westerling et al. (2006) and Romps et al. (2014), which explore the consequences of reduced precipitation on wildfire risks, demonstrating that drier conditions can lead to larger and more intense fires. Together, these hypotheses are intended to provide insights into the dynamics of wildfires in Portugal. With H1 as the focus, and H2a and H2b providing additional depth, the study will use advanced machine learning models, using historical data and future climate projections, to analyze and understand the complex relationships between climate variables and wildfire risks.

5.4 Methodology

The hypotheses form the basis for a broad empirical investigation. This leads to the methodology section of the study, which outlines the specific procedures and techniques used to collect, process and analyze the relevant data.

5.4.1 Prediction Data

To comprehensively evaluate the impacts of global warming on wildfires, four distinct datasets were created from different Shared Socioeconomic Pathways (SSPs) within the Coupled Model Intercomparison Project Phase 6 (CMIP6). Each dataset represents diverse future climatic scenarios for Portugal, with varying environmental outcomes characterized by a unique set of monthly weather data. These outcomes have been meticulously detailed in the Intergovernmental Penal on Climate Change (IPCC) Working Group summary for



policymakers (IPCC, 2021). The first dataset, which conforms to the SSP5-8.5 pathway with high greenhouse gas emissions, presents a scenario wherein CO2 emissions almost double by 2100, leading to a future with increased temperatures and reduced rainfall. This dataset plays an important role in scrutinizing the risk factors related to extreme wildfires given low climate change mitigation efforts (O'Neill et al., 2017). The second dataset, based on SSP3-7.0, predicts a future with significant GHG emissions, although slightly lower than the first dataset. The model predicts a moderate rise in temperature and a slight decrease in rainfall. It provides a perspective on the impact of significant yet not extreme change in climate patterns. The third dataset is based on SSP2-4.5 and represents a medium-level scenario for greenhouse gas emissions, where CO2 levels stabilize by the middle of this century. This scenario projects a less severe temperature rises and relatively stable precipitation patterns (O'Neill et al., 2017). Finally, the data from SSP1-2.6, the most ecologically sustainable projection, is in accordance with the objectives of the Paris Agreement. This dataset projects a decrease in temperature rise and an increase in precipitation rates (Group, 2021; O'Neill et al., 2017).

Figure 7: Different SSP scenarios., Source: ("Portugal - Mean Projections Expert | Climate Change Knowledge Portal", 2021)

Figure 11 depicts annual precipitation patterns and average temperatures in Portugal. On the left it showcases precipitation tendencies from 1995 to 2014 and predicts the same for 2014 to 2100 under the impact of four separate Shared Socioeconomic Pathways (SSPs). These projections introduce plausible changes in precipitation patterns, a vital element in the evaluation of future hazards related to wildfires. The right side illustrates temperature trends during the same period, differentiated by the SSPs and highlighting the varied climatic alterations. Precipitation and temperature build the foundation of the predictive data.

5.4.2 Data Cleaning and Preprocessing

Data cleaning and preprocessing is crucial to ensure the accuracy and reliability of subsequent model predictions. This phase involves rectifying inconsistencies or inaccuracies in weather datasets, which is important to maintain the dataset's integrity (Levine and Wilks 2000). An integral aspect of this stage consists of identifying hot days, categorized as exceeding 30 degrees Celsius, which offer a substantial indication of the summer's severity, a pivotal factor in the manifestation of wildfires (Westerling et al. 2006). Additionally, an evaluation of the yearly precipitation is undertaken, acknowledging that the dispersion of rainfall over prolonged periods is widely recognized to have a crucial impact on wildfire hazard. This analysis is complemented by reviewing monthly rainfall data, particularly in summer months where wildfires historically occur at higher rates.

5.4.3 Feature Selection and Engineering

In the wildfire prediction model, a careful selection of meteorological variables has been made, focusing on those that not only have a strong correlation with fire occurrence, but also remain predictable over long periods of time. This balance is essential to improve model accuracy and reliability. Certain variables characterized by high unpredictability over longer time scales were deliberately excluded. For example, wind speed, although relevant to wildfire spread, was omitted due to its high variability and the challenges of accurately predicting it over long periods. The inclusion of such unpredictable variables could introduce a level of uncertainty that could compromise the overall accuracy and robustness of the model. The variables that were considered are listed below:

mean_temperature_month: Reflects average monthly temperatures, relevant for their impact on drying vegetation. *precipitation_month:* Indicates total monthly rainfall, essential for understanding moisture levels. *hot_days_year:* Counts days with temperatures above a high threshold (30° maximum temperature), indicative of extreme fire risk conditions. *precipitation_year:* Annual precipitation total, influencing long-term environmental moisture. *precipitation_jja:* Measures summertime rainfall, critical during the peak wildfire season. *average_max_temperature_month:* The monthly average of peak temperatures, highlighting periods of extreme heat. To ensure the integrity of the model and to address potential multicollinearity, a Variance Inflation Factor (VIF) analysis was conducted. The results, detailed in the Appendix (see Table B1), show that the VIF values for the selected features are below the critical threshold of 5-10, suggesting that multicollinearity is not a significant problem, thus reinforcing the statistical reliability of the model.

5.4.4 Model Selection and Training

To accurately forecast the impact of climate change on wildfire intensity and frequency, it is mandatory to choose a machine learning model that captures historical data nuances and boasts dependable predictive capabilities. This selection procedure consists of comparing various models using established error metrics. Mean-Squared Error (MSE) is a commonly used measure of the quality of an estimator. It is computed by averaging the squared deviations between the predicted values ($\hat{y}_i$) and the actual values (y_i) (Chicco, Warrens, and Jurman 2021):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

where *n* is the number of observations. The smaller the MSE, the better the model fits the data because the prediction/actual differences are smaller. R square, or coefficient of determination, measures how well the independent variables explain the variation in the dependent variable. It is calculated as:

$$R^2 = 1 - \frac{Sum\ of\ Squared\ Residuals\ (SSR)}{Total\ Sum\ of\ Squares\ (SST)}$$

Where:

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

and $\bar{y}$ is the mean of the actual values. R-squared values range from 0 to 1, with higher values are indicating a better fit (Chicco, Warrens, and Jurman 2021). The analysis compared multiple machine learning models, evaluating each based on MSE and R-squared values. The basic framework of any prediction model and the simplest, is linear regression. The linear regression model employed for the prediction is represented by the following equation:

y = β_0 *+ (*β_1** mean_temperature_month) + (*β_2** precipitation_month) + (*β_3** hot_days_year) + (*β_4** precipitation_year) +*β_5 ** precipitation_jja) + (*β_6** average_max_Temperature_Month)*

For the Random Forest, Gradient Boosting, Extreme Gradient Boosting (XGBoost), and Simple Feed-Forward Neural Network (FFNNs) models, the equations are not as straightforward as for linear regression due to their complex, non-linear nature. However, these models were evaluated based on the same error metrics, as seen in Table 5.

Table 3: Evaluation of the models

Model name Error code	*Linear Regression*	*Random Forest*	*Gradient Boosting*	*Extreme Gradient Boosting (XGBoost)*	*Simple feed-forward neural network (FFNNs):*
Mean-Squared Error	13206621856.6	12737.8	4866097238.5	48838467	2532536907.9
R-squared error	0.0345	0.9999	0.6643	0.9964	0.8149

After completing the training for all models, a comparison was made between the various error codes. Linear regression, gradient boosting, XGBoost, random forest and simple feed-forward neural networks were evaluated. However, following findings in related studies (Gao, Lin, and Hu 2023; Latifah et al. 2019) the random forest model demonstrated the highest accuracy. This was indicated by its low Mean-Squared Error and near-perfect R-squared value, demonstrating a superior fit compared to the other models.

A train/test split approach was used to train these models using the historical wildfire data. This allowed the performance of the model to be evaluated on unseen data, ensuring the robustness of the predictive capabilities. The trained models were then used to analyze and predict future wildfire scenarios.

Table 4: Feature Importance

Rank	Feature name	Importance score
1	precipitation_jja	0.43476
2	precipitation_year	0.31181
3	hot_days_year	0.25103
4	mean_temperature_month	0.00233
5	precipitation_month	0.00007
6	average_max_temperature_month	0.00000

The feature importance analysis of the Random Forest, as seen in Table 6, model showed that the three most predictive variables for wildfire risks were summer precipitation, yearly rainfall, and the number of hot days annually. Monthly mean temperature and precipitation had a lower impact, whereas the average maximum temperature per month had the least influence.

Building on the comprehensive evaluation of the machine learning models, the research study also includes an Ordinary Least Squares (OLS) regression analysis to further explore the climatic factors influencing wildfire areas in Portugal, the exact values can be retrieved from the Appendix (see Table B2). This analysis, which yielded an R-squared value of 0.035, complements the machine learning results by quantifying the specific contributions of individual climate variables. In this detailed analysis, *hot_days_year* emerged with a positive correlation to wildfire areas, reinforcing the importance of extreme heat conditions in escalating wildfire risks. In contrast, an unexpected negative correlation was observed for *mean_temperature_month*, suggesting that mean temperature alone may not fully capture the nuances of wildfire risk, possibly due to other underlying environmental factors. Precipitation patterns, represented by *precipitation_month* and *precipitation_jja* (summer precipitation), showed variable effects on wildfire areas, adding another layer to the complex interplay of climatic influences.

5.5 Results

For the results section, the trained Random Forest model, utilizing historical data, was applied to four different future weather datasets. These datasets each represent a future climate scenario, as previously mentioned. Figure A1 in the appendix presents the findings of this analysis. The dark blue line displays historical wildfires from 1980 to 2020, and from 2020 onwards, it exhibits the projected annual burned area in hectares for each scenario until 2100. The SSP 6.0 and 8.5 scenarios (represented by the red and green lines) exhibit a significant incidence of severe wildfires comparable to the devastating fires seen in Portugal in 2017. By contrast, such severe wildfires are significantly less likely in the best-case scenario, SSP 2.6, which reflects the Paris Agreement's aims. The upcoming section will test the hypotheses against the prediction results.

5.5.1 Hypothesis testing

H1 (Main Hypothesis): *Adherence to the Paris Climate Agreement's objectives causes a reduction or stabilization in the intensity of wildfires in Portugal, as evidenced by decreased burned areas.*

This hypothesis was tested by examining the projected impacts of different shared socio-economic pathways (SSPs) on wildfire severity in Portugal. Each SSP reflects a different level of commitment to the goals of the Paris Climate Agreement, with SSP 2.6 closely aligned with the ambitious climate goals of the Agreement. The analysis used a comprehensive model to simulate future wildfire scenarios under each SSP. The focus was on contrasting the SSP 2.6

scenario, which indicates strong adherence to the Paris Agreement, with other SSPs that represent less stringent climate action. The model's projections under SSP 2.6 showed a reduction of 26% in the average area burned in comparison to the historical burned area, as

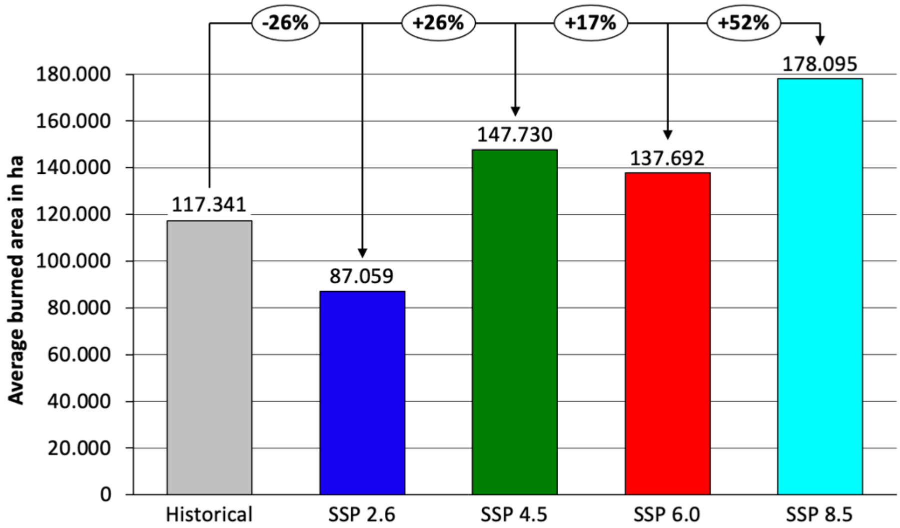


Figure 12: Average burned area per year.

showed in figure 12. This suggests that strong adherence to the Paris Agreement could effectively reduce the severity of wildfires in Portugal. This scenario assumes robust climate action leading to lower emissions and a consequent mitigation of factors that exacerbate wildfires, such as extreme temperatures and drought. However, a critical aspect of the analysis was the observation that all other SSP scenarios that were not consistent with the Paris Agreement targets indicated an increase in wildfire severity. This was particularly evident in the high-emissions SSP 8.5 scenario, which predicted a significant increase in area burned (+52%). However, it is noteworthy that even the medium and high emission scenarios such as SSP 4.5 (+26%) and SSP 6.0 (+17%), which represent lower levels of climate change mitigation compared to SSP 2.6, also projected increases in wildfire severity. This trend was clearly illustrated in the data, which showed a direct correlation between the level of greenhouse gas emissions and the extent of wildfire damage. The results provide strong support for hypothesis H1. The observed inverse relationship between compliance with the Paris Agreement targets and the severity of wildfires in Portugal highlights the critical role of comprehensive climate policies in mitigating wildfire risks. The analysis shows that only the scenario that closely reflects the Paris Agreement (SSP 2.6) predicts a reduction in wildfire severity, while all other scenarios that deviate from these targets predict an increase in wildfire risks. In conclusion, the study reinforces the importance of the Paris Agreement and similar international efforts to address the negative impacts of climate change, especially wildfires. Meeting the agreement's ambitious targets appears to be crucial to reducing the severity of wildfires in Portugal.

Hypothesis H2a: proposed a significant and direct relationship between increased temperatures due to climate change and the area burned by wildfires in Portugal. This hypothesis assumed that higher temperatures, as a consequence of ongoing climate change, would contribute to larger wildfire extents. To test this hypothesis in detail, a combination of advanced machine learning techniques and Ordinary Least Squares (OLS) regression analysis was used. In the feature importance analysis of the machine learning model, *hot_days_year* emerged as a critical predictor of wildfire extent, ranking prominently after two key precipitation variables. This variable, which represents the number of days with exceptionally high temperatures, was identified as a stronger indicator of wildfire risk than more generalized temperature metrics such as *mean_temperature_month* and *average_max_temperature_month*. The significant importance score assigned to *hot_days_year* underscored its relevance in capturing conditions highly conducive to wildfires, particularly those associated with extreme heat events that can exacerbate vegetation drought and flammability. Complementing these findings, OLS regression analysis provided a more nuanced perspective on the relationship between temperature and wildfire. The model showed a positive correlation for *hot_days_year* with area burned, supporting the hypothesis. However, it also revealed a counter-intuitive negative correlation for *mean_temperature_month*. This finding suggests that the relationship between mean temperature and area burned is not straightforward. It implies that average temperature as a single metric may not capture the full range of thermal conditions that influence wildfire behavior, particularly in the context of sporadic but intense heat events.

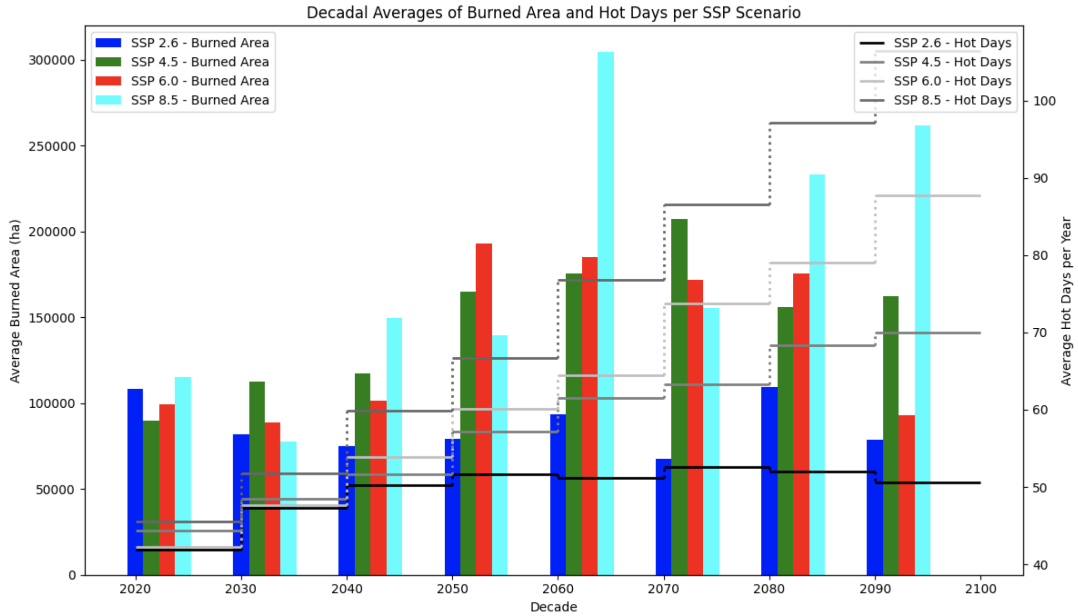


Figure 8: Correlation between hot days per year and area burned per decade.

Figure 12, which shows the relationship between *hot days per year* and wildfire occurrence under different SSP scenarios up to 2100, further illustrates this complexity. Under scenarios such as SSP 8.5, which predict significant temperature increases, both the frequency of hot days and the resulting areas burned increased significantly. However, this relationship showed a non-linear pattern, suggesting the interaction of other influencing factors. These include the variability and intensity of heat waves, the presence and nature of ignition sources, changes in land use, and the interaction of high temperatures with other climatic elements such as humidity and wind patterns. Thus, the comprehensive analysis of hypothesis H2a, integrating machine learning and OLS regression results, confirms a positive but complex relationship between increased temperatures and areas burned by wildfires in Portugal. The results highlight the complexity inherent in modelling wildfire risk, where temperature is a significant factor but is linked to a variety of other environmental and climatic variables. This multifaceted approach highlights the need to incorporate a wide range of climatic indicators, beyond simple temperature metrics, to accurately assess and predict wildfire dynamics in the evolving landscape of climate change.

Hypothesis H2b: assumes that a decrease and change in precipitation patterns, consistent with climate change scenarios, would lead to an increase in the area burned by forest fires. To validate this hypothesis, an analysis was performed using machine learning feature ranking and Ordinary Least Squares (OLS) regression. In the prediction model, summer precipitation (*precipitation_jja*) and annual precipitation (*precipitation_year*) were identified as the most important predictors of wildfire area, outweighing the influence of temperature variables. This prioritization is consistent with empirical observations that lower precipitation in summer, a period typically associated with higher wildfire occurrence, leads to drier conditions which favor fire. OLS regression analysis provided additional quantitative insights. The negative correlation of *precipitation_year* and *precipitation_jja* with area burned suggests that increased precipitation, especially during the summer, tends to mitigate wildfire extent. This finding is consistent with hypothesis H2b and highlights the influence of precipitation patterns on wildfire dynamics. However, the complexity of this relationship is highlighted in Figure 13, which

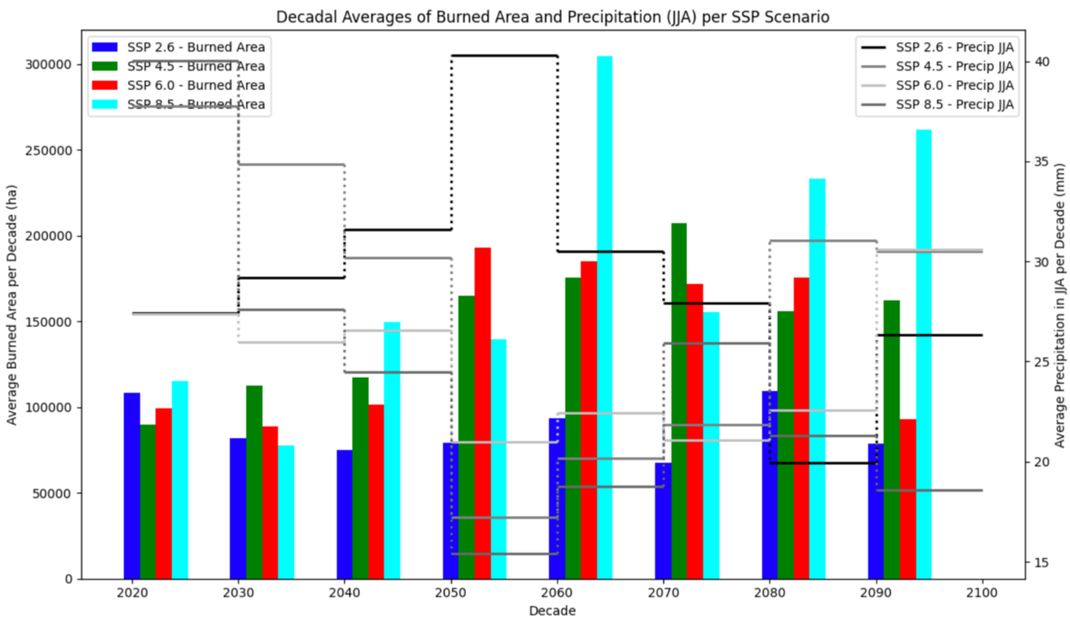


Figure 9: Correlation between precipitation and area burned per decade.

shows significant variation in the relationship between summer precipitation and area burned across different climate scenarios. This lack of a linear trend indicates the nuanced impact of precipitation on wildfire risk. For example, in scenarios with higher total summer precipitation, this precipitation may occur as intense, concentrated downpours, leading to extended dry periods that increase wildfire risk. This observation is essential for understanding that the relationship between precipitation patterns and wildfire areas is far from being straightforward. Therefore, testing hypothesis H2b not only supports the proposed relationship between reduced precipitation and increased wildfire risk, but also highlights the complex nature of this relationship. The complexity lies in the fact that this relationship is influenced by multiple environmental and climatic factors, beyond just the total amount of precipitation. Reduced precipitation, particularly during the critical summer months, is indeed associated with increased wildfire risk, but this relationship is modulated by the distribution and intensity of precipitation events, soil moisture levels, and their interaction with other climatic variables such as temperature. These findings highlight the importance of considering the nuances of precipitation patterns when predicting wildfire risk, and the need for comprehensive and multifaceted wildfire management and climate adaptation strategies in Portugal.

5.6 Conclusion

With a focus on the climate goals of the Paris Agreement, this thesis investigated the impact of climate change on the risk of wildfires in Portugal. Using advanced machine learning models, it analyzed the relationship between variations in temperature and precipitation and the frequency and severity of wildfires. The study revealed a complex, non-linear correlation between higher temperatures and increased wildfire areas, highlighting the significant role of extreme heat in increasing wildfire risks. It also found a positive correlation between reduced precipitation, particularly during the critical summer months, and increased wildfire severity. Crucially, the research showed that meeting the targets of the Paris Agreement could significantly reduce the severity of wildfires in Portugal. The SSP 2.6 scenario, which is consistent with the Agreement's stringent climate targets, predicted a decrease in wildfire severity. In contrast, scenarios with less stringent climate measures, such as SSP 4.5, SSP 6.0 and the high-emission SSP 8.5, predicted increased wildfire risks. The findings of this thesis highlight the importance of climate action, as outlined in the Paris Agreement, in managing and mitigating wildfire risks in a changing climate. This research provides a basis for policy makers and environmental managers to develop effective wildfire management strategies in the context of global climate change.

6 Conclusion and Recommendation for Actions

This thesis, at the intersection of machine learning and wildfire management, brings together a comprehensive analysis of various individual research components, leading to a unified conclusion and set of recommendations for effective wildfire risk management in Portugal. This synthesis of findings and methodologies reveals the intricate dynamics of wildfires, highlighting the complexity and multifaceted nature of effective management strategies.

Central to the thesis is the comparison between traditional wildfire prediction methods and the advanced machine learning models developed in this study. Traditional methods often fall short due to their simplistic assumptions and generalizations. In contrast, the machine learning models used here offer a significant improvement in accuracy. They take into account a wide range of factors, including geographical details, forest composition and human factors such as population density. This approach goes beyond traditional cause-and-effect models to provide a more detailed and accurate picture of wildfire risk. These models effectively integrate and interpret policy data, demonstrating the potential of machine learning to improve risk assessment and management tools in wildfire studies.

Analysis of the National Plan for Integrated Wildland Fire Management reveals a complex picture. While increased investment in fire management and the establishment of Special Protection Areas (SPAs) have proven beneficial in some regions, the general effectiveness varies considerably between areas. This inconsistency highlights the need for more region-specific and nuanced policy approaches. The study also highlights the importance of transparency in policy implementation. Clear information on spending and specific measures in national plans would allow a deeper understanding of what works best in managing wildfires.

Further a link between fire management expenditure and the severity of wildfires was found, suggesting that consistent and possibly increased investment in fire management is essential.

The paper then examines the relationship between environmental, demographic and socio-economic factors in shaping wildfire risk. It shows that while environmental conditions such as weather and topography are critical, demographics, particularly in rural areas, play an important but often underestimated role. The current models suggest other influential factors not yet captured, indicating the complexity of wildfire dynamics. Future research should be expanded to include these additional elements, aiming for a more comprehensive understanding of the drivers of wildfire risk. This could include examining specific land management practices, the broader socio-economic context and policy frameworks.

A significant part of the paper focuses on the impacts of climate change on wildfire risk, highlighting the exacerbating effects of extreme heat and reduced precipitation during the fire season. The study examines different climate scenarios, in line with the goals of the Paris Agreement and finds that strong climate action could significantly reduce the severity of wildfires. This finding underscores the critical role of global climate action in managing wildfire risks and provides a strong foundation for policymakers to develop effective strategies in response to a changing climate. Drawing on these various components, the paper argues for an integrated approach to wildfire risk management that combines advanced predictive models with in-depth policy analysis and a nuanced understanding of environmental, demographic and socio-economic factors. Such a comprehensive strategy is essential for developing effective wildfire management policies that not only reduce risk, but also support community resilience and sustainability, especially considering the challenges posed by climate change and demographic shifts.

The research culminates in several key recommendations for policymakers and practitioners. These include the adoption of advanced machine learning models for more accurate risk assessment, the development of tailored and nuanced wildfire management policies and the importance of consistent investment and transparency in fire management practices. The thesis also encourages the inclusion of additional socio-economic and environmental factors in future research and emphasizes the alignment of wildfire management strategies with global climate goals. In conclusion, this thesis represents a significant step forward in understanding and predicting wildfire dynamics. It calls for a comprehensive approach to wildfire mitigation, combining technological innovation in machine learning with a deep understanding of policy implications and environmental conditions. By recognizing the complexity of these interactions and the need for an integrated approach, the study opens avenues for the development of more effective strategies that address wildfire risks while supporting the well-being and sustainability of rural communities. The findings and recommendations of this thesis provide a robust framework for future research and action in wildfire management, potentially influencing policy and practice in Portugal and regions facing similar wildfire challenges.

7 Limitations

In our broad exploration of wildfire management and prediction, it's essential to consider the range of inherent limitations that affect various aspects of this study, which extend beyond predictive modelling to the broader scope of the research.

A primary limitation is the scale and diversity of the data used. While the thesis integrates climatic, socio-economic, policy and geospatial data, this integration is not complete. Crucial elements such as local economic conditions, cultural practices, and detailed demographic shifts, which could significantly influence wildfire dynamics, have not been fully explored. These overlooked factors could provide important insights into the human dimension of wildfire risk and management, potentially modifying predictive results and policy recommendations.

In addition, the geographical scope of the study, while extensive, does not cover all the diverse ecosystems and cultural landscapes affected by wildfire. Different regions have unique patterns of climate, vegetation, human activity and land use that significantly influence wildfire behavior. A more localized approach might reveal nuances that a broader analysis might miss, thereby affecting the generality of the results. A notable limitation in examining the effectiveness of national wildfire management policies is the lack of detailed data on policy implementation, expenditure, and specific measures. This gap limits the ability to conduct a thorough evaluation of policy impact and effectiveness. Furthermore, the temporal scope of policy analysis limits the assessment of long-term effects, especially in the evolving context of environmental conditions and climate change. Policies and their impacts often unfold over decades, a timeframe not fully captured in this study. The study's reliance on current technological and methodological approaches introduces additional limitations. As technology advances, new data collection methods, more sophisticated algorithms and advanced modelling techniques are likely to emerge, providing more accurate insights into wildfire dynamics. Current methodologies are constrained by the limitations of current technology and understanding of complex natural systems.

A key limitation of predictive modelling is the use of historical data to train and test models. While this approach is fundamental to predictive modelling, it is inherently limited when predicting future events. The effectiveness of models validated against historical data in accurately predicting future conditions is uncertain, particularly in the context of rapidly changing climate patterns. Consequently, there is an inherent challenge in assessing whether these models can accurately reflect future conditions, as they cannot be validated against future data yet to be observed.

The interdisciplinary nature of the study, integrating climatology, sociology, policy analysis and machine learning, presents its own set of challenges. Each discipline brings different methodologies, assumptions, and data types, which can be difficult to reconcile in a coherent analysis. Integrating these different perspectives and insights can lead to oversimplifications or misinterpretations, especially when drawing broad conclusions.

Furthermore, a focus on empirical and quantitative analysis may overlook important ethical and social considerations. The impact of wildfires and management policies on local communities requires understanding beyond data and models. Addressing these dimensions may require a qualitative approach that engages directly with affected communities to understand their perspectives and experiences. Forest fires are very complex and individual in their origin, as almost all of them can be traced back to human activity, this is difficult to reflect in data.

To address these limitations, future research should aim for a more holistic approach. This includes broadening the range of socio-economic factors considered, incorporating more localized and culturally specific data, and enhancing policy analysis with more detailed and long-term data. Embracing new technologies and methodologies will also be key to refining forecasting models and analytical frameworks. In addition, integrating qualitative research and engaging with diverse stakeholders can provide deeper insights into the social and ethical dimensions of bushfire management.

In summary, while this work represents a significant advance in understanding and predicting wildfire dynamics, it is critical to acknowledge its limitations. These limitations highlight the need for continued, comprehensive research in this area, bringing together different disciplines and methodologies to develop a more nuanced and effective approach to wildfire management and policy.

8 References

Ager, Alan A., Nicole M. Vaillant, and Mark A. Finney. 2010. "A Comparison of Landscape Fuel Treatment Strategies to Mitigate Wildland Fire Risk in the Urban Interface and Preserve Old Forest Structure." *Forest Ecology and Management* 259 (8): 1556–70. https://doi.org/10.1016/j.foreco.2010.01.032.

AGIF. 2020. "National Plan for Integrated Wildland Fire Management." In . https://www.agif.pt/en/action/national-plan-for-integrated-wildland-fire-management.

AGIF. 2023. "2018-2022 Policy Review." In. https://www.agif.pt/en/2018-2022-review.

Brotons, Lluís, Núria Aquilué, Miquel de Cáceres, Marie-Josée Fortin, and Andrew Fall. 2013. "How Fire History, Fire Suppression Practices and Climate Change Affect Wildfire Regimes in Mediterranean Landscapes." *PLOS ONE* 8 (5): e62392. https://doi.org/10.1371/journal.pone.0062392.

Carmo, Miguel, Francisco Moreira, Pedro Casimiro, and Pedro Vaz. 2011. "Land Use and Topography Influences on Wildfire Occurrence in Northern Portugal." *Landscape and Urban Planning* 100 (1): 169–76. https://doi.org/10.1016/j.landurbplan.2010.11.017.

Catry, Filipe X., Francisco Moreira, Enrique Cardillo, and Juli G. Pausas. 2012. "Post-Fire Management of Cork Oak Forests." In *Post-Fire Management and Restoration of Southern European Forests*, edited by Francisco Moreira, Margarita Arianoutsou, Piermaria Corona, and Jorge De las Heras, 195–222. Managing Forest Ecosystems. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2208-8_9.

Chan, Jireh Yi-Le, Steven Mun Hong Leow, Khean Thye Bea, Wai Khuen Cheng, Seuk Wai Phoong, Zeng-Wei Hong, and Yen-Lin Chen. 2022. "Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review." *Mathematics* 10 (8): 1283. https://doi.org/10.3390/math10081283.

Chicco, Davide, Matthijs J. Warrens, and Giuseppe Jurman. 2021. "The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation." *PeerJ Computer Science* 7 (July): 1–24. https://doi.org/10.7717/PEERJ-CS.623/SUPP-1.

Chung, Chang-Jo, and Andrea G. Fabbri. 2008. "Predicting Landslides for Risk Analysis — Spatial Models Tested by a Cross-Validation Technique." *Geomorphology* 94 (3–4): 438–52. https://doi.org/10.1016/j.geomorph.2006.12.036.

Collins, Ross D., Richard de Neufville, João Claro, Tiago Oliveira, and Abílio P. Pacheco. 2013. "Forest Fire Management to Avoid Unintended Consequences: A Case Study of Portugal Using System Dynamics." *Journal of Environmental Management* 130 (November): 1–9. https://doi.org/10.1016/j.jenvman.2013.08.033.

Cruz, Miguel G., Martin E. Alexander, Andrew L. Sullivan, James S. Gould, and Musa Kilinc. 2018. "Assessing Improvements in Models Used to Operationally Predict Wildland Fire Rate of Spread." *Environmental Modelling & Software* 105 (July): 54–63. https://doi.org/10.1016/j.envsoft.2018.03.027.

Delbeke, Jos, Artur Runge-Metzger, Yvon Slingenberg, and Jake Werksman. 2019. "The Paris Agreement." *Towards a Climate-Neutral Europe: Curbing the Trend*, October, 24–45. https://doi.org/10.4324/9789276082569-2/PARIS-AGREEMENT-JOS-DELBEKE-ARTUR-RUNGE-METZGER-YVON-SLINGENBERG-JAKE-WERKSMAN.

Dieterle, Gerhard. 2009. "Sustaining the World's Forests: Managing Competing Demands for a Vital Resource – The Role of the World Bank." In *Sustainable Forest Management in a Changing World: A European Perspective*, edited by Peter Spathelf, 19:9–32. Managing Forest Ecosystems. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-3301-7_2.

EFFIS. 2023. "Wildfires in the EU: 2022 Was the Second-Worst Year, a Warning from a Changing Climate." European Forest Fire Information System. https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/wildfires-eu-2022-was-second-worst-year-warning-changing-climate-2023-11-22_en.

F. Chung, Chang-Jo, A. G. Fabbri, and C.J. van Westen. 1995. “Multivariate Regression Analysis for Landslide Hazard Zonation.” In *Geographical Information Systems in Assessing Natural Hazards Selected Contributions from an International Workshop Held in Perugia on September 20-22, 1993. (Advances in Natural and Technological Hazards Research ; 5)*, edited by A. Carrara and F. Guzzetti, 107–33. Kluwer Academic.

Fernandes, Paulo M., Carlos Loureiro, Nuno Guiomar, Gianni B. Pezzatti, Filipa T. Manso, and Luís Lopes. 2014. “The Dynamics and Drivers of Fuel and Fire in the Portuguese Public Forest.” *Journal of Environmental Management* 146 (December): 373–82. https://doi.org/10.1016/j.jenvman.2014.07.049.

Gao, Chao, Honglei Lin, and Haiqing Hu. 2023. “Forest-Fire-Risk Prediction Based on Random Forest and Backpropagation Neural Network of Heihe Area in Heilongjiang Province, China.” *Forests 2023, Vol. 14, Page 170* 14 (2): 170. https://doi.org/10.3390/F14020170.

García-Ruiz, José M., Ignacio I. López-Moreno, Sergio M. Vicente-Serrano, Teodoro Lasanta-Martínez, and Santiago Beguería. 2011. “Mediterranean Water Resources in a Global Change Scenario.” *Earth-Science Reviews* 105 (3–4): 121–39. https://doi.org/10.1016/J.EARSCIREV.2011.01.006.

Gillies, Sean and others. 2013. “Rasterio: Geospatial Raster I/O for Python Programmers.” Mapbox. https://github.com/mapbox/rasterio.

GmbH (https://www.klokantech.com/), Klokan Technologies. n.d. “ETRS89 / Portugal TM06 - EPSG:3763 — Epsg.Io.” https://epsg.io/3763.

Gómez-González, Susana, Fernando Ojeda, and Paulo M. Fernandes. 2018. “Portugal and Chile: Longing for Sustainable Forestry While Rising from the Ashes.” *Environmental Science & Policy* 81 (March): 104–7. https://doi.org/10.1016/j.envsci.2017.11.006.

González, José Ramón, Osmo Kolehmainen, and Timo Pukkala. 2007. “Using Expert Knowledge to Model Forest Stand Vulnerability to Fire.” *Computers and Electronics in Agriculture* 55 (2): 107–14. https://doi.org/10.1016/j.compag.2006.12.005.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. “Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone.” *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2017.06.031.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, et al. 2020. “The ERA5 Global Reanalysis.” *Quarterly Journal of the Royal Meteorological Society* 146 (730): 1999–2049. https://doi.org/10.1002/QJ.3803.

Hunter, Molly E., Melanie M. Colavito, and Vita Wright. 2020. “The Use of Science in Wildland Fire Management: A Review of Barriers and Facilitators.” *Current Forestry Reports* 6 (4): 354–67. https://doi.org/10.1007/s40725-020-00127-2.

IEA. n.d. “Hungary Climate Resilience Policy Indicator – Analysis - IEA.”

IPCC. 2019. “Global Warming of 1.5 ºC.” *One Earth* 1 (3): 374–81.

Jain, Piyush, Sean C.P. Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D. Flannigan. 2020. “A Review of Machine Learning Applications in Wildfire Science and Management.” *Environmental Reviews* 28 (4): 478–505. https://doi.org/10.1139/er-2020-0019.

Jaiswal, Rajeev Kumar, Saumitra Mukherjee, Kumaran D. Raju, and Rajesh Saxena. 2002. “Forest Fire Risk Zone Mapping from Satellite Imagery and GIS.” *International Journal of Applied Earth Observation and Geoinformation* 4 (1): 1–10. https://doi.org/10.1016/S0303-2434(02)00006-5.

Karssenberg, Derek, Oliver Schmitz, Peter Salamon, Kor De Jong, and Marc F.P. Bierkens. 2010. “A Software Framework for Construction of Process-Based Stochastic Spatio-

Temporal Models and Data Assimilation." *Environmental Modelling & Software* 25 (4): 489–502. https://doi.org/10.1016/j.envsoft.2009.10.004.

Latifah, Arnida L., Ayu Shabrina, Intan N. Wahyuni, and Rifki Sadikin. 2019. "Evaluation of Random Forest Model for Forest Fire Prediction Based on Climatology over Borneo." *2019 International Conference on Computer, Control, Informatics and Its Applications: Emerging Trends in Big Data and Artificial Intelligence, IC3INA 2019*, October, 4–8. https://doi.org/10.1109/IC3INA48034.2019.8949588.

Levine, Richard A., and Daniel S Wilks. 2000. "Statistical Methods in the Atmospheric Sciences." *Journal of the American Statistical Association* 95 (449): 344. https://doi.org/10.2307/2669579.

Martínez, Jesús, Cristina Vega-Garcia, and Emilio Chuvieco. 2009. "Human-Caused Wildfire Risk Rating for Prevention Planning in Spain." *Journal of Environmental Management* 90 (2): 1241–52. https://doi.org/10.1016/j.jenvman.2008.07.005.

Mateus, Paulo, and Paulo M. Fernandes. 2014. "Forest Fires in Portugal: Dynamics, Causes and Policies." In *Forest Context and Policies in Portugal: Present and Future Challenges*, edited by Fernando Reboredo, 97–115. World Forests. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-08455-8_4.

"Metodologia Para a Produção Da Carta de Perigosidade de Incêndio Rural de Cariz Estrutural — Icnf.Pt." 2020. https://www.icnf.pt/api/file/doc/96bb210ebf341cda.

Moreira, Francisco, Davide Ascoli, Hugh Safford, Mark A. Adams, José M. Moreno, José M. C. Pereira, Filipe X. Catry, et al. 2020. "Wildfire Management in Mediterranean-Type Regions: Paradigm Change Needed." *Environmental Research Letters* 15 (1): 011001. https://doi.org/10.1088/1748-9326/ab541e.

Moreira, Francisco, Olga Viedma, Margarita Arianoutsou, Thomas Curt, Nikos Koutsias, Eric Rigolot, Anna Barbati, et al. 2011. "Landscape – Wildfire Interactions in Southern Europe: Implications for Landscape Management." *Journal of Environmental Management* 92 (10): 2389–2402. https://doi.org/10.1016/j.jenvman.2011.06.028.

Nunes, Adélia N. 2012. "Regional Variability and Driving Forces behind Forest Fires in Portugal an Overview of the Last Three Decades (1980–2009)." *Applied Geography* 34 (May): 576–86. https://doi.org/10.1016/j.apgeog.2012.03.002.

Oliveira, Sandra, Ana Gonçalves, and José Luís Zêzere. 2021. "Reassessing Wildfire Susceptibility and Hazard for Mainland Portugal." *Science of The Total Environment* 762 (March): 143121. https://doi.org/10.1016/j.scitotenv.2020.143121.

Oliveira, Sandra, Friderike Oehler, Jesús San-Miguel-Ayanz, Andrea Camia, and José M.C. Pereira. 2012. "Modeling Spatial Patterns of Fire Occurrence in Mediterranean Europe Using Multiple Regression and Random Forest." *Forest Ecology and Management* 275 (July): 117–29. https://doi.org/10.1016/j.foreco.2012.03.003.

Oliveira, Tiago M., Ana M. G. Barros, Alan A. Ager, Paulo M. Fernandes, Tiago M. Oliveira, Ana M. G. Barros, Alan A. Ager, and Paulo M. Fernandes. 2016. "Assessing the Effect of a Fuel Break Network to Reduce Burnt Area and Wildfire Risk Transmission." *International Journal of Wildland Fire* 25 (6): 619–32. https://doi.org/10.1071/WF15146.

Pacheco, Renata Martins, and João Claro. 2021. "Prescribed Burning as a Cost-Effective Way to Address Climate Change and Forest Management in Mediterranean Countries." *Annals of Forest Science* 78 (4): 1–11. https://doi.org/10.1007/s13595-021-01115-7.

Pandey, Pooja, Gabriela Huidobro, Luis Filipe Lopes, Anne Ganteaume, Davide Ascoli, Conceição Colaco, Gavriil Xanthopoulos, et al. 2023. "A Global Outlook on Increasing Wildfire Risk: Current Policy Situation and Future Pathways." *Trees, Forests and People* 14 (December): 100431. https://doi.org/10.1016/j.tfp.2023.100431.

Parente, Joana, and Mário G. Pereira. 2016. "Structural Fire Risk: The Case of Portugal." *Science of The Total Environment* 573 (December): 883–93. https://doi.org/10.1016/j.scitotenv.2016.08.164.

Paton, Douglas, and F. Tedim. 2014. "Enhancing Forest Fires Preparedness in Portugal: Integrating Community Engagement and Risk Management," January. https://figshare.utas.edu.au/articles/journal_contribution/Enhancing_Forest_Fires_Preparedness_in_Portugal_Integrating_Community_Engagement_and_Risk_Management/22908065/1.

Paveglio, Travis B., Jesse Abrams, and Autumn Ellison. 2016. "Developing Fire Adapted Communities: The Importance of Interactions Among Elements of Local Context." *Society & Natural Resources* 29 (10): 1246–61. https://doi.org/10.1080/08941920.2015.1132351.

Pereira, Mario Gonzalez, Jose Aranha, and Malik Amraoui. 2014. "Land Cover Fire Proneness in Europe." *Forest Systems* 23 (3): 598–610. https://doi.org/10.5424/fs/2014233-06115.

Pinto, Lígia Costa, Sara Sousa, and Marieta Valente. 2022. "Forest Bioenergy as a Land and Wildfire Management Tool: Economic Valuation under Different Informational Contexts." *Energy Policy* 161 (February): 112765. https://doi.org/10.1016/j.enpol.2021.112765.

Platt, John and others. 1999. "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods." *Advances in Large Margin Classifiers* 10 (3): 61–74.

"Portugal - Mean Projections Expert | Climate Change Knowledge Portal." n.d. Accessed November 9, 2023. https://climateknowledgeportal.worldbank.org/country/portugal/climate-data-projections.

Portugal Government. 2021. "National Climate Change Adaptation Planning and Strategies Legal Instrument: Regulation on the Governance of the Energy Union and Climate Action Obligation: National Climate Change Adaptation Planning and Strategies-GovReg."

QGIS Development Team. 2023. "QGIS Geographic Information System." QGIS Association. https://www.qgis.org.

Ribeiro, Luís M., Domingos X. Viegas, Miguel Almeida, Tara K. McGee, Mário G. Pereira, Joana Parente, Gavriil Xanthopoulos, Vittorio Leone, Giuseppe Mariano Delogu, and Holy Hardin. 2020. "2 - Extreme Wildfires and Disasters around the World: Lessons to Be Learned." In *Extreme Wildfire Events and Disasters*, edited by Fantina Tedim, Vittorio Leone, and Tara K. McGee, 31–51. Elsevier. https://doi.org/10.1016/B978-0-12-815721-3.00002-3.

Sayad, Younes Oulad, Hajar Mousannif, and Hassan Al Moatassime. 2019. "Predictive Modeling of Wildfires: A New Dataset and Machine Learning Approach." *Fire Safety Journal* 104 (March): 130–46. https://doi.org/10.1016/j.firesaf.2019.01.006.

Schillaci, Calogero, Andreas Braun, and Jan Kropáček. 2015. "Terrain Analysis and Landform Recognition." https://doi.org/10.13140/RG.2.1.3895.2802.

Schultz, Courtney A., Matthew P. Thompson, and Sarah M. McCaffrey. 2019. "Forest Service Fire Management and the Elusiveness of Change." *Fire Ecology* 15 (1): 13, s42408-019-0028–x. https://doi.org/10.1186/s42408-019-0028-x.

Scott, Andrew C, David MJS Bowman, William J Bond, Stephen J Pyne, and Martin E Alexander. 2013. *Fire on Earth: An Introduction*. John Wiley & Sons.

Sebastián-López, Ana, Raymond Salvador-Civil, Julián Gonzalo-Jiménez, and Jesús SanMiguel-Ayanz. 2008. "Integration of Socio-Economic and Environmental Variables for Modelling Long-Term Fire Danger in Southern Europe." *European Journal of Forest Research* 127 (2): 149–63. https://doi.org/10.1007/s10342-007-0191-5.

Simon, Benjamin, Christian Crowley, and Fabiano Franco. 2022. "The Costs and Costs Avoided From Wildfire Fire Management—A Conceptual Framework for a Value of Information Analysis." *Frontiers in Environmental Science* 10. https://www.frontiersin.org/articles/10.3389/fenvs.2022.804958.

Skulska, I., I. Duarte, F. C. Rego, and C. Montiel-Molina. 2020. “Relationships Between Wildfires, Management Modalities of Community Areas, and Ownership Types in Pine Forests of Mainland Portugal.” *Small-Scale Forestry* 19 (2): 231–51. https://doi.org/10.1007/s11842-020-09445-6.

“The Natura 2000 Protected Areas Network — Eea.Europa.Eu.” n.d. https://www.eea.europa.eu/themes/biodiversity/natura-2000#: :text=Natura%202000%20is%20a%20network,on%20land%20and%20at%20sea.

Tien Bui, Dieu, Kim-Thoa Le, Van Nguyen, Hoang Le, and Inge Revhaug. 2016. “Tropical Forest Fire Susceptibility Mapping at the Cat Ba National Park Area, Hai Phong City, Vietnam, Using GIS-Based Kernel Logistic Regression.” *Remote Sensing* 8 (4): 347. https://doi.org/10.3390/rs8040347.

Turco, Marco, Sonia Jerez, Sofia Augusto, Patricia Tarín-Carrasco, Nuno Ratola, Pedro Jiménez-Guerrero, and Ricardo M. Trigo. 2019. “Climate Drivers of the 2017 Devastating Fires in Portugal.” *Scientific Reports* 9 (1): 13886. https://doi.org/10.1038/s41598-019-50281-2.

Vapnik, Vladimir. 1999. *The Nature of Statistical Learning Theory*. Springer science & business media.

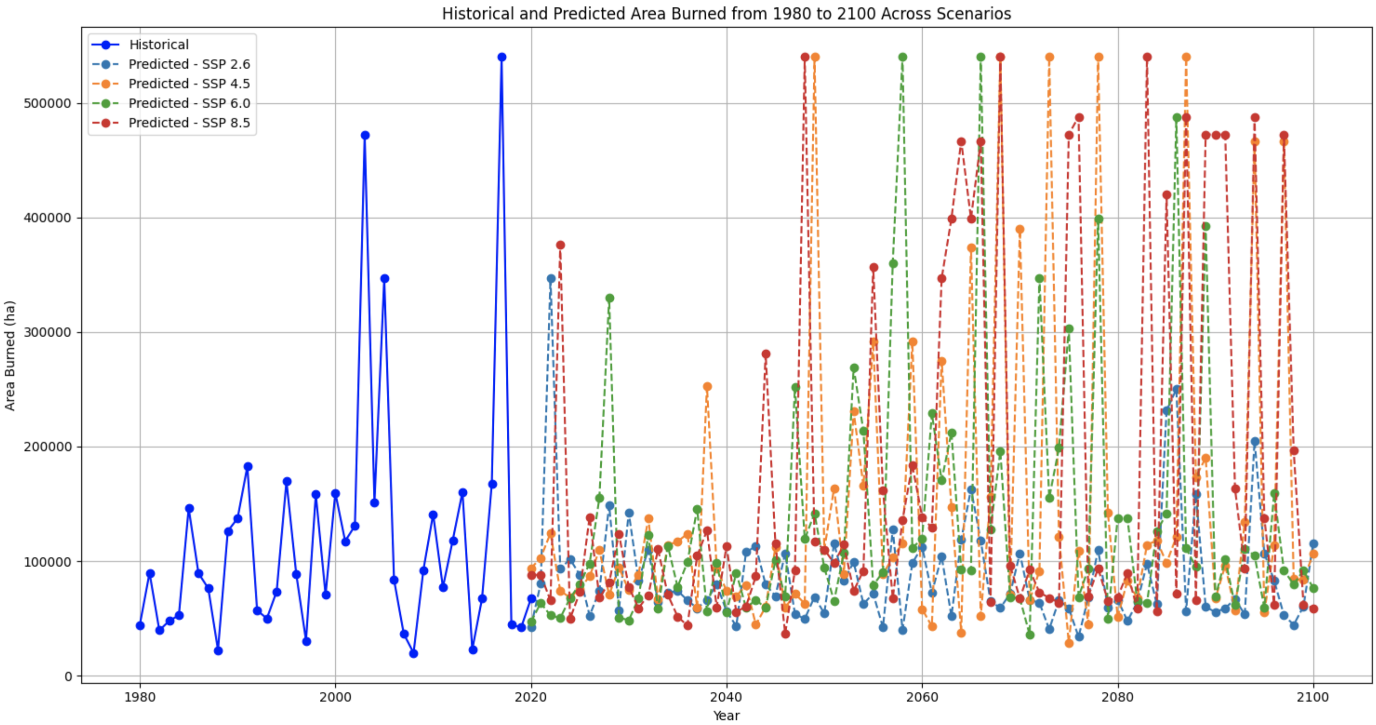
Westerling, A. L., H. G. Hidalgo, D. R. Cayan, and T. W. Swetnam. 2006. “Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity.” *Science* 313 (5789): 940–43. https://doi.org/10.1126/SCIENCE.1128834/SUPPL_FILE/WESTERLING.SOM.PDF.

“WorldPop :: Geospatial Covariate Data Layers.” n.d. Accessed November 8, 2023. https://hub.worldpop.org/geodata/summary?id=17723.

9 Appendix

9.1 Appendix A: Figures

A1: Results of the Predictive Modelling



9.2 Appendix B: Tables

B1: Variance Inflation Factor (VIF) Analysis

Feature	VIF
const	55.960730
mean_temperature_month	4.533996
precipitation_month	1.286178
hot_days_year	1.116649
precipitation_year	1.876623
precipitation_jja	1.822617
Average_Max_Temperature_Month	4.133120

B2: Regression Coefficients and Statistics

Variable	Coefficient	Std. Error	t-value	P-value	95% Conf. Interval
const	122329.2484	953.778	128.258	0.000	[120000, 124000]
mean_temperature_month	-1052.1528	63.620	-16.538	0.000	[-1176.846, -927.460]
precipitation_month	50.9447	3.906	13.044	0.000	[43.290, 58.600]
hot_days_year	691.9655	4.748	145.743	0.000	[682.660, 701.271]
precipitation_year	-5.1162	0.226	-22.590	0.000	[-5.560, -4.672]
precipitation_jja	-94.6154	2.119	-44.641	0.000	[-98.770, -90.461]
Average_Max_Temperature_Month	1151.3798	64.332	17.897	0.000	[1025.291, 1277.469]

B3: Model Summary Statistics

Statistic	Value
R-squared	0.035
Adjusted R-squared	0.035
F-statistic	4833.19
Prob (F-statistic)	0.0000
Log-Likelihood	-1.0501e+07
AIC	2.100e+07
BIC	2.100e+07
No. Observations	803692
Df Residuals	803685
Df Model	6