1 2 9 0

UNIVERSIDADE Ð
COIMBRA

Nuno Pires

Intelligent system for localising and monitoring forest fires

Dissertation in the context of the Master in Informatics Engineering, specialization in Information Systems, advised by Professor Alberto Cardoso and Professor Jacinto Estima and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

January 2024

1 2 9 0

DEPARTAMENTO DE
ENGENHARIA INFORMÁTICA
FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE Đ
COIMBRA

Nuno Pires

Intelligent system for localising and monitoring forest fires

Dissertation in the context of the Master in Informatics Engineering, specialization in Information Systems, advised by Professor Alberto Cardoso and Professor Jacinto Estima and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

January 2024

1290

DEPARTAMENTO DE
ENGENHARIA INFORMÁTICA
FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE Ð
COIMBRA

Nuno Pires

Sistema inteligente para localização e monitorização de incêndios florestais

Dissertação no âmbito do Mestrado em Engenharia Informática, especialização em Sistemas de Informação, orientada pelo Professor Alberto Cardoso e Professor Jacinto Estima e apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Janeiro 2024

Abstract

Fire can have disastrous consequences. Decision-support systems play a central role in dealing with forest fires. Its early warning capacity and real-world impact help to protect forests, species, and communities from wildfire.

The presented work proposes a system for forecasting and monitoring forest fires using multiple data sources. Data fusion, aggregation, and enhancement techniques are also mentioned.

The main purpose of the system is to provide important information for emergency decision-making, such as the geolocation, severity, and temporal evolution of a wildfire. It will employ statistical and machine learning methodologies to predict and determine fire occurrence, susceptibility, and risk.

Finally, the system, with the help of data visualisation tools, will show findings and insights.

The document also presents current approaches and obstacles to forest fire prediction, as well as the suggested methodology and analysis of risk.

Keywords

Decision support system, Fire management, Fire forecasting, Machine learning, Spatial and temporal prediction

Resumo

Os incêndios podem ter consequências desastrosas. Os sistemas de apoio à decisão desempenham um papel central na luta contra os incêndios florestais. As suas capacidades de alerta e o seu impacto no mundo real ajudam a proteger as florestas, as espécies e as comunidades.

O trabalho apresentado propõe um sistema de previsão e monitorização de incêndios florestais que utiliza fontes diversas de dados. Onde são utilizadas técnicas de fusão, agregação e melhoramento de dados.

O principal objetivo do sistema é fornecer informações importantes para a tomada de decisões de emergência, tais como a geolocalização, a gravidade e a evolução temporal de um incêndio florestal. O sistema empregará metodologias estatísticas e de aprendizagem automática para prever e determinar a ocorrência, a suscetibilidade e o risco de incêndio.

Finalmente, com a ajuda de ferramentas de visualização de dados, o sistema será capaz de apresentar informações e resultados.

No documento também são analisadas as abordagens actuais e os obstáculos à previsão de incêndios florestais, bem como a metodologia sugerida e a análise de risco.

Palavras-Chave

Sistema de apoio à decisão, Gestão de incêndios, Previsão de incêndios, Aprendizagem automática, Previsão espacial e temporal

Contents

Acronyms

ANN Artificial Neural Networks.

AUC Area Under the Roc Curve.

BRT Boosted Regression Tree.

CNN Convolutional neural networks.

DC Drought Code.

DMC Duff Moisture Code.

ECMWF European Centre for Medium-Range Weather Forecasts.

FbSP Fuzzy Logic Supervised Approach.

FCT Foundation for Science and Technology.

FFDWI Forest Fire Danger Weather Index.

FFMC Fine Fuel Moisture Content.

FWI Fire Weather Index.

GLM Generalised Linear Model.

ISI Initial Spread Index.

KBDI Keetch-Byram Drought Index.

LVQ Learning Vector Quantization.

MDA Mixed Discriminant Analysis.

MODIS Moderate Resolution Imaging Spectroradiometer.

NDVI Normalized difference vegetation index.

OLI Operational Land Imager.

RWSB Rechargeable Wireless Sensor Network.

SVM Support Vector Machines.

TSS True Skill Statistic.

USDA United States Department of Agriculture.

XAI Explainable artificial intelligence.

List of Figures

List of Tables

Chapter 1

Introduction

Por ler:

- FWI features - `https://www.for.gov.bc.ca/ftp/!Project/FireBehaviour/Canadian%20Fire%20Behaviour%20for%20AU/FWI%20Tables.pdf`

In 2017, a forest fire broke out in Pedrógão Grande, located in the central Region of Portugal. This disaster claimed the lives of 66 citizens and burned an area of 47 thousand hectares. The wildfire was significantly shaped by a thunderstorm and an intense heatwave, with temperatures exceeding 40ºC and a relative humidity bellow 20%. Within an hour, the fire spread from 2940 to 6740 burned hectares. The majority of fatalities were caused by flame exposure and smoke inhalation as residents attempted to flee the encroaching fires that threatened their houses (Viegas, 2018).

What if Portugal had better means of predicting forest fires? Could a disaster like this be avoided? First and foremost, it is necessary to comprehend what constitutes a forest fire and the threat it poses.

A forest fire is a wildfire that burns uncontrollably in a forest, meadow, scrubland, or cultivated land (of Encyclopaedia Britannica, 2023). It can be caused by weather and erupt spontaneously as a result of lightning or the heat of the sun. Forest fires are becoming more frequent as a result of climate change and human activity, wreaking havoc on the environment, economy, and human health.

Given the complexities of fire and the challenges it poses in modern times, the goal is to develop a system that processes and aggregates data related to wildfires from multiple sources to classify forest fire severity.

The risk of fire in a wilderness environment fluctuates with the weather and is influenced by topography. Factors such as drought, heat, and wind all contribute to the ferocity of a fire. Human activities like unsupervised campfires, abandoned cigarettes, and other human influences also exacerbate forest fire ignition (Arif et al., 2021; Gaikwad et al., 2022; Ivanchuk, 2023; of Encyclopaedia Britannica, 2023). Forest fires are harmful to people and the environment because they pollute the air, destroy infrastructure, compel people to evacuate, and endanger biodiversity (Economic, 2023).

Forest fires are a common occurrence in many areas throughout the world. Portugal is a geographical location that stands out not only in terms of the number of occurrences, but also in terms of the extent of the burned lands. (Silva and Santos, 2012). According to the "Commission report on forest fires", in Europe more than 340,000 hectares were burned in 2020. Romania being the most affected country, followed by Portugal in second place (Europeia, 2021).

The fires burning today in many parts of the world are bigger, more intense, and last longer than they used to. In Portugal, only after 1990 were flames larger than 10,000 hectares documented, which shows that in recent years, the burned area has been steadily increasing(Viegas, 2018).

1.1 Context

Decision support systems are critical in a variety of emergency scenarios, notably in early wildfire forecasts, which will lessen the disaster's damage and ease the control of wildfires through civil protection and firefighting efforts (Cardoso and Estima, 2023).

The availability of data derived from volunteer contributions allows the development of models for recognising wildfire incidents. This involves localising and tracking the event over time and space. Anticipating the likelihood of a forest fire breakout before it begins by calculating the link between fire risk and other relevant factors such as meteorological conditions or topography (Cardoso and Estima, 2023).

Providing adequately structured and validated information will aid in a increasing understanding of a wildfire event and will make real-time decision-making easier (Abid, 2021; Cardoso and Estima, 2023).

In addition to being mentioned previously, where the hazards of fire and their consequences were examined. Table 1.1 illustrates why it is important to develop solutions that will understand wildfires and avoid disaster. Table 1.1 shows that Portugal is one of the European nations most afflicted by fire. In the last ten years, Portugal has had an average of 13298 fires per year and 123876 hectares of land burned.

The aim is to develop a set of modules that take on another viewpoint regarding decision-making support systems in forest fire management. Providing spatial, temporal and risk monitoring that can be used in an organised way to ease decision-making. It is therefore necessary to understand how fire ignition factors correlate to each other, how and why a fire ignites.

1.1.1 FireLoc

FireLoc was a project funded by the Foundation for Science and Technology (FCT), which has now come to an end. This thesis is a follow-up to the stud-

Table 1.1: Number of rural fires and corresponding extent of burnt area in mainland Portugal, per year, between 1 January and 15 October 2023 January and 15 October 2023 (da Conservação da Natureza e das Florestas, 2023)

Year	No. of rural fires	Burnt area (ha)			
		Settlements	Forest	Agricultural area	Total
2013	21917	54905	94564	7858	157327
2014	9095	8701	10889	2954	22544
2015	18945	23461	39538	3796	66795
2016	14980	77390	82505	6290	166185
2017	19104	328851	168611	39669	537131
2018	11451	21873	19114	3091	44078
2019	10528	21411	15831	4608	41850
2020	9182	31682	27826	6315	65823
2021	7452	8077	16105	2947	27129
2022	10323	55304	43591	11015	109910
2023	7635	19281	12994	2145	34420
Average 2013-2022	13298	63165	51857	8854	123876

ies that were carried out as part of the project. The modules to be developed will therefore be integrated with the project's previous material.

FireLoc had the goal of creating a system that would enable any volunteer citizen with a smartphone to report a spotted fire. It enabled fire spottery by taking the location of the observation point automatically. It also permitted the capture of an image during the observation (such as a smartphone photo) and provided information that allowed the observed phenomenon to be georeferenced, such as the orientation (which the device automatically determines) (CISUC, 2023; Silva et al., 2020).

1.1.2 Main Issues

The currently ongoing, global change-induced, intensification of the fire regime has escalated from being primarily an ecological problem to also becoming a civil protection issue (Resco de Dios and Nolan, 2021).

Every biome is impacted by wildfires, including grasslands, tundra, savannahs, and forests. Every year, more than 400 million hectares of land worldwide experience fire damage, with savannahs and grasslands accounting for 70% of these incidents (Lang and Moeini-Meybodi, 2023).

Issues regarding forest fires can influence multiple stakeholders and biodiversity. The following are issues under which it is important to carry out forest fire mitigation:

- Wildfires have impacts on the environment, society, and economy (Lang and Moeini-Meybodi, 2023);
- Wildfires are an issue that need an effective strategy for predicting and preventing them (Thakkar et al., 2022);

- Effective fire control needs precise timely information regarding fire occurrence, spread, and environmental effect and how environmental elements impact the risk of forest fires (Arif et al., 2021; Mekala et al., 2023);
- When compared to the pace of wildfire propagation, current surveillance technologies are slow (Thakkar et al., 2022);
- Forest fire control is essential for mitigating the detrimental effects of wildfires on the environment and communities (Mekala et al., 2023);
- False alarms are one of the most serious issues with forest fire protection systems (Thakkar et al., 2022);
- Wildfires wreak havoc on people, the environment, and the economy (Mekala et al., 2023) (Sharma et al., 2020);
- Wildfires cause irreversible environmental and atmospheric harm. One-third of all carbon dioxide residing in the atmosphere comes from wildfires (Alkhatib, 2014);
- Investments and measures to reduce the danger and effects of wildfires have not been enough to address this expanding hazard, despite the evidently negative effects of wildfires (Lang and Moeini-Meybodi, 2023).

1.2 Objectives

Given the extent of wildfire issues and the complex nature of fire, it is important to provide a solution that tackles forest fire management at its core. Therefore, the goal is to provide a set of system modules that will:

- Analyse, evaluate and collect volunteer contributions;
- Build a data pipeline capable of extracting, transforming, and loading data from a variety of sources;
- Categorise the severity of a wildfire event using machine learning and statistical methodologies, as well as to determine the most relevant elements for fire occurrence, susceptibility, and risk;
- Aggregate contributions related to the same wildfire;
- Create new insights due to aggregation analysis;
- Apply data fusion methodologies to wildfire data;
- Determine the best way to visualise the results and deliver findings in a clear, engaging, and intuitive manner that can improve decision-making in emergency situations with data visualisation tools.

1.2.1 Benefits of system implementation

The creation of the different modules will result in the following implementation benefits:

- Increased Accuracy: By utilising machine learning techniques, the system is able to create more accurate predictions based on past data;
- Improved Emergency Response: Aiding decision-making in emergency circumstances. Making them more prone to respond more effectively and efficiently by offering geographical and temporal surveillance of incidents;
- Early notice: The system gives early notice of probable forest fires, making it possible to take preventive actions to lessen the danger of fire;
- Improved Forest Fire Management: The system is meant to process, validate, and aggregate volunteer contributions, categorise the severity of a forest fire event, identify its geolocation, and track its temporal evolution;
- Economical: The suggested approach is less expensive than standard techniques of monitoring and forecasting forest fires, which frequently rely on manual labour due to its automated nature;
- Data-Driven: Because the system is data-driven, it is possible to continuously enhance the prediction model based on fresh data;
- Data Visualisation: Utilising the best approach to visualise the findings, making it simpler to grasp and analyse the data.

1.2.2 Document Outline

This document is organised into a total of five chapters. The first chapter provides a contextualization of wildfires, fire issues, objectives, and implementation benefits.

The second chapter supplies a background description of multiple topics related to the project.

In the third chapter, a state-of-the-art analysis of fire occurrence, risk prediction, and fire susceptibility is performed. The chapter also includes literature in regards to data aggregation, fusion, enhancement, and data visualisation tools.

In the fourth chapter, a solution proposal is defined and the problem is categorised. Finally, the fifth chapter concludes the document.

Chapter 2

Background

The first way of anticipating forest fires may be traced back to native American cultures that utilised smoke signals to relay critical information across large distances. These tribes could tell whether a fire was nearing their settlements by monitoring the colour, density, and direction of the smoke (Frąckiewicz, 2023).

In the present day, there are multiple approaches to study fire and its consequences. Forest fire research can be divided into four main categories (Arif et al., 2021):

- Fire occurrence prediction (spatial and temporal);
- Detection of an ongoing fire incident (Future Burned Area);
- Prediction of Wildfire Spread;
- Fire-caused Burned Area Detection.

This chapter will comprise: understanding forest fire management and its key roles in the thesis. The importance of decision support systems and the importance of volunteer contributions. These last two are all follow-ups from the Fireloc project, mentioned in 1.1.1.

At last, the concept of spatial and temporal prediction will be described. These are key concepts for this thesis, as almost everything related to wildfires revolves around the binary relationship between spatial and temporal prediction.

2.1 Forest Fire Management

Fire management is a type of risk management. It is the process of organising, averting, and combating fires in order to save individuals, assets, and forest resources (Canada, 2021; Jain et al., 2020).

The fire management topics that will be covered in this thesis are fire occurrence, susceptibility, and risk. These three factors can all use statistical and machine learning methods to model fire ignition likelihood and frequency, and map fire risk and susceptibility based on environmental factors.

The aim of modern fire management is to maintain the right quantity of fire on the landscape. This can be achieved by controlling vegetation, which includes controlled burning, controlling human activity (prevention), and suppressing fires. This is a critical aspect because fire is a natural occurrence, and nature has evolved in response to its presence (Society, 2023). The following list comprises some of the advantages of wildfire, and the importance of wildfire occurrence at a managed scale:

- Many ecosystems benefit from periodic fires because they sweep away decaying organic material;
- Some plant and animal populations rely on fire to survive and reproduce;
- Several plants rely on fire to complete their life cycles;
- Fires can assist in the eradication of invasive species that have not adapted to frequent wildfires.

2.2 Importance of decision support systems in emergency situations

In the context of forest fire management, a decision support system is a set of software used to assist decisions, assessments, and actions. Massive data sets are sorted through and analysed, which then outputs detailed information that may be used in decision-making and problem-solving (Sutton et al., 2020).

Decision supports systems help to protect the population and territories from natural emergencies such as wildfires (Nemtinov et al., 2021).

They aid in understanding the spatial distribution of fires and identifying the human and environmental elements that contribute to the occurrence of fires in various areas, delivering crucial information throughout the decision-making process for fire control (Gao et al., 2023).

2.3 Importance of voluntary contributions

Contributing to fire localization is important when there is no official, planned, or organised disaster preparedness (Smith et al., 2016).

Informal volunteering is a vital and valuable resource for emergency and crisis management (Aminizade et al., 2017). Successful implementation of fire control endeavours has been acknowledged to depend on the active participation of the local population (Goldammer et al., 2001).

Volunteering can provide economic and social benefits for society, reducing vulnerability and supporting disaster risk reduction (Aminizade et al., 2017).

Volunteers can help gather and report data on weather conditions, vegetation status, and other factors that can influence the likelihood of forest fires. This data can feed into predictive models to help anticipate where and when fires might occur (Artés et al., 2019).

Therefore, volunteers play a key role in raising awareness about forest fire risk and prevention strategies within their communities. This can lead to more accurate reporting of risk factors and quicker responses to wildfire occurrences (Li et al., 2023).

2.4 Spatial and Temporal prediction

The first measure of fire control is forest fire prediction, which is essentially the prediction of forest fire occurrence, anticipating the forest fire breakout likelihood prior to its first ignition. Modelling the link between fire risk and relevant elements like weather or topography (Abid, 2021).

Prediction entails not only knowing the precise moment of the initiation of a forest fire, but also where it will occur, as a result, it is a hybrid of temporal and spatial prediction (Arif et al., 2021).

Spatial and temporal prediction models are at the heart of the wildfire prediction problem definition. Spatial prediction models use methods to anticipate where a fire could begin. This can be based on historical data about fires or regions where circumstances are favourable for a fire, such as areas with a lot of dry trees (Cilli et al., 2022). For example, the output of a spatial prediction can be a pin pointing to a location on a map.

Temporal prediction models consider when the fire will occur. By using previous meteorological data, it is possible to anticipate, for example, the month or season in which the fire will take place (Dong et al., 2022).

Chapter 3

State of the Art

Forest fire occurrence prediction is a crucial task for lessening the impact of wildfires on the environment, economy, and human health. Various factors, such as weather, topography, vegetation, and human activities, influence the likelihood and frequency of fire ignition and spread. To model and map these factors, different mathematical and machine learning techniques have been proposed and applied in the literature.

This chapter reviews the state of the art methods for forest fire occurrence, susceptibility, and risk prediction, as well as the data sources, tools, and challenges involved in this task.

At last, it is addressed methodologies and techniques for data aggregation, fusion, and enhancement and data visualisation tools.

3.1 Importance of Machine Learning in Forest fire prediction

Machine learning algorithms have the ability to automatically anticipate and identify fire incidents (Abid, 2021). They are able to simulate how fire danger is influenced by variables like topography and weather.

Wildfire occurrence prediction will often take multiple approaches to decision-making by utilising multiple machine learning models in conjunction with mathematical models. This aids in predicting the likelihood and risk of a forest fire spreading before it starts (Abid, 2021). Additionally, machine learning models are also capable of predicting the behaviour of forest fires, or the development of their spread (Abid, 2021).

Taking Convolutional neural networks (CNN) and other machine learning models as an example, these models have been used to conduct spatial prediction of forest fire susceptibility (Zhang et al., 2019).

Predicting fire risks is crucial to minimize monthly fire emissions across large regions and reduce the consequences of fires on human health, air quality, and climate change (Wang et al., 2022). Machine learning models can help local authorities and fire services use their resources more wisely in addition to averting disasters and saving lives in the event of a fire (Surya, 2017).

3.2 Mathematical models in forest fire prediction

Sometimes machine learning models are assembled in conjunction with mathematical models. These models estimate the chance and probability of a forest fire occurrence using functions and algorithms. Some models can rely on meteorological characteristics such as temperature, humidity, wind speed, and rainfall to forecast forest fires. The most famous mathematical model is the Fire Weather Index (FWI).

The canadian FWI and other forest weather systems give numerical indices for predicting and avoiding fires. It is a technique for calculating indices based on temperature, relative humidity, rain, and other variables (Mohammed et al., 2020). This rating incorporates the chance of fire ignition and the pace of fire spread (Alkhatib, 2014). This approach is employed not just in Canada, but also in various European nations, like Portugal (Mohammed et al., 2020).

Another example is the Angstrom index, which is a basic fire index that calculates the likelihood of a fire depending on relative humidity and temperature. It was created in Sweden and is used in some regions of Scandinavia to predict when fires will start on a particular day (Zacharakis and Tsihrintzis, 2023).

Some models discussed in the state of the art will even tackle the Keetch-Byram Drought Index (KBDI), which is a tool that assesses drought conditions. It is frequently applied to forecast the probability and intensity of wildfires (Service, 2023).

3.3 Influencing factors

Forest fires require an ignition triangle of oxygen, heat source, and fuel, which can be found in trees, grasses, dry bushes, and forest litter. Lightning, scorching winds, and even the sun are the most common natural sources of ignition (Naderpour et al., 2021).

Influencing factors regarding wildfires can be split into two main categories of risk: Long-term risk that is defined by the components of a territory that do not vary over the long term such as landscape factors and population factors. Short-term risk is defined by factors that change, such as weather conditions and temporal factors (Novo et al., 2020).

The following list will encompass the main factors concerning wildfire incidence:

- Weather factors: temperature, humidity, wind speed, precipitation, relative humidity, water vapour pressure, thunderstorms, rainfall, oxygen, CO level (Arif et al., 2021; Gao et al., 2023; Mekala et al., 2023; Sharma et al., 2020);

- Temporal factors: season, month, time of day (Arif et al., 2021);
- Population factors: population density, human activities in the forest, human behaviour, short circuits in power lines that cross the forest (Arif et al., 2021);
- Landscape factors: tree types, slope, distance from agricultural lands, road distance, settlement distance, fuel mode types, vegetation, topography (Arif et al., 2021; Bountzouklis et al., 2023; Novo et al., 2020).

3.4 Forest fire occurrence prediction

Estimating how many and where fires may occur in the next days is essential to making advance plans for dealing with fire crises. Numerous variables, including the weather, lightning, and other elements impact the risk of fire and influence prediction (see section 3.3 for a more complete list of factors regarding fire ignition and spread).

Forest fire occurrence models utilize statistical methods and machine learning to establish a relationship between an output (such as fire reports or hotspots) and influencing factors. This relationship is used to predict the likelihood of forest fires in specific regions or geographical areas (Jain et al., 2020).

3.4.1 Decision tree

The authors of (Abid and Izeboudjen, 2020) presented an algorithm that uses a decision tree classifier and meteorological data to forecast the likelihood of forest fires in Algeria. The input elements were temperature, relative humidity, wind speed, and rain.

The output classifications included fire and non-fire. The model yielded an accuracy of 82.92% and a recall of 0.92 for the fire class when tested on a dataset gathered from two areas in Algeria.

3.4.2 Random Forest Algorithm

The study shown in (Thakkar et al., 2022) investigated the influence of various factors such as humidity, temperature, wind speed, and rain using a hybrid approach with a random forest algorithm and components like Fine Fuel Moisture Content (FFMC), Duff Moisture Code (DMC), Drought Code (DC), and Initial Spread Index (ISI) from the fire weather index. The study made use of a dataset gathered over three years, from January 2000 to December 2003, from the Montesinho natural park in Trás-os-Montes (Cortez and Morais, 2008).

A correlation matrix was utilised and indicated that temperature, month, FFMC, DMC, DC, and ISI are directly related or proportional, while rain and temperature show no correlation. The random forest regression model obtained a mean absolute error of 0.6664 and a regression score of 0.9799, which means the model can explain 97.99% of the variation in the data.

The confusion matrix showed that the model could correctly classify most of the samples as fire or no fire. The paper addressed the fact that weather forecasting is useful to predict favourable climatic conditions for wildfire occurrences.

3.4.3 Boosted Decision trees

The paper "A smart approach for fire prediction under uncertain conditions using machine learning" (Sharma et al., 2020) conducted an experiment on the (Cortez and Morais, 2008) dataset with 8 different classification models, and out of them, boosted decision trees had the highest Area Under the Roc Curve (AUC) value. Results containing the comparison between the classification models can be seen in table 3.1.

Table 3.1: Classification models results on (Cortez and Morais, 2008) dataset.

Algorithm	**Precision**	**Recall**	**F-score**	**Accuracy**	**AUC**
Boosted Decision Trees	76%	76%	76%	72%	0.787302
Decision Forest Classifier	75%	71%	73%	69%	0.753968
Decision Jungle Classifier	70%	80%	75%	69%	0.752381
Averaged Perceptron	67%	90%	77%	69%	0.634921
2 Class Bayes Point Machine	60%	80%	69%	58%	0.514286
Local Deep SVM	81%	61%	70%	69%	0.68254
Logistic Regression	60%	100%	75%	69%	0.749226
Binary Neural Network	60%	70%	64%	63%	0.69969

3.4.4 Logistic Regression

The authors of (Mekala et al., 2023) used a logistic regression model and concluded that there were only three variables: humidity, temperature, and oxygen. It was possible to conclude that high humidity and low temperatures are correlated with a low probability of fire occurrence. The model obtained on the test dataset achieved an accuracy of 93.75%, a precision of 94.12%, a recall of 93.75%, and an F1-score of 93.94%.

3.4.5 ANN

Artificial Neural Networks (ANN) are statistical models that draw inspiration from biological neural networks and are partially based on them. They are able to represent and handle nonlinear interactions between inputs and outputs. The linked algorithms can be applied in a variety of ways and are a component of machine learning (Al-Zebda et al., 2021).

Comparing ANN and SVM

A unique approach for creating a dataset based on data from remote sensing and data mining algorithms to forecast the incidence of wildfires was suggested by (Sayad et al., 2019). Big data, remote sensing and data mining algorithms were combined to process data collected from satellite images and extract insights to predict the occurrence of wildfires.

The dataset includes three factors relating to crop health, soil temperature, and fire indicators that were retrieved from Moderate Resolution Imaging Spectroradiometer (MODIS) tools. It consists of 804 cases from various zones in the centre of Canada, mostly in British Columbia and Quebec, where wildfires had previously occurred.

ANNs and Support Vector Machines (SVM) were employed as data mining methods. According to the study, the suggested model performs better than other models in terms of sensitivity, specificity, precision, and F-score and achieves high accuracy (98.32% for neural networks and 97.48% for SVM).

The table 3.2 shows the comparison of results between the different models, STIFF and ISTFF are frameworks mentioned in other papers, that were compared to the solution proposal of (Sayad et al., 2019).

Table 3.2: Results comparison between ANN and other models

Model	**Accuracy**	**RANK**
ARIMA	83.5%	4
STIFF	65%	5
ISTFF	95%	3
ANN	98.32%	1
SVM	97.48%	2

ANN with KBDI

The authors of (Sadatrazavi et al., 2022) developed a novel model for forecasting wildfires using meteorological parameters and ANNs. A wildfire database was gathered from the United States Department of Agriculture (USDA), and the European Centre for Medium-Range Weather Forecasts (ECMWF) provided satellite data for the model's inputs and outputs.

The inputs include temperature, relative humidity, total pressure, evaporation, soil moisture, snow storage, precipitation, wind speed, Normalized difference vegetation index (NDVI), and the Keetch-Byram drought index. The outputs consist of two values: one (1) for wildfire incidences and zero (0) for records that do not include a fire.

The study made conclusions about the factors influencing fire ignition for temperate forests and boreal forests. It concluded that while precipitation was also significant for boreal forests, the most significant factors for both types of forests were determined to be relative humidity, total pressure, wind speed, and KBDI.

3.4.6 Explainable Artificial Intelligence (XAI)

Explainable artificial intelligence (XAI) approaches are utilised to determine which factors contribute most substantially to the occurrence of wildfires while making models more interpretable for stakeholders. In this regard, the paper "Predicting wildfire ignition causes in Southern France using eXplainable Artificial Intelligence (XAI) methods" (Bountzouklis et al., 2023) used an explainable artificial intelligence framework to classify the source of fire ignition based on environmental and anthropogenic variables.

This study aimed to determine if the sources of unknown-caused fires could be predicted using machine learning. The study's results explained that natural fire prediction accuracy (F1-score 0.87) is higher than human-caused fire prediction, such as accidental (F1-score 0.74) and arson (F1-score 0.64) (Bountzouklis et al., 2023). While this study does not focus entirely on the spatial and temporal prediction of wildfire occurrences, it is worth quoting the feasibility of a method like this for the study of wildfire forecasting.

3.5 Forest fire susceptibility prediction

Creating susceptibility maps from forest fire occurrence prediction is a technique for identifying regions prone to forest fires. It is concerned with the prospect of harm to human health and property, as well as environmental health and its potential implications. This is accomplished by examining the many elements that contribute to the chance of a fire happening in a certain location, as well as the dependency relationships between the components. These components can be both natural and produced by humans (**?**).

3.5.1 BRT, GLM and MDA

The authors of (Pourghasemi et al., 2020) developed forest fire susceptibility maps from Landsat-8 Operational Land Imager (OLI) and MODIS satellite photos from 358 sites in Fars Province, Iran, using three machine-learning approaches to examine the determinants and spatial patterns of forest fire susceptibility.

Elevation, slope, topographical wetness index, aspect, distance from metropolitan areas, annual mean temperature, land usage, distance from the road, annual mean rainfall, and distance from the river were recognised as the 10 most important feature for fire susceptibility.

The geographical correlations between the 358 historical fire sites and the 10 contributing factors were analysed using the Boosted Regression Tree (BRT), Generalised Linear Model (GLM), and Mixed Discriminant Analysis (MDA). Various criteria, such as the ROC curve, accuracy, overall accuracy, and True Skill Statistic (TSS), were used to assess prediction accuracy. These three models provide a forest fire susceptibility map with four risk levels: very high, high, moderate, and low.

The study discovered that among the three models, BRT had the best prediction accuracy (AUC = 0.882) and the lowest standard error, followed by MDA (AUC = 0.856) and GLM (AUC = 0.825). The researchers also employed Learning Vector Quantization (LVQ) to determine the relative relevance of the contributing elements, and they discovered that land use, annual mean rainfall, and slope were the most important determinants for forest fire vulnerability.

According to the study, the BRT and MDA models were adequate and trustworthy for mapping forest fire susceptibility in Fars Province, and the produced maps might aid in the planning and management of forest resources and ecological balances.

3.6 Forest fire risk prediction

Forest fire risk assessment is a scientific process for quantifying risk levels, allowing decision-makers to weigh the benefits and drawbacks of various measures. It includes data about the likelihood and extent of future forest fires based on current fire behaviour such as ignition, spread, suppression, and longevity (Naderpour et al., 2021).

Forest fire risk assessment can be split into four stages (Naderpour et al., 2021):

(a) identification of hotspot locations;

(b) calculation of forest fire susceptibility;

(c) identification of forest fire sensitive areas;

(d) assessment of likely forest fire risk.

3.6.1 Multilayer perceptron and Fuzzy logic supervised approach

The authors of (Naderpour et al., 2021) developed a solution to output a risk index in a geographical area near Sydney, Australia. The risk was determined by a combination of three factors: social vulnerability, physical vulnerability, and a hazard index. Social and physical vulnerability were computed given a set of social and demographical features. While the hazard index was created given a combination of forest fire influencing factors (such as NDVI, slope, rainfall, temperature, and land cover) and fire ignition points, These two elements contributed to the creation of the susceptibility map. The ignition points were based on a historical inventory database of Australian forest fire ignition spots.

The susceptibility model for the harzard index was made with a multilayer perceptron model, which consists of multiple layers of neurons connected by weighted links (Bento, 2021).

The model was then optimised using a Fuzzy Logic Supervised Approach (FbSP) to find the best values for the hyper-parameters.

The result of the study can be seen in Fig. 3.1, where a map for forest fire risk is shown.

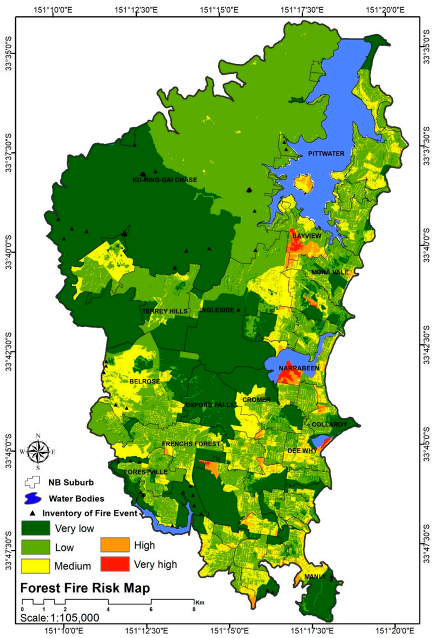


Figure 3.1: Forest fire risk map

3.6.2 Ant-miner algorithm

A study conducted by (Zheng et al., 2020) mined decision rules using historical fire data and risk factor data from Chongqing, China. The authors stated risk factor data as a combination of the following parameters:

- Meteorological data: wind speed, air temperature, relative humidity, accumulated precipitation, continued non-precipitation days;
- Land cover data: land cover type;
- human activity data: distance from nearest road and residential area.

An ant-miner algorithm was used to mine decision rules from the data and forecast the danger level of forest fire across the study region. A confusion matrix was also used in the study to assess the accuracy of the proposed model, which was then compared to existing models such as artificial neural networks, and support vector machines.

The model was also compared to a risk prediction model based on meteorological data that calculated a Forest Fire Danger Weather Index (FFDWI) according to China's national standard using five meteorological factors (wind speed, air temperature, relative humidity, accumulated precipitation, and continuous non-precipitation days).

Another risk prediction model was based on the artificial neural network (ANN) technique, which employed the same risk elements (meteorological data, land cover data, and human activity data) to train a multilayer perceptron network for forecasting the risk level of forest fire. A support vector machine (SVM) was used and employed the same risk parameters as the proposed model to train a kernel-based classifier for predicting the risk level of forest fire.

The model's performance was compared using a confusion matrix. In terms of overall accuracy and Kappa coefficient, the ant-miner model surpassed the others.

The study acknowledged that the proposed model may lack generalisability across different geographical locations and may be unable to describe the risk distribution of forest fires as a specific mathematical function when multiple factors' joint probability distribution functions are considered.

3.6.3 Fuzzy Inference

The authors of (Lin et al., 2018) offer a novel approach for predicting and assessing forest fire risk using fuzzy inference (the process of applying fuzzy logic to create a mapping from a given input to an output (Kalogirou, 2009)), and large data analysis. Figure 3.2 explains visually the basic logic of the approach.

The project was built with a Rechargeable Wireless Sensor Network (RWSB) in mind that is collecting continuous 24-hour meteorological data, and installed in a forest region in the Nanjing City region of China. Temperature, humidity, wind speed, rainfall, season, date, time, historical fire data, population density, fuel type, and road density are all input parameters.

The method generates the fire rating output by using triangular fuzzy numbers to reflect the uncertainty and ambiguity of the input parameters. The output fire rating is divided into five categories: low, moderate, high, very high, and extreme. It was concluded that the system can accurately forecast possible fire danger and provide helpful information for forest fire prevention and management. Figure 3.3 depicts fire probability given the oscillation of influencing factors.

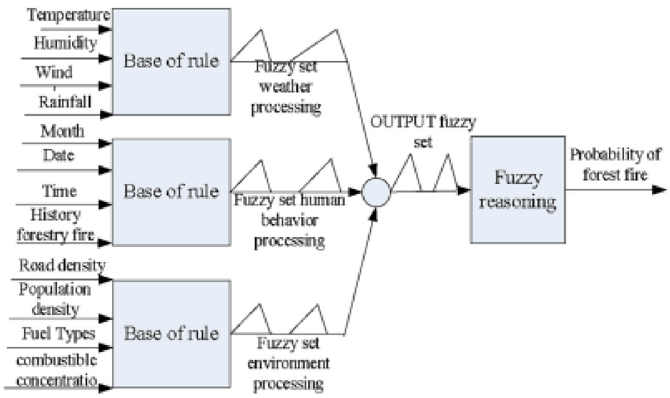


Figure 3.2: Prediction of forest fire risk using a fuzzy reasoning algorithm (Lin et al., 2018).

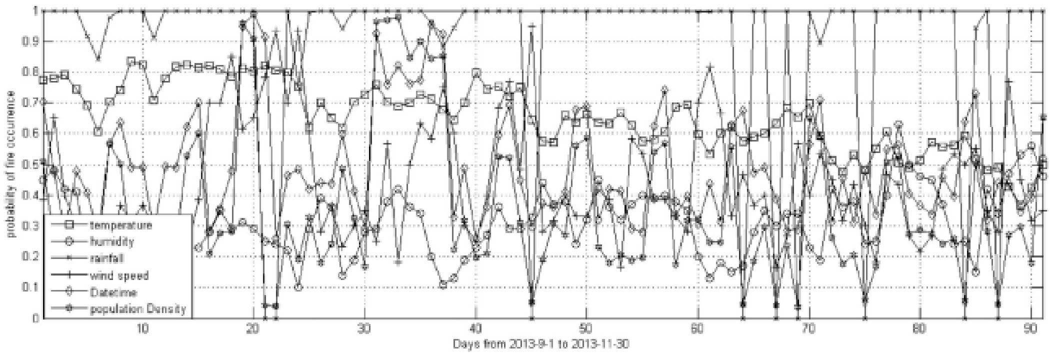


Figure 3.3: Meteorological data, datetime, and people density to predict the likelihood of a forest fire (Lin et al., 2018).

3.6.4 Auto-sklearn

An autonomous machine learning framework for forecasting the danger of forest fires based on meteorological data was proposed by (Qu and Cui, 2020). This framework is capable of adapting to varied datasets and geographies. When it comes to binary unbalanced datasets, the machine learning techniques currently in use for forest fire prediction are not very compatible.

The dataset used in the article comes from Montesinho Nature Park (Cortez and Morais, 2008) and contains historical data about fire occurrences in the place. The framework optimises auto-sklearn, in terms of data preprocessing, parameter learning, and loss function.

The experimental findings demonstrated that the framework has strong geographic flexibility and performs better than other state-of-the-art machine learning techniques. Table 3.3 illustrates how Auto-Sklearn Pro performs better than other models when meta-learning is modified and a new bayesian optimisation index is included.

Table 3.3: Forest district classifier accuracy comparison

Forest District	Classifier	Accuracy (Overall)	Accuracy (Minority sample)
District1	AUTO-SKLEARN PRO	87.3%	94.2%
	AUTO-SKLEARN	86.4%	72.5%
	SVM	84.6%	67.3%
	RandomForest	75.5%	69.2%
	AUTO-SKLEARN PRO	85.6%	90.4%
District2	AUTO-SKLEARN	89.5%	74.3%
	SVM	82.3%	63.2%
	RandomForest	71.1%	73.0%
	AUTO-SKLEARN PRO	82.4%	87.1%
District3	AUTO-SKLEARN	85.3%	79.8%
	SVM	73.1%	54.4%
	RandomForest	69.5%	72.1%

3.6.5 DFP-MnBpAnn

The authors behind (Tien Bui et al., 2018) proposed DFP-MnBpAnn, a new hybrid machine learning technique for spatial modelling of forest fire threat based on Artificial Neural Network (Ann) and a unique hybrid training algorithm of Differential Flower Pollination (DFP) and mini-match backpropagation (MnBp).

Flower Pollination technique is a a nature-inspired, metaheuristic optimisation technique that mimics flower pollination. Solutions are viewed as flowers in FPA, and the process of identifying the optimum solution is analogous to pollination.

Differential Evolution (DE), on the other hand, is a population-based optimisation technique that employs the evolutionary notion. It begins with a randomly generated population of solutions and improves them repeatedly using mutation, crossover, and selection procedures.

The two algorithms are merged in such a way that they compliment each other in a hybrid training algorithm of DE and FPA (Abdel-Basset and Shawky, 2019).

The factors influencing forest fire were slope, aspect, elevation (m), land use, NDVI, distance to road (m), distance to residential area (m), temperature (°C), wind speed (m/s), and rainfall (mm). They were generated from GIS data (see Figure 3.4 for a complete view of the proposed framework) to apply the suggested strategy to a tropical forest in Vietnam's Lam Dong province.

The DFP-MnBpAnn model outperformed the other six approaches in terms of prediction accuracy (88.43%) and AUC (0.94), highlighting its superiority and promise for large-scale forest fire threat mapping. The table 3.4 presents the complete results of DFP-MnBpAnn's performance metrics against other models.

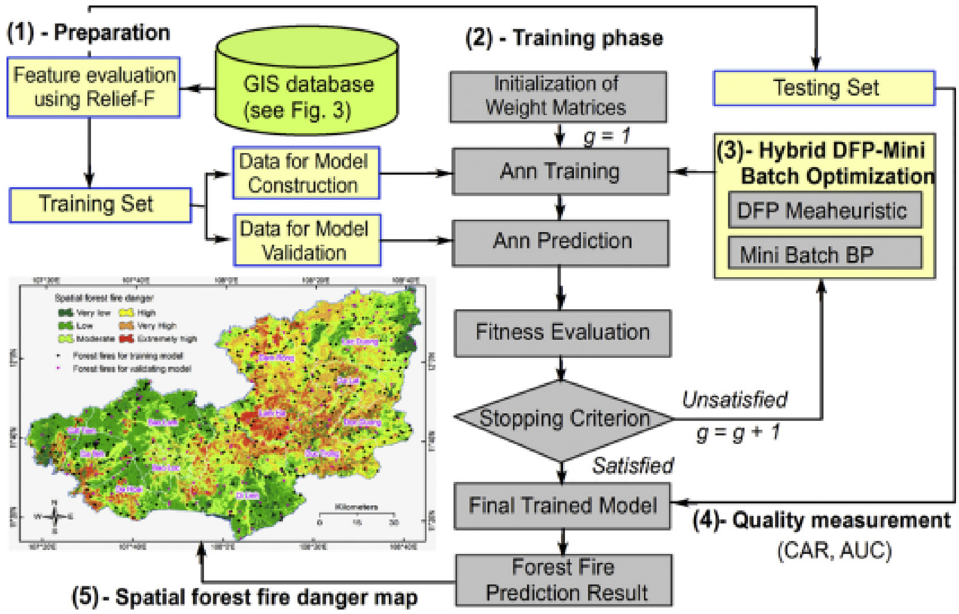


Figure 3.4: Methodology used with DFP-MnBpAnn(Lin et al., 2018).

Table 3.4: DFP-MnBpAnn performance metrics against other prediction models (Tien Bui et al., 2018)

Phase	Model	Performance Metrics					
		CAR (%)	AUC	TP	FP	FN	TN
Training	DFP-MnBpAnn	87.30	0.91	0.94	0.20	0.06	0.80
	PSO-NF	89.30	0.93	0.96	0.17	0.04	0.83
	RF	86.40	0.91	0.91	0.19	0.09	0.81
	SVM	86.20	0.89	0.92	0.19	0.08	0.81
	LSSVM	86.24	0.89	0.90	0.18	0.10	0.82
	BpAnn	89.02	0.94	0.93	0.15	0.07	0.85
Testing	DFP-MnBpAnn	89.20	0.94	0.94	0.15	0.06	0.85
	PSO-NF	85.80	0.92	0.92	0.20	0.08	0.80
	RF	85.20	0.91	0.90	0.19	0.10	0.81
	SVM	84.90	0.88	0.88	0.19	0.12	0.81
	LSSVM	86.42	0.88	0.92	0.19	0.08	0.81
	BpAnn	85.03	0.89	0.91	0.21	0.09	0.79

3.7 Methodologies for data aggregation, fusion, and enhancement

3.7.1 Data aggregation

The process of merging information from several sources into a single dataset is known as data aggregation. These sources may consist of databases, sensor networks, or other types of data repositories containing Weather, location, and historical fire event data (Cai et al., 2019). Data aggregation can be split into two key-categories (Papčo et al., 2021):

- Temporal aggregation: Combining data across predetermined time frames. Finding trends linked to fire seasonality, such as daily meteorological data that may be combined to produce monthly or seasonal averages (Papčo et al., 2021).

- Spatial aggregation: Combining data across geographic regions. This can aid in comprehending how fire risk elements are distributed spatially (Papčo et al., 2021).

3.7.2 Data fusion

Data fusion refers to the need to combine information obtained from many sources (sensors, databases, information gathered by people, apis) in order to present complementary interpretations of the same phenomenon. It provides more exact results than those obtained by using a single dataset, eliminating ambiguity and redundancy. (Chatzichristos et al., 2022; Lacroix, 2003)

Figure 3.5 describes visually two data fusion techniques that will be utilised. The following list explains each one:

- Decision-Level fusion: In this technique, characteristics are extracted from the data and one or more intermediate analyses are carried out until an appropriate result for the problem is achieved (Schmitt and Zhu, 2016).

- Feature-level fusion: This technique involves collecting features from each data source and then merging them (Schmitt and Zhu, 2016).

3.7.3 Data enhancement

The process of changing, converting, or adding to data in order to make it better and easier to use is known as data enhancement (Clickworker, 2023). It can be achieved by using multiple paths such as:

- Data Enrichment: Using numerous third-party data to supplement the original data (mParticle, 2021).

- Synthetic Data Generation: Created artificially without utilising the original dataset (Awan, 2022).

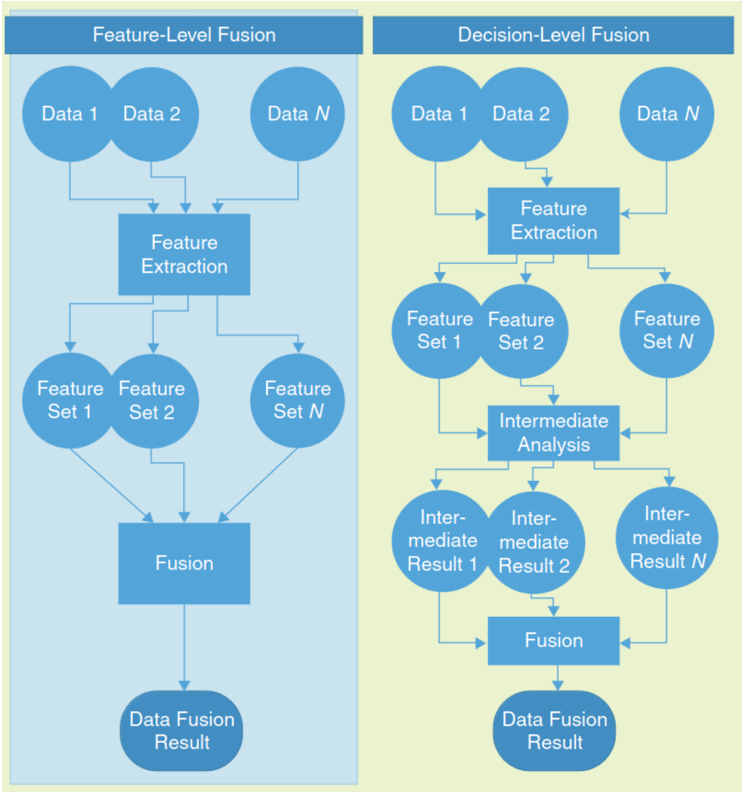


Figure 3.5: Methodologies for Data fusion (Schmitt and Zhu, 2016).

3.8 Tools for data visualization

3.8.1 Seaborn

Seaborn is a high-level interface for creating visually appealing statistical visuals and is developed on top of Matplotlib, which is a charting toolkit and allows the creation of all kinds of plots and charts (Waskom, 2021)(Hunter, 2007).

3.8.2 Folium

Folium is a Python wrapper for Leaflet.js map tool. It helps in spatial data visualisation (Story, 2013).

3.8.3 GeoPlot

It is based on Matplotlib and GeoPandas. Geoplot offers a high-level mapping creation interface. GeoPandas is a library extension for Pandas that facilitates mapping and has geographic functions. It manages and produces maps using geospatial data (Bilogur and Contributors, 2016; Hunter, 2007; Jordahl et al., 2020).

3.8.4 Plotly

Plotly is a robust and interactive charting library. It works with many different kinds of charts and is helpful for making interactive visualisations. (Inc., 2015).

3.8.5 Rasterio

Rasterio is used to read and write geographic raster data, such as satellite images. (Gillies et al., 2013).

3.8.6 Pydeck

Pydeck is a WebGL-powered, high-layer framework that is used to create dynamic maps covering sizable regions (Contributors, 2021).

3.8.7 Rasterstats

Geospatial raster datasets may be summarised using rasterstats based on vector geometry. It is also used to compute statistics within certain regions (Perry and Contributors, 2015).

3.8.8 Fiona

The goal of Fiona is to simplify the reading and writing of geographic data files. It makes dealing with vector data easier and is based on GDAL which is a potent open-source library for reading and writing geographic data formats in both raster and vector forms. (GDAL/OGR contributors, 2024; Gillies and Contributors, 2011).

3.9 State-of-the-art conclusion

Throughout this chapter, various methodologies for predicting occurrence, susceptibility, and risk of fires have been discussed. As this problem is highly dependent on the data that can be obtained from multiple sources, some methodologies like (Sadatrazavi et al., 2022) and (Bountzouklis et al., 2023) can be discarded, as they lack generalisation and are only suited for one purpose. (Sadatrazavi et al., 2022) carried out a study between the susceptibility of two types of forests and concluded the factors that influence the ignition of fire in each type, and (Bountzouklis et al., 2023) showed that naturally ignited fire is the easiest one to predict.

The paper conducted by (Mekala et al., 2023) should also not be taken as guidance. Although the employed logistic regression obtained an accuracy value of 93.75%, its approach is very narrow as only three variables were used.

For an initial prototyping phase, random forests seem to be a good candidate, as shown in (Abid and Izeboudjen, 2020) and (Thakkar et al., 2022). The authors of (Abid and Izeboudjen, 2020) used a decision tree classifier and obtained an accuracy of 82.92%. While (Thakkar et al., 2022) didn't explicitly mention the accuracy metric, its random forest classifier obtained a regression score of 97.99%.

ANN is a great option overall for fire occurrence prediction. (Sayad et al., 2019) compared the accuracy between ANN and SVM, while these two had very high accuracy ratings. ANN got the upper hand with 98.32% against SVM's 97.42%. Once the prototyping phase with random forests has been completed, ANN could be the de facto option for forest fire occurrence prediction.

The study carried out by (Pourghasemi et al., 2020) demonstrated that for fire susceptibility, boosted regression trees are better than GLM and MDA. Fire susceptibility can be inherently connected to risk prediction. It can be a step in the prediction of risk. Therefore, it is not necessary to provide a specific solution for susceptibility like the authors of (Pourghasemi et al., 2020) did.

Most of the studies conducted in the field of risk prediction employed a hybrid approach. These approaches lack generalisation, as seen in (Zheng et al., 2020), and are much harder to implement.

The authors of (Tien Bui et al., 2018) used a novel approach to map the forest fire ignition risk. This method used a hybrid model and yielded a high AUC score of 0.94. Although this method uses a hybrid approach, the underlying model is still an ANN. Emphasising once again the usefulness of an ANN approach for both occurrence and risk prediction.

The study conducted by (Qu and Cui, 2020) created a framework with auto-sklearn capable of adapting itself to multiple geographical locations. This is a very important feature for risk prediction, and this approach can also be taken into consideration as it uses an easy to obtain dataset, so result metrics can be easily compared. This study had an accuracy of 87.3%.

The tools for data visualisation are also dependable on the obtained data. But tools like plotly (3.8.4) and seaborn (3.8.1) will always be used due to their generic capabilities. These libraries serve the purpose of plotting data.

The methodologies for data aggregation, fusion, and enhancement presented previously will also be used, as they represent standard procedures for machine learning data preparation.

Chapter 4

Solution Proposal

This chapter will define the problem, establish how success will be evaluated, define the methodology for developing the solution, and present the associated risks.

4.1 Problem definition

The challenge is to design and develop a comprehensive system that can effectively process, validate, and aggregate voluntary contributions supported by various types of data. This system would need to be able to classify the severity of forest fire events based on aggregated data.

Moreover, it would need to identify the geolocation and track the temporal evolution of a wildfire event. Providing properly organised and validated information is crucial, as it can help increase knowledge about the event, facilitate real-time decision-making, and potentially save lives and resources.

Finally, the system would need a set of plotting tools to aid in information gathering and posterior analysis.

Figure 4.1 shows in rudimentary way, multiple geographic location that don't have correlation to each other. Each tile represents a different biome, and each one is named after a letter. Fire risk is composed of four distinct categories, red which is associated with extreme risk of fire, orange is high risk, yellow is moderate risk, and green is low risk of fire. Tile C is on the worst-risk. There are several active fire fronts, and it is windy.

Tile A has high risk and has a forest near a rural area, this may have enhanced fire ignition due to human activities. Tile B is near a town, but since it isn't windy or very hot, it has a moderate risk of fire. Tile D is the least likely to catch on fire, it is raining and the weather is cold. It has a risk score of green signalling a low risk of fire.

This imaginary exercise can be converted into the real world, because it entangles the problem at its core. Create a fire risk and occurrence analysis with multiple sources of data and then show it in a easy manner.

The tiles would be geographical areas, and risk would be associated with it given each region's fire influencing factors.

In essence, the goal is to create a comprehensive system that can handle a wide range of data types, make sense of this data, and provide actionable insights about forest fire events.

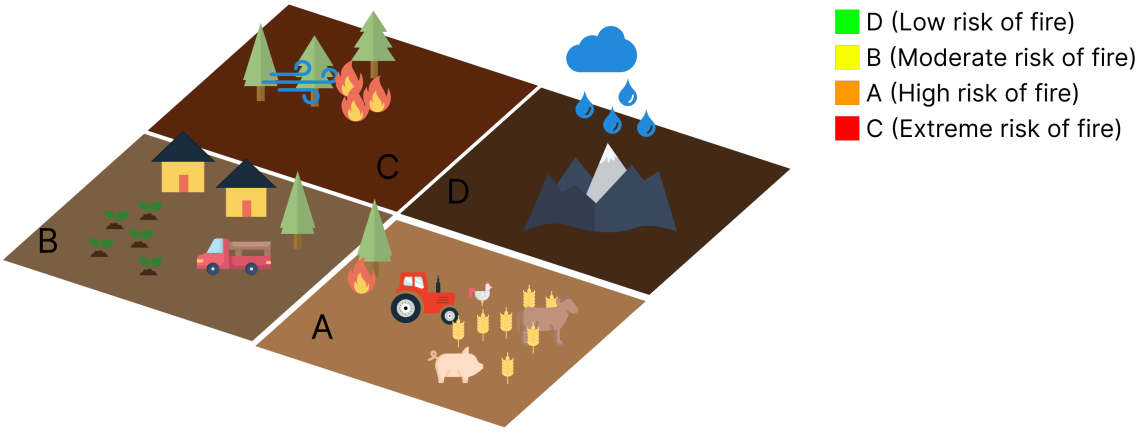


Figure 4.1: Tiles depicting geographical areas and forest fire risk

4.2 Proposed Methodology

The methodology section will cover the steps for carrying out the solution. Firstly, data is collected from various sources to create a data pipeline. This is followed by the step of merging the data according to its features. The fusion can be done recursively until there is satisfaction with the information obtained.

The data will be combined with other external sources, and depending on the scarcity of the data, it may be necessary to generate synthetic data. Once the data has been cleaned, it will be placed in a database.

The models will use the database to obtain classification results. The results will focus on risk, susceptibility, and fire occurrence. Finally, with the help of data visualisation tools, the findings will be displayed.

4.2.1 Data Pipeline

Data pipeline is the process of extracting, transforming, and loading data from diverse sources, such as volunteer contributions, geographical data, and meteorological data. The following steps comprise the data pipeline:

Data extraction

Data is gathered from a variety of sources, including websites, APIs, and databases.

Data transformation

Following extraction, the data is cleaned (duplicates are removed and missing values are handled), normalized (ensuring that is in a consistent format), and enriched.

Data loading

The transformed data is then loaded into an appropriate data storage system, such as a database. This enables efficient searching and retrieval of data when it is required for analysis.

4.2.2 Data Fusion

Data fusion can happen at the data transformation step. At this level, decision-level and feature-level fusion will occur (see 3.7.2 for a more detailed explanation of each one).

4.2.3 Data enhancement

Data enhancement follows the fusion step. At this level, data can be enriched with external sources or generated.

4.2.4 Classification Models

Experiments will be conducted on data with multiple machine learning models to convey the ones with better metrics and performance and highlight the most useful features in forest fire event classification. For fire occurrence, Random Forest and SVM are strong candidates for a first approach. These two models are easy to implement and highlighted high results in (Thakkar et al., 2022) and (Sayad et al., 2019). For fire risk, an ANN model is the strongest candidate, as shown in (Naderpour et al., 2021).

4.2.5 Data Visualisation

Data visualisation will enable the plotting and presentation of forest fire severity classification results, as well as the display of aggregated and validated information.

4.2.6 Outcome

The project's output would be the information derived from the classification models. If in a given area or given some meteorological indices, the system will determine the likelihood of fire (fire or no fire) and the associated risk (low, moderate, high, or extreme). This information will be presented using data visualisation tools.

4.3 Planning and organisation

Figure 4.2 exhibits the time taken for each step in the first semester. It has a comparison between the time expected for each task and the actual time taken to complete it. The comprising steps in the first Gantt chart are:

1. Learn Key Concepts - Forest Fire Management, Issues, Decision Support systems;
2. Learn Key Concepts - Machine learning applied to forest fire, Mathematical models, and Influencing factors;
3. Learn models used to classify Forest Fire;
4. Research data visualisation tools and libraries;
5. Study of the problem;
6. Define the solution, objectives, success criteria, risks and methodology;
7. Write document.

Figure 4.3 shows the proposed Gantt chart for the second semester. The following list will explain each step in figure 4.3:

1. Collect data from multiple sources: The first step involves the gathering of multiple data sources;
2. Extract, Transform, and Load Data: Apply extract, transform, and load techniques to the gathered data. In this step, the data pipeline is created.
3. Fusion techniques on data: This step undertakes the fusion methodologies inside the pipeline;
4. Conduct forest fire event classification with RF and ANN models: Random forest and ANN have been chosen as the first candidates to tackle the problem of forest fire occurrence and risk;
5. Utilise fused and classification outputs to visualise the information: This is regarded as the last step in the system's implementation. The findings discovered in the previous steps are now interpretable information ready to be used in a decision-support system;

6. Assemble the scientific paper: Besides the thesis, this project should result in a scientific paper;
7. Write document: The last step is the completion of the thesis document.

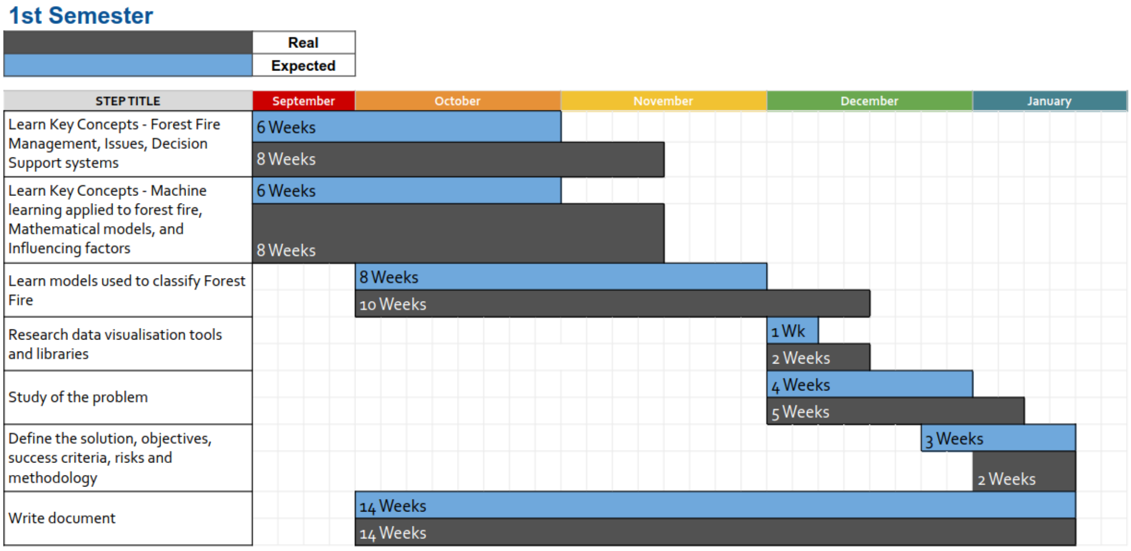


Figure 4.2: Gantt chart for first semester

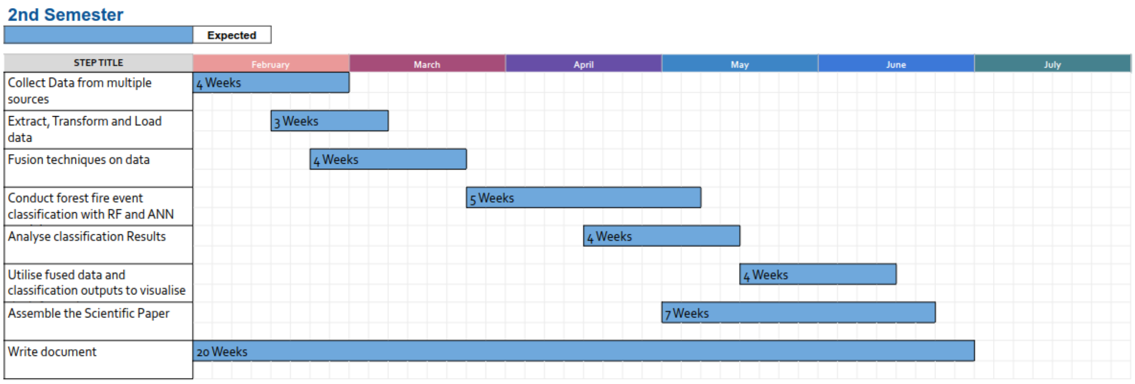


Figure 4.3: Proposed gantt chart for the second semester

4.4 Risk analysis

In this section, the potential risk that may affect the success of the approach is identified, and mitigation and impact plans for the risk are proposed. A qualitative probability and impact level were also assigned using the following scales: low, medium, and high. Table 4.1 outlines the only identified risk.

4.5 Success Evaluation Criteria

The success of the project can be measured in two dimensions: the realisation of the proposed objectives and how closely the outcomes resemble reality. That

Table 4.1: Risk - Data unavailability

Risk	Data unavailability
Description	Due to technological challenges, privacy limitations, or other factors, some data sources may not be available.
Mitigation Plan	To provide data redundancy and variety, many data sources are employed.
Impact Plan	Identify alternative data sources. Generate data. Adapt the prediction model to work effectively with the available data. Adapt the experiment to the available data.
Probability	High
Impact	High

is, whether the system is able to assess whether there is a likelihood of fire or not. Also, regarding success outcomes, data visualisation should play a key role in showing findings in a clean manner. Therefore, the following steps must be considered when evaluating the success of the outcomes:

4.5.1 Classification performance

The classification model's accuracy, precision, recall, F1-score, and AUC will be evaluated. The models will be also compared to other classification results and sources on historical data. Different machine learning algorithms' performance will be examined, and feature selection approaches will be utilised to discover the best model and the most significant characteristics for the classification problem.

4.5.2 Visualization effectiveness

The visualisation should be simple, and easy to follow. This can be achieved by following design guidelines and making comparisons with other data visualisation interfaces.

Chapter 5

Implementation

5.1 Data Sources

5.1.1 Historical record of fires from 1980 to 2015 in mainland Portugal (centraldedados, 2017)(Central de Dados Github Repository, 2017) (ICNF, 2024)

//describe dataset here, how many entries?

5.1.2 Historical record of fires from 2013 to 2023 in mainland Portugal (ICNF, 2024)

Data retrieved from the API endpoint. //describe dataset here, how many entries?

5.1.3 Open-meteo hourly weather variables (Zippenfenig, 2023)

Historical Weather (Zippenfenig, 2023) from (Hersbach et al., 2023), (Muñoz Sabater, 2019) and (Schimanke et al., 2021)

5.1.4 Fire danger indices historical data from the Copernicus Emergency Management Service (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019)

The dataset "Fire danger indices historical data from the Copernicus Emergency Management Service" (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019) provided by ECMWF, contains full historical reconstruction of weather conditions suitable for the origin, spread, and sustainability of natural occurring fires.

Table 5.1: Field description of Historical fires from 1980 to 2015

Variable	Description
ano	Year of fire occurrence
codigo_sgif	Unique identifier for the fire occurrence
tipo	Kind of wildfire. The available options are: forest, slash-and-burn, false alarm, and agricultural.
distrito	District of fire occurrence.
concelho	Municipality of fire occurrence.
freguesia	Parish of fire occurrence.
local	Location of fire occurrence.
ine	Not described or explained anywhere.
x | y	Wildfire location.
data_alerta	Wildfire warning date.
hora_alerta	Wildfire warning hour.
data_extincao	Wildfire complete extinguishment date.
hora_extincao	Hour of wildfire complete extinguishment.
data_primeira_intervencao	Date of first fire intervention
hora_primeira_intervencao	Hour of first fire intervention
fonte_alerta	Authority or group of people who reported the fire first.
nut	A Unique identifier for a given nomenclature of territorial units for statistics.
area_povoamento	Burnt settlement area.
area_mato	Burnt bush area.
area_agricola	Burnt agricultural area.
area_pov_mato	Sum of burned area from the burnt settlement area and burnt agricultural area.
area_total	Total burnt area.
reacendimento	Describes if a given fire is a re-ignition from a previous wildfire.
queimada	Identifies if a fire is a slash-and-burn.
falso_alarme	Identifies if it is a false alarm.
fogacho	Identifies if it is a specific type of fire named a blaze.
incendio	Identifies if it is a fire.
causa	Numerical identifier for the fire cause.
tipo_causa	Description of fire cause. The available options are unknown, deliberate, natural, negligent fire, and undefined.

It embodies fire danger indices from three distinct models created in Canada, United States and Australia. The fire danger indices are obtained from historical simulations and weather forecast provided by the dataset ECMWF ERA5 reanalysis.

The available data starts from January 1940 and it extends all the way through 2023, but the data records are regularly extended with time as ERA5 forcing data becomes available. The variables contained in the dataset are expressed in the table 5.4.

5.1.5 Forest Inventory 2015 (GBIF.Org User, 2024; Uva et al., 2021)

Forest Inventory 2015 contains 579422 occurrences of forest inventory for mainland Portugal. The data was gathered using aerial images and ground surveys that covered mainland Portugal. //describe dataset here

5.2 Additional sources of Data

The python library geopy (Community, 2023) was used to geolocate multiple locations, resolving district, parish, municipalities, and localities to sets of coordinates. Geopy utilises multiple geocoding web services like OpenStreetMap Nominatim and Google Geocoding API to resolve locations.

The Google Maps service (Google, 2024) was used to manually check if the extracted data from Open-Meteo corresponded to the intended location. It was also used to analyse some errors that were found in the location of some entries.

5.3 Creating the dataset

The dataset described in 5.1.1 is composed of multiple files describing historical occurrences since 1980 until 2015. Prior to 2001, the fields from each file became unstandardized, and there's no explicit parameter mentioning a natural wildfire cause. Therefore, the time frame considered was from 2001 to 2012. The latter years were rejected due to the fact that entries from 5.1.1 do not contain any explicit latitude and longitude. They rely on territorial entities such as districts, municipalities, parishes, and NUTS to describe locations.

The second historical wildfire dataset 5.1.2 is also composed of multiple files. Its time frame is from 2013 until 2023. Unlike dataset 5.1.1, entries do contain an explicit latitude and longitude values. It also features descriptive territorial entities.

5.3.1 Entry selection

Entries whose cause was deliberate or negligent fire were excluded. The fire causes contained, by order of importance, in the dataset are: natural, reignition, unknown, and undefined. Entries that were undefined as causes differed from those with unknown causes because their cause field was blank, and entries that had unknown causes were explicitly described as unknown.

falta: tabela com o número de entradas antes e depois

5.3.2 Geocoding places from 2001 to 2012 historical wildfire locations

The dataset entries featured in 5.1.1 contain no direct field leading up to the real site location coordinates. To tackle this issue, an algorithm with the help of the geopy library (Community, 2023) was made to resolve the names of historial wildfire places to a set of coordinates.

Using multiple combinations, attempts were made to geocode the location, featuring the combinations in the table 5.6. The district, municipality, parish, and local (if available) of each entry were utilized for this purpose. Sometimes, the name of the exact wildfire locality was enclosed in brackets, requiring processing using strings to extract it.

These combinations caused errors in the location of some entries because the geocoders returned coordinates in other countries, such as Spain and Brazil, due to similar names in some locations. The entries that produced errors underwent recalculation, with the addition of "Portugal" at the end. An example of this usage is *Parish, District, Portugal*.

After each entry was resolved, their latitude and longitude were added as values in the columns LAT and LON of each corresponding file.

A very minor sample of entries couldn't be geocoded using this method. Therefore they were manually geocoded from the Google Maps service.

5.3.3 Retrieving historical meteorological data

In order to retrieve historical meteorological data, a Python script was made. It went through each historical fire location and downloaded weather data about the entire day regarding the wildfire. The weather data contained all the fields described in 5.1.3.

5.3.4 Linking historical wildfires with historical weather data

Dizer número de instâncias Every unit from each field was specified in the retrieved data, so it had to be removed. Dizer que NA ficou com NC no dataset.

5.3.5 Matching each historical wildfire with tree species

O que mudei no dataset inicial das trees dfTreesDRP['stateProvince'] = dfTreesDRP['stateProvince'].replace('Bragança District', 'Bragança') dfTreesDRP['locality'] = dfTreesDRP['locality'].replace('Ovadas e Panchora', 'Ovadas e Panchorra') Continha um erro, Ovadas e Panchora não existe.

variáveis das trees que foram usadas: scientificName,locality,stateProvince,occurrenceStatus,individ haversine formula da distncia (SimonKettle, 2017).

Multiple tests were made for distance, 120metres, 500 metres.

Especies que estavam na mesma freguesia ou concelho foram associadas sem fazer cálculo de distancia. Para combater espécies duplicadas, só se adicionava uma espécie se esta não estivesse contida na entrada do fogo. Distrito, usava-se uma distância de 1000 metros.

As restantes entradas que não obtiveram correspondencia com os outros métodos anteriores, foi feita uma análise das espécies que estavam mais perto, neste passo foram detetados erro, alguns valores tabulados do icnf não correspondiam à realidade, e alguns valores em 5.3.2 foram mal calculados.

Devido ao tamanho do dataset das arvores o script de python utilizado dividiu os anos em chunks e com multiprocessamento foi calculado as especies de arvores perto do fogo.

5.3.6 Locations in the middle of the sea.

Between 2013 and 2023, some of the featured locations were in the middle of the ocean. Although having a real-life location set explicitly in the file when using services like Google Maps, its coordinates were undeniably wrong. These multiple geolocation errors were discovered when trying to pin multiple species of trees 5.3.5 to a single location with a distance function calculator. The algorithm yielded values that were outside of the range spectrum of 1500km. Leading to the manual confirmation of these errors with the help of the Google Maps service.

5.3.7 Dataset description

2001: 25982 2002: 25650 2003: 25138 2004: 21189 2005: 34578 2006: 19175 2007: 15615 2008: 9905 2009: 17399 2010: 14431 2011: 14913 2012: 11841 2013: 11899 2014: 3833 2015: 8431 2016: 7782 2017: 10412 2018: 5559 2019: 4040 2020: 3953 2021: 2610 2022: 4040 2023: 2499 TOTAL : 300874

Natural Fires 2001: 44 2002: 13 2003: 96 2004: 16 2005: 3 2006: 67 2007: 50 2008: 28 2009: 106 2010: 138 2011: 102 2012: 56 2013: 77 2014: 38 2015: 138 2016: 67 2017: 104 2018: 114 2019: 128 2020: 95 2021: 103 2022: 115 2023: 72 TOTAL: 1770

Reignition Fires - Averiguar os zeros 2001: 0 2002: 0 2003: 0 2004: 0 2005: 0 2006: 0 2007: 0 2008: 0 2009: 0 2010: 0 2011: 0 2012: 2256 2013: 2430 2014: 305 2015: 1505 2016: 1347 2017: 1714 2018: 712 2019: 580 2020: 524 2021: 201 2022: 480 2023: 247 TOTAL: 12301

Unknown Fires 2001: 215 2002: 221 2003: 234 2004: 268 2005: 377 2006: 1557 2007: 3577 2008: 3253 2009: 4422 2010: 6071 2011: 5915 2012: 4041 2013: 4133 2014: 2273 2015: 3834 2016: 3497 2017: 5266 2018: 2949 2019: 2506 2020: 2480 2021: 2101 2022: 3154 2023: 1863 TOTAL: 64207

NA Fires

5.4 Python libraries used in the conception of the dataset

requests pandas os to check if files already existed.

5.5 Entry Selection

Specifity how many entries raw files have.

Table 5.2: Field description of Historical fires from 2013 to 2024

Variable	**Description**
CODIGO id	Unique identifier for the fire occurrence
DISTRITO	District of fire occurrence.
TIPO	Kind of wildfire. The available options are: forest and agricultural fire.
ANO	Year of fire occurrence
AREAPOV	Burnt settlement area.
AREAMATO	Burnt bush area.
AREAAGRIC	Burnt agricultural area.
AREATOTAL	Total burnt area.
REACENDIMENTOS	Boolean value for reignition.
FOGACHO	Boolean value for small fire.
NCCO	Non specified identifier.
NOMECCO	Not described or explained anywhere.
DATAALERTA	Wildfire warning date.
HORAALERTA	Wildfire warning hour.
LOCAL	Location of fire occurrence
CONCELHO	Municipality of fire occurrence.
FREGUESIA	Parish of fire occurrence
FONTEALERTA	Authority or group of people who reported the fire first.
INE	Not described or explained anywhere.
X | Y	Wildfire location.
DIA	Day of the fire occurrence.
MES	Month of fire occurrence.
HORA	Fire hour of occurrence.
OPERADOR	Not described or explained anywhere.
PERIMETRO	Not described or explained anywhere.
APS	Not described or explained anywhere.
CAUSA	Numerical identifier for the fire cause.
TIPOCAUSA	Description of fire cause. The available options differ from year to year.
distrito	District of fire occurrence.
local	Local of fire occurrence
data_alerta	Wildfire warning date.
hora_alerta	Wildfire warning hour.
data_extincao	Wildfire complete extinguishment date.
hora_extincao	Hour of wildfire complete extinguishment.
data_primeira_intervencao	Date of first fire intervention
hora_primeira_intervencao	Hour of first fire intervention
fonte_alerta	Authority or group of people who reported the fire first.
nut	A Unique identifier for a given nomenclature of territorial units for statistics.
area_pov_mato	Sum of burned area from the burnt settlement area and burnt agricultural area.
reacendimento	Describes if a given fire is a re-ignition from a previous wildfire.
queimada	Identifies if a fire is a slash-and-burn.
falso_alarme	Identifies if a false alarm.
fogacho	Identifies if it is a specific type of fire named a blaze.
incendio	Identifies if it is a fire.
causa	Numerical identifier for a fire cause.

Table 5.3: Hourly weather variables from Open-meteo

Variable	Unit	Description
Temperature	°C	Air temperature 2 metres above ground.
Relative Humidity	%	Relative humidity 2 metres above ground.
Dew	°C	Dew point 2 metres above ground.
Apparent Temperature	°C	Apparent temperature is the result of a wind chill factor, relative humidity, and solar radiation.
Pressure	hPa	Atmospheric air pressure reduced to mean sea level.
Surface Pressure	hPa	Surface pressure reduced to mean sea level.
Precipitation	mm	Sum of preceding hour precipitation including rain, showers, and snow.
Rain	mm	Preceding hour of liquid precipitation.
Snowfall	cm	Preceding hour of snowfall amount.
Cloud cover low	%	Fog and low level clouds up to an altitude of 2 kilometres.
Cloud cover mid	%	Clouds floating at a medium level with altitudes ranging from 2 kilometres to six kilometres.
Cloud cover high	%	Clouds floating at an altitude of 6 kilometres.
Shortwave radiation	W/m^2	Shortwave solar radiation.
Direct radiation	W/m^2	Direct solar radiation.
Direct normal irradiance	W/m^2	Direct solar irradiance.
Diffuse radiation	W/m^2	Diffuse solar radiation.
Global tilted irradiance	W/m^2	Total radiation received on a tilted pane.
Sunshine duration	Seconds	Duration of sunshine in seconds.
Wind speed at 10m	km/h	Speed of the wind, 10 metres above ground.
Wind speed at 100m	km/h	Speed of the wind, 100 metres above ground.
Wind direction at 10m	°	Wind direction at 10 metres above ground.
Wind direction at 100m	°	Wind direction at 100 metres above ground.
Wind gusts	km/h	Wind gusts at 10 metres above ground.
Evapotranspiration	mm	Evapotranspiration value for the required irrigation for plants calculated from temperature, wind speed, humidity, and solar radiation.
Weather code	WMO code	Numeric codes for weather conditions.
Snow depth	meters	The depth of snow on the ground.
Vapour pressure deficit	kPa	Vapour pressure deficit in kilopascal.
Soil temperature	°C	Average soil temperature ranging from 0 to 7cm, 7 to 28cm, 28 to 100cm, and 100 to 255cm below ground.
Soil moisture	m^3/m^3	Average soil moisture ranging from 0 to 7cm, 7 to 28cm, 28 to 100cm, and 100 to 255cm depths.

Table 5.4: Fire danger indices from historical data

Variable	Unit	Description
Build-up index	Dimensionless	Weighted combination of the Duff moisture code and Drought code.
Burning index	Dimensionless	Measure that explains how difficult it is to control a fire.
Danger rating	Dimensionless	Equivalent to the FWI but with class level definitions of very low, low, medium, high, very high and extreme.
Drought code	Dimensionless	Component representing fuel availability, and the influence of recent temperatures and rainfall events on fuel availability.
Duff moisture code	Dimensionless	Moisture content in loosely-compacted organic layers of moderate depth. Duff moisture code fuels are affected by rain, temperature and relative humidity.
Energy release component	J/m^2	Available energy within the burning front at the head of a fire.
Fine fuel moisture code	Dimensionless	Moisture content in litter. Representative of the top litter layer less than 1-2 cm deep.
Fire daily severity index	Dimensionless	A numerical assessment of the difficulty of controlling flames.
Fire danger index	Dimensionless	Metric that hold the chances of a fire starting.
Fire weather index	Dimensionless	Combination of Initial spread index and Build-up index. Numerical rating of the potential fire intensity.
Ignition component	%	Probability of a firebrand that will require suppression action.
Keetch-Byram drought index	Dimensionless	The total impact of evapotranspiration and precipitation in causing cumulative moisture shortage in deep duff and higher soil layers.
Spread component	Dimensionless	Measure of the spead at which a headfire would spread.

Table 5.5: Forest Inventory 2015

Variable	Description
gbifID | datasetKey | ocurrenceID	Identifiers for the occurrence of trees and the dataset.
kingdom	Kingdom classification of a given Tree.
phylum	Phylum classification of a given Tree.
class	Taxonomic class.
order	Taxonomic Order of a Tree.
genus	Tree genus.
species	Data containing the species of a given tree.
taxonRank	Data containing the highest taxonomic rank available for a given tree group.
scientificName | verbatimScientificName	Scientific name for the available taxonomic classification.
verbatimScientificNameAuthorship	Scientific name authorship for the available taxonomic classification.
countryCode	Contry code of Portugal.
locality	Name of a locality containing a given tree.
stateProvince	Name of a district containing a given tree.
occurrenceStatus	Describes if a tree is still present.
decimalLatitude	Latitude for the tree occurrence.
decimalLongitude	Longitude for the tree occurrence
coordinateUncertaintyInMeters	Uncertainty for a given tree location in metres.
eventDate | year	Year of event record.
taxonKey	Taxonomic key for the highest available classification for a tree
speciesKey	Individual key for a given tree species if available.
speciesKey	Individual key for a given tree species if available.
institutionCode	Unique identifier for ICNF.
collectionCode	Unique collection identifier for the institutionCode.

Table 5.6: Combinations for local geocoding

Combination
Local, District
Local, Parish, District
Local, Parish, Municipality
Local, Parish, Municipality, District
Parish, Municipality, District
Local, Parish, District

Chapter 6

Graphs

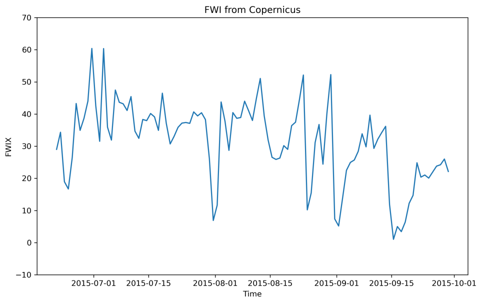
6.1 Sample Description

2022: FWI:21.95067024230957 Drought Code (DC): 418.90625 Duff Moisture Code (DMC): 160.0711212158203 Fine Fuel Moisture Code (FFMC): 86.57350158691406

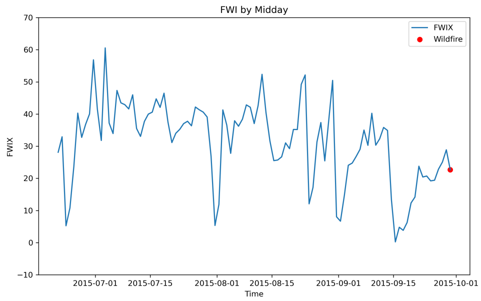
6.2 Simulated FWI variables

Este local não teve incêndio

Figure 6.1: Comparison of FWI calculated values and Copernicus

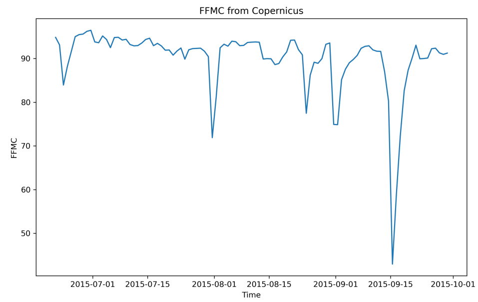


(a) FWI - Copernicus

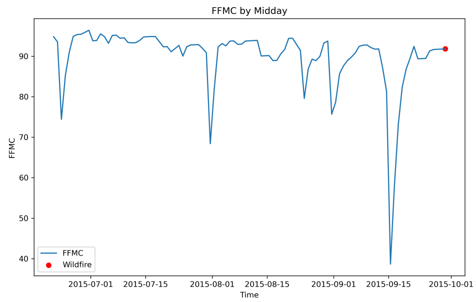


(b) FWI - Calculated value

Figure 6.2: Comparison of FFMC calculated values and Copernicus

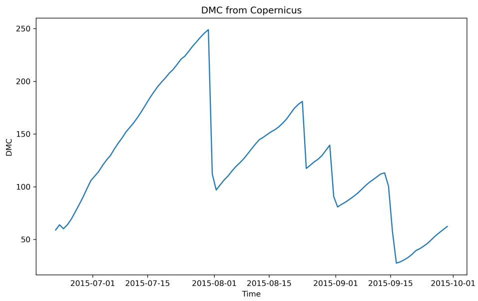


(a) FFMC - Copernicus

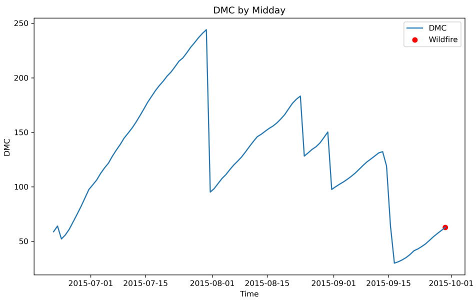


(b) FFMC - Calculated value

Figure 6.3: Comparison of DMC calculated values and Copernicus

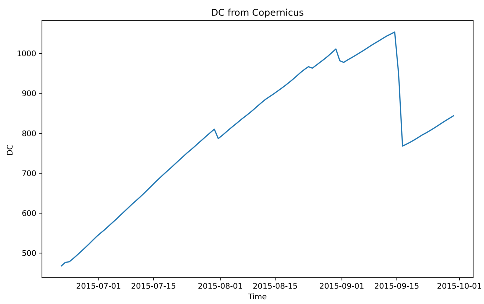


(a) DMC - Copernicus

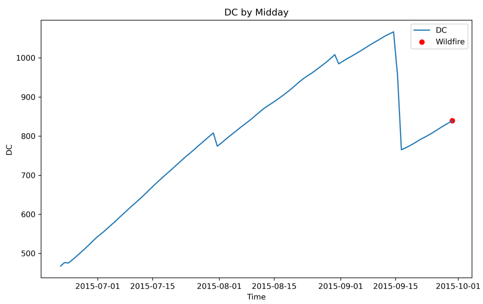


(b) DMC - Calculated value

Figure 6.4: Comparison of DC calculated values and Copernicus

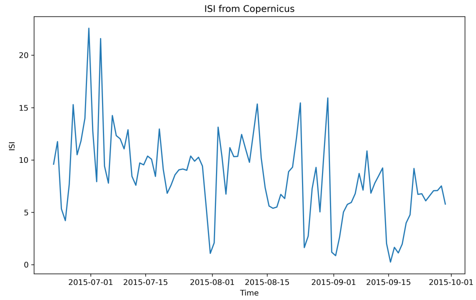


(a) DC - Copernicus

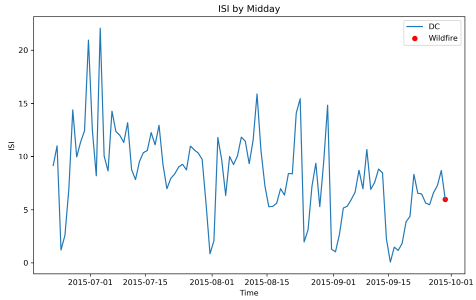


(b) DC - Calculated value

Figure 6.5: Comparison of ISI calculated values and Copernicus

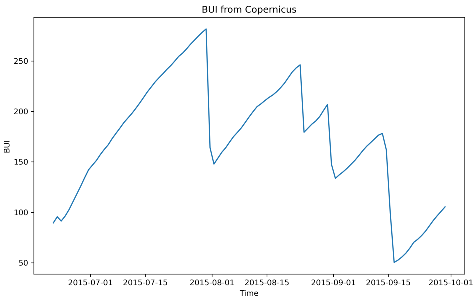


(a) ISI - Copernicus

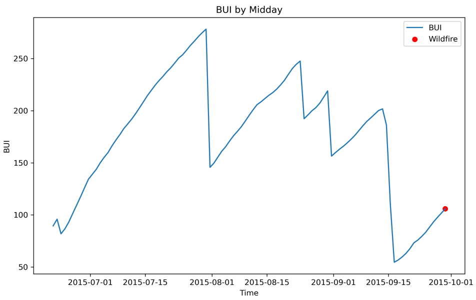


(b) ISI - Calculated value

Figure 6.6: Comparison of BUI calculated values and Copernicus



(a) BUI - Copernicus

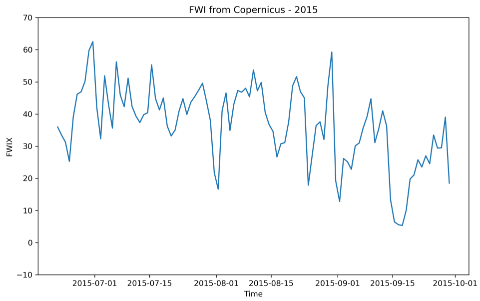


(b) BUI - Calculated value

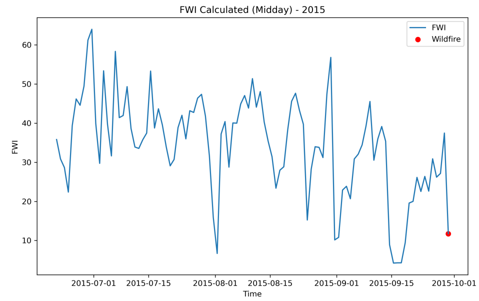
6.3 Comparison of Copernicus and Simulated FWI

6.3.1 Fogo de 2015

Figure 6.7: Comparison of FWI calculated values and Copernicus at midday

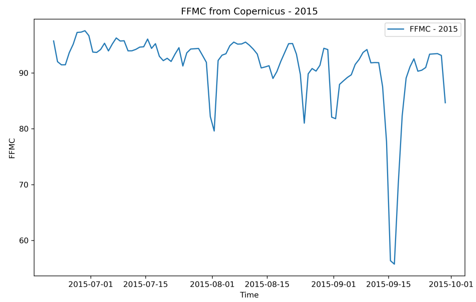


(a) FWI - Copernicus

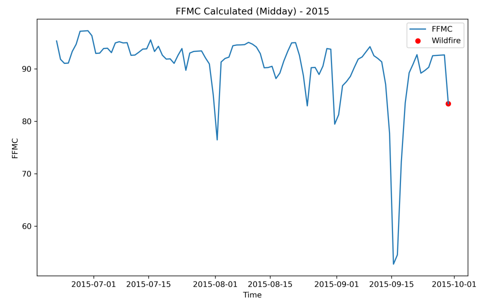


(b) FWI - Calculated value

Figure 6.8: Comparison of FFMC calculated values and Copernicus at midday

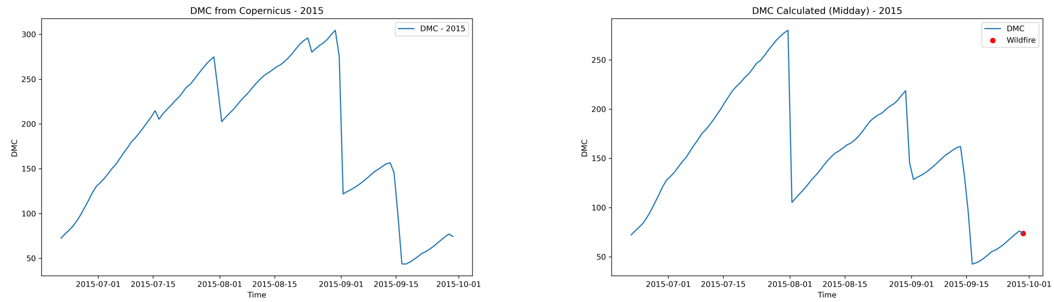


(a) FFMC - Copernicus



(b) FFMC - Calculated value

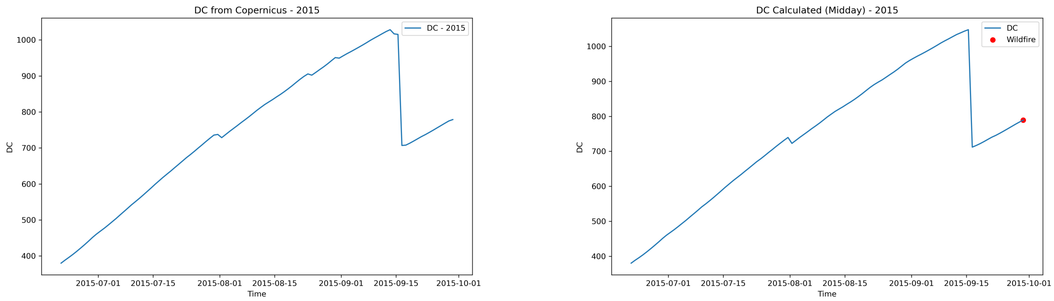
Figure 6.9: Comparison of DMC calculated values and Copernicus at midday



(a) DMC - Copernicus

(b) DMC - Calculated value

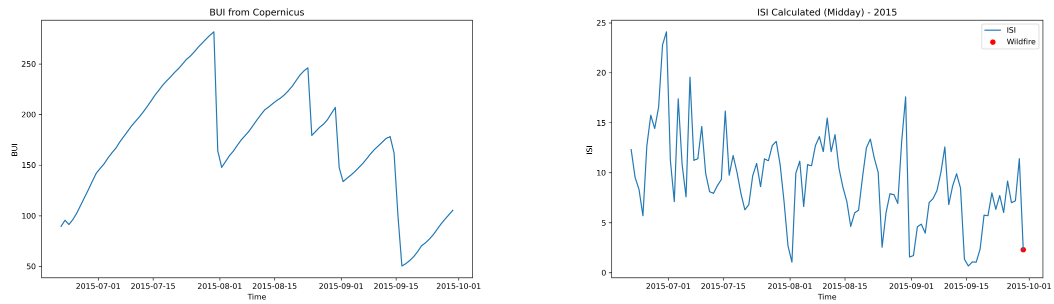
Figure 6.10: Comparison of DC calculated values and Copernicus at midday



(a) DC - Copernicus

(b) DC - Calculated value

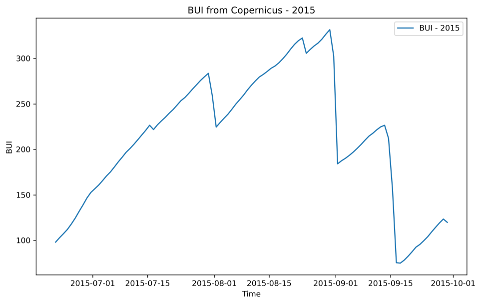
Figure 6.11: Comparison of ISI calculated values and Copernicus at midday



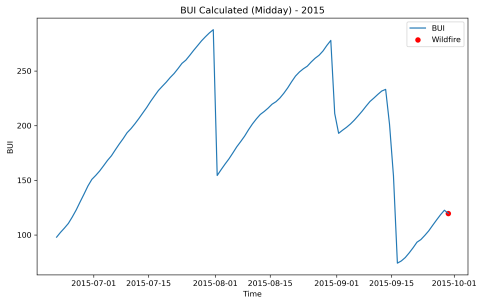
(a) ISI - Copernicus

(b) ISI - Calculated value

Figure 6.12: Comparison of BUI calculated values and Copernicus at midday



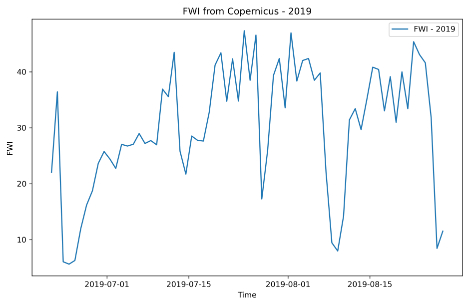
(a) BUI - Copernicus



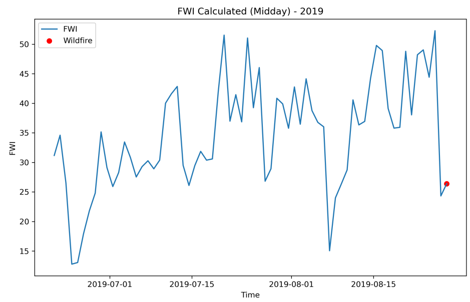
(b) BUI - Calculated value

6.3.2 Fogo de 2019

Figure 6.13: Comparison of FWI calculated values and Copernicus at midday - 2019



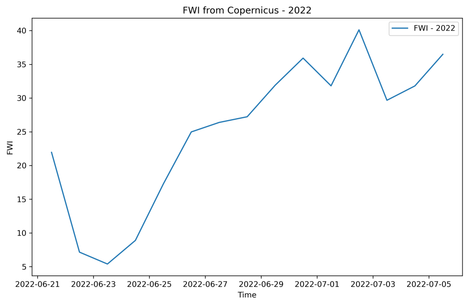
(a) FWI - Copernicus



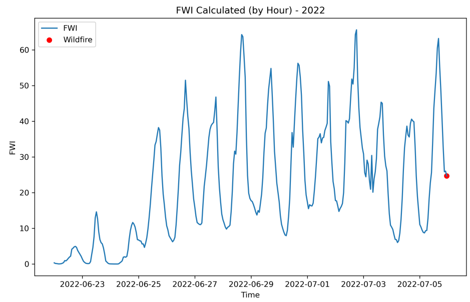
(b) FWI - Calculated value

6.3.3 Fogo de 2022

Figure 6.14: Comparison of FWI calculated values and Copernicus at midday - 2022



(a) FWI - Copernicus



(b) FWI - Calculated value

6.4 Hourly FWI variables

Figure 6.15: Calculated hourly FWI value for 2015, 2019, and 2022

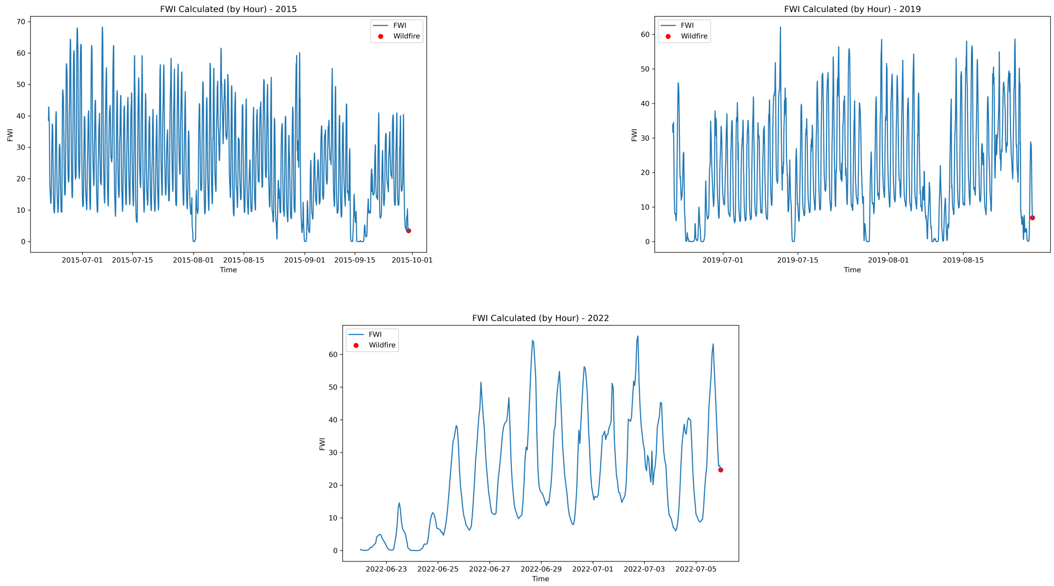


Figure 6.16: Calculated hourly FFMC value for 2015, 2019, and 2022

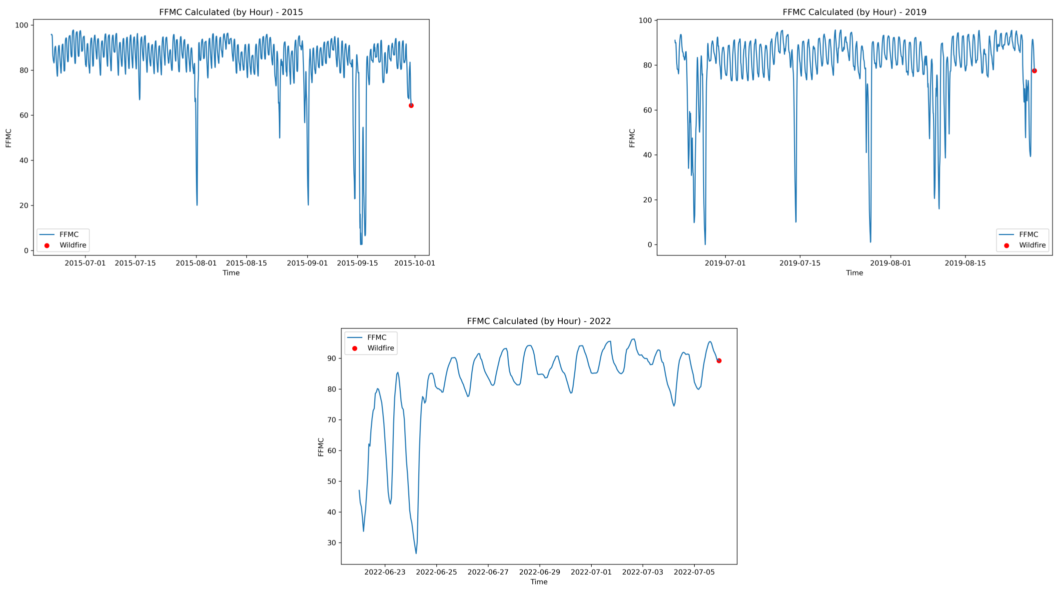


Figure 6.17: Calculated hourly DMC value for 2015, 2019, and 2022

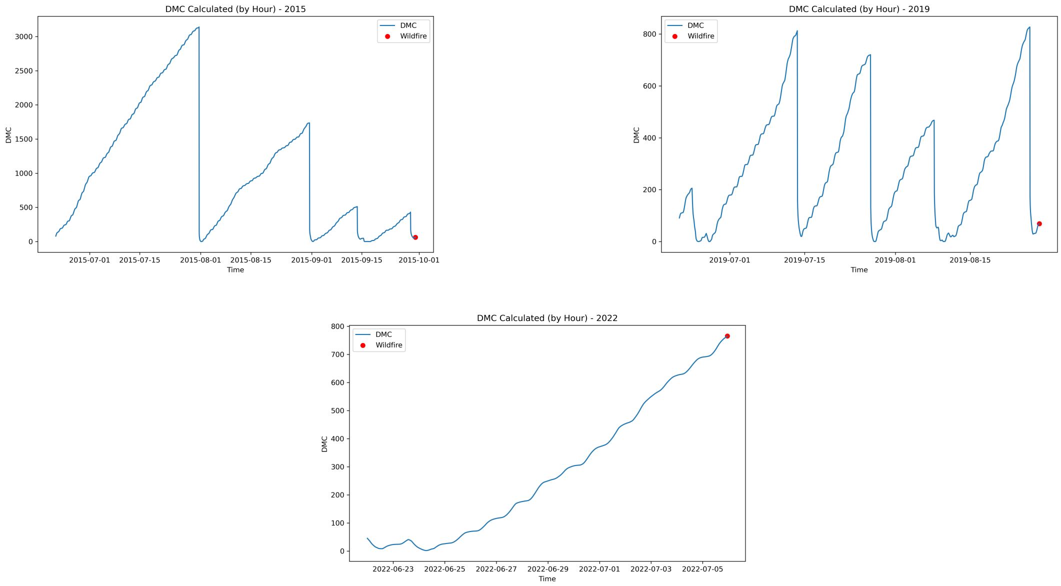


Figure 6.18: Calculated hourly DC value for 2015, 2019, and 2022

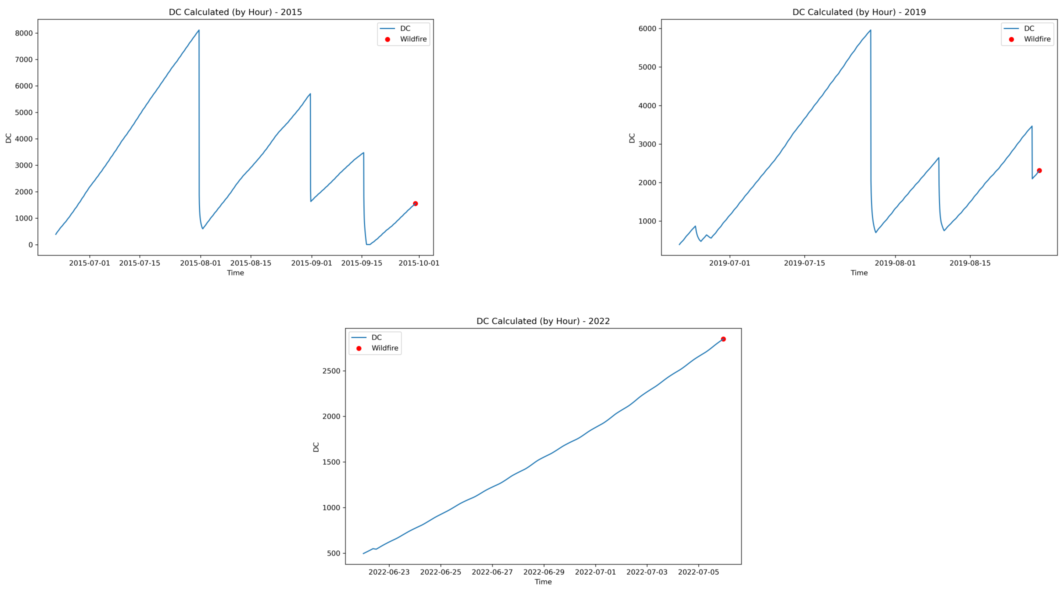


Figure 6.19: Calculated hourly ISI value for 2015, 2019, and 2022

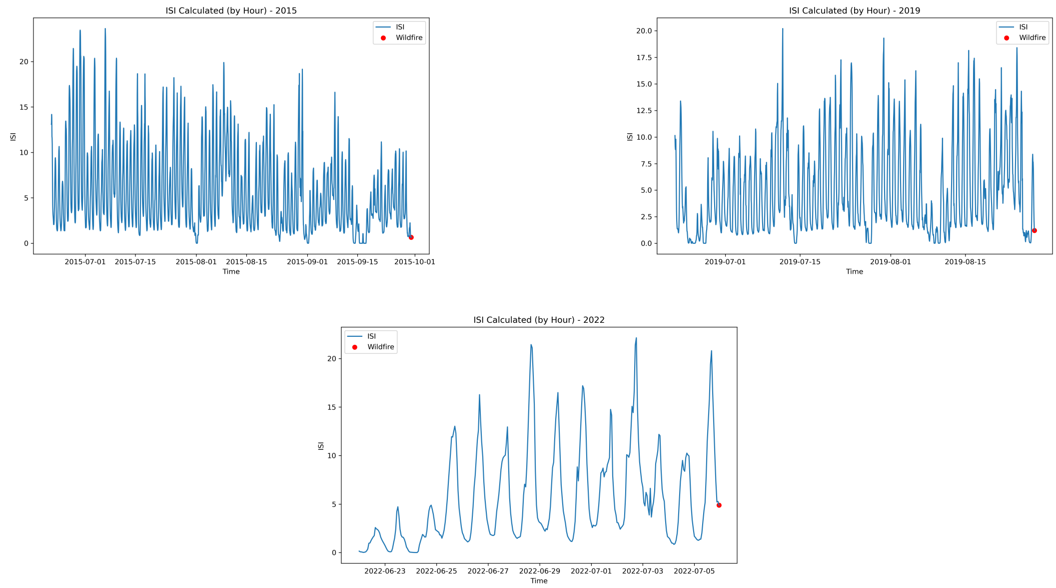
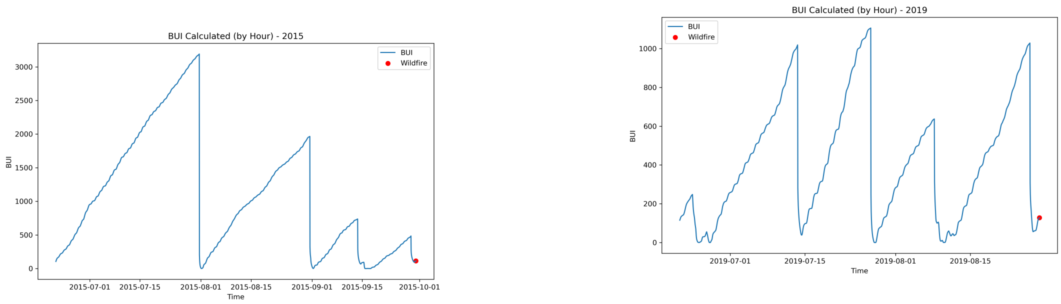
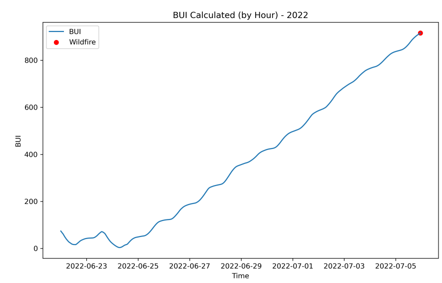


Figure 6.20: Calculated hourly BUI value for 2015, 2019, and 2022

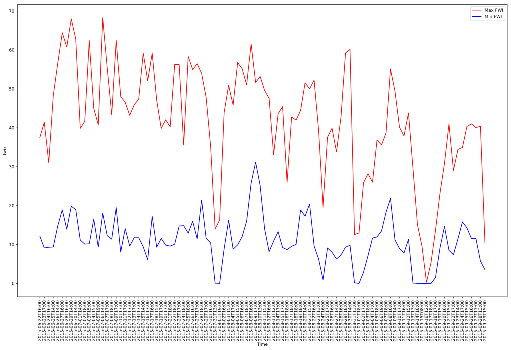


(a) Caption for image 2

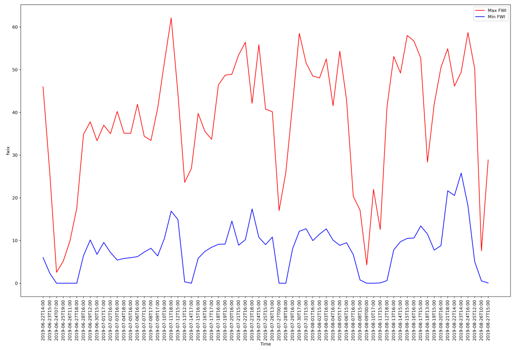


6.5 Evolution of maximum and minimum daily values of FWI variables

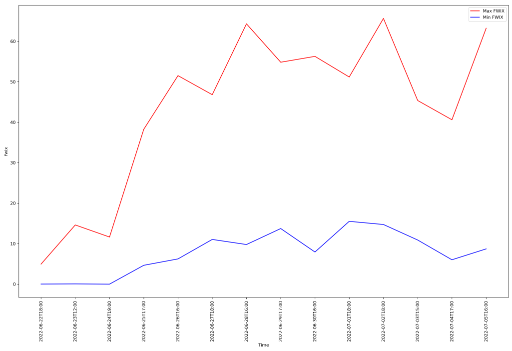
Figure 6.21: Daily max and min FWI values



(a) 2015

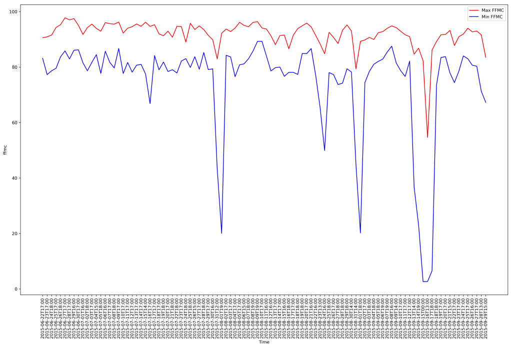


(b) 2019

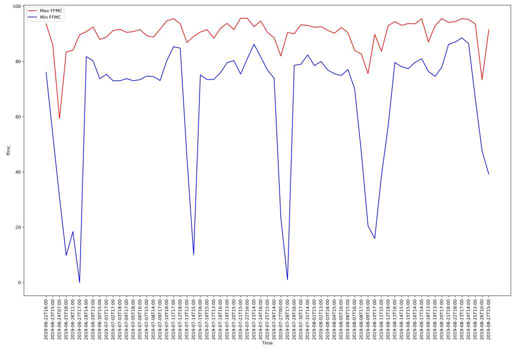


(c) 2022

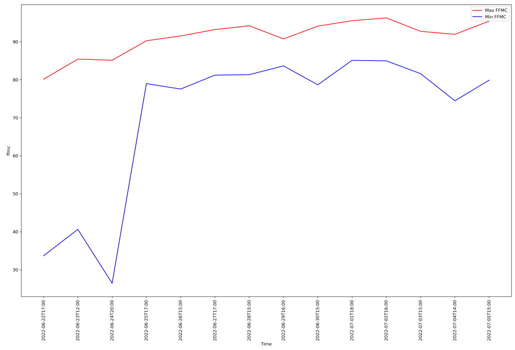
Figure 6.22: Daily max and min FFMC values



(a) 2015

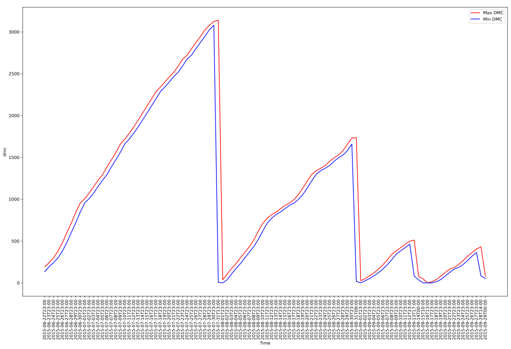


(b) 2019

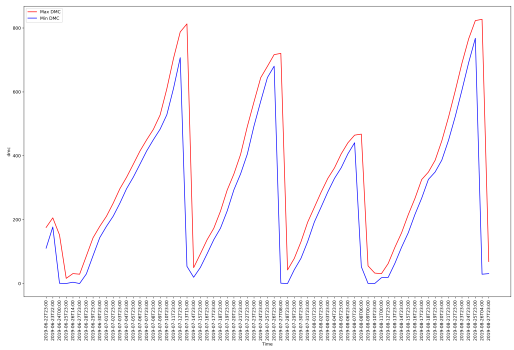


(c) 2022

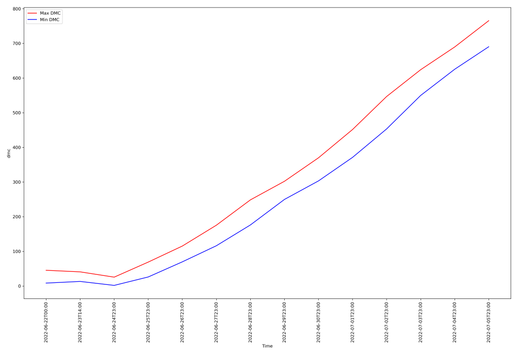
Figure 6.23: Daily max and min DMC values



(a) 2015

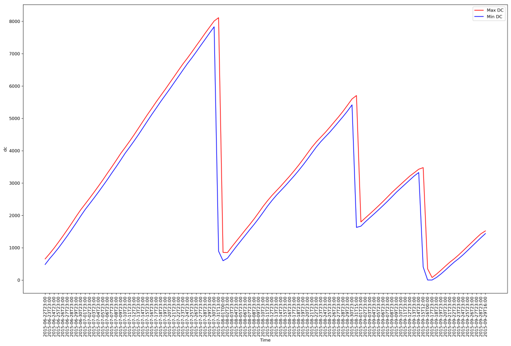


(b) 2019

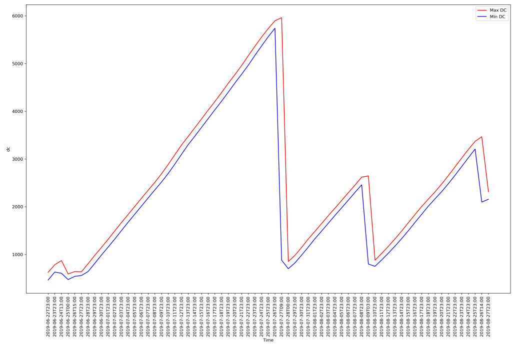


(c) 2022

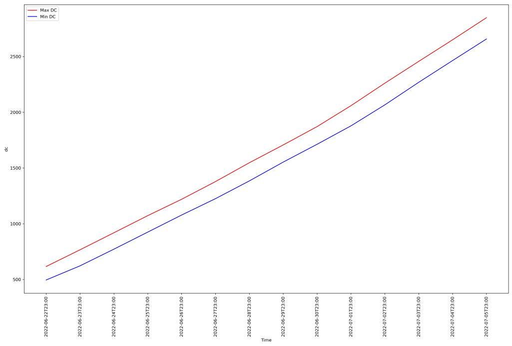
Figure 6.24: Daily max and min DC values



(a) 2015

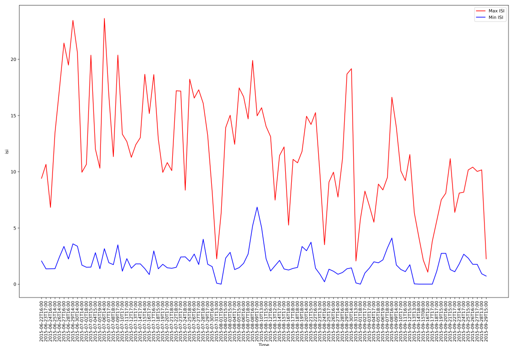


(b) 2019

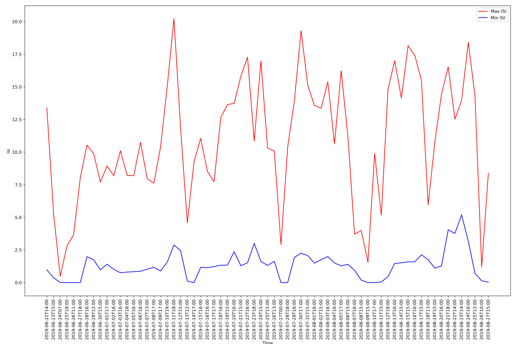


(c) 2022

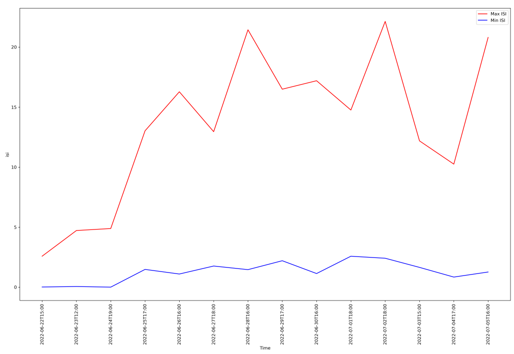
Figure 6.25: Daily max and min ISI values



(a) 2015

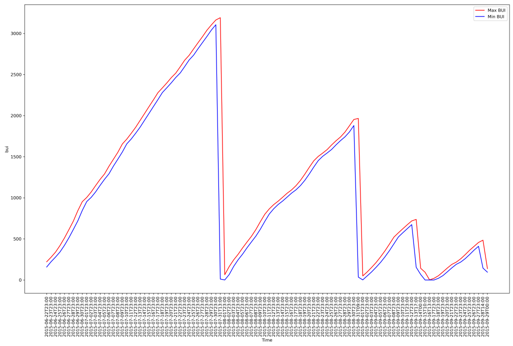


(b) 2019

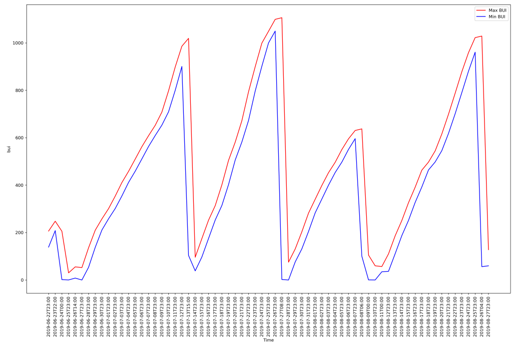


(c) 2022

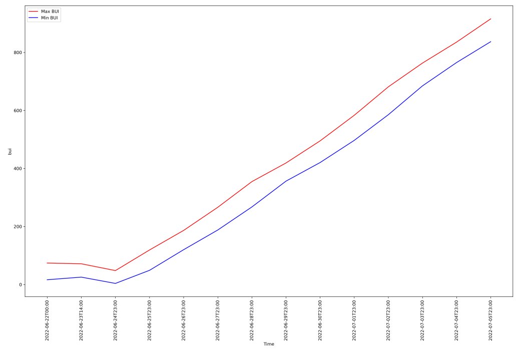
Figure 6.26: Daily max and min BUI values



(a) 2015



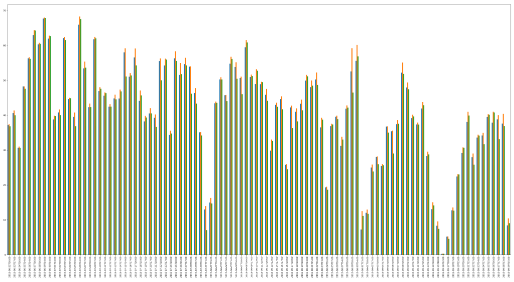
(b) 2019



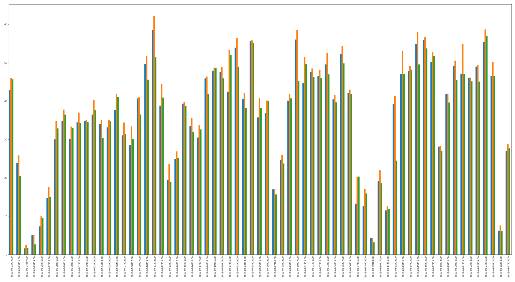
(c) 2022

6.6 Before, after and daily maximum value

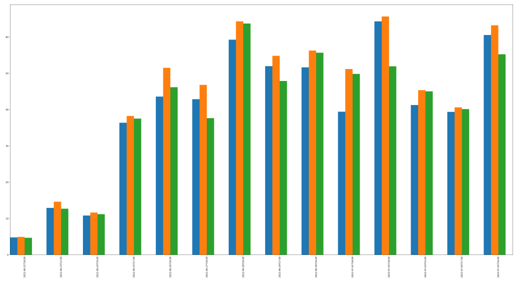
Figure 6.27: Before, after and daily FWI maximum value



(a) 2015

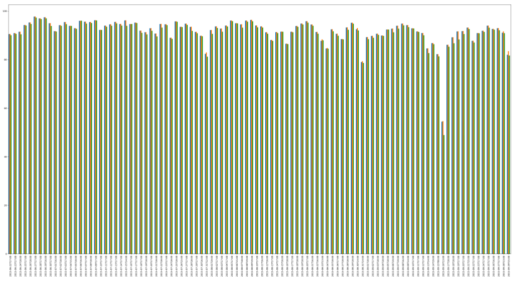


(b) 2019

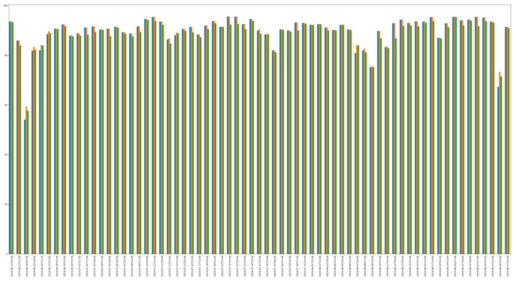


(c) 2022

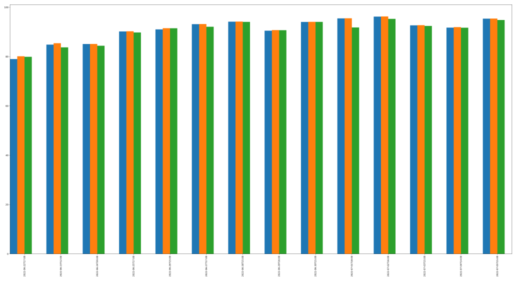
Figure 6.28: Before, after and daily FFMC maximum value



(a) 2015

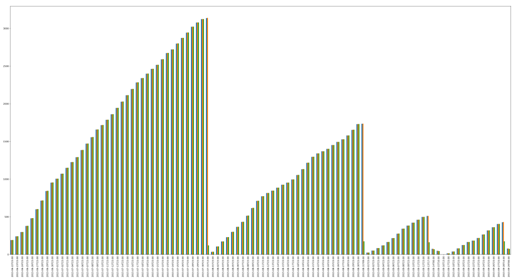


(b) 2019

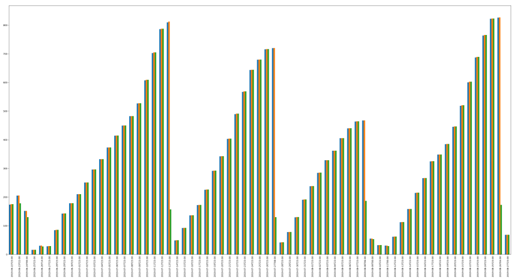


(c) 2022

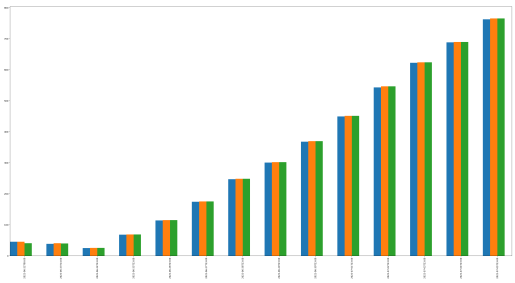
Figure 6.29: Before, after and daily DMC maximum value



(a) 2015

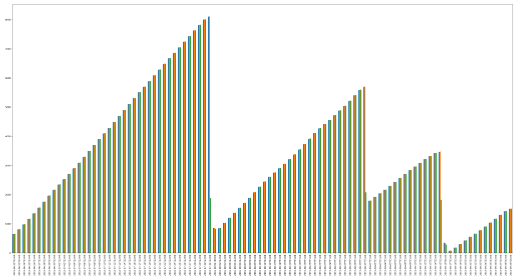


(b) 2019

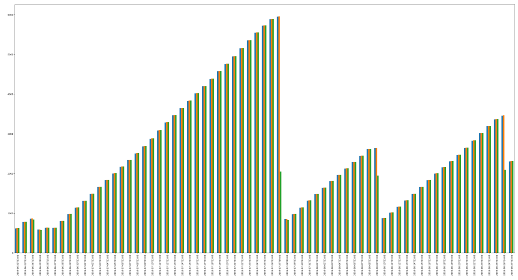


(c) 2022

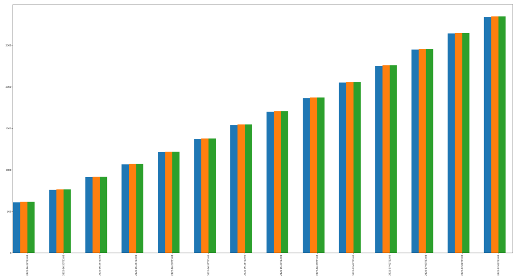
Figure 6.30: Before, after and daily DC maximum value



(a) 2015

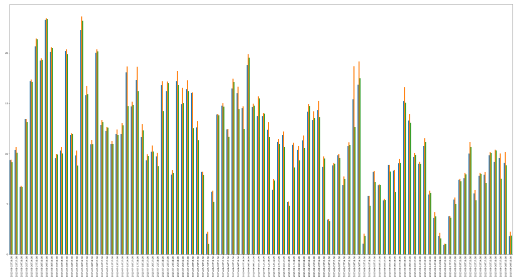


(b) 2019

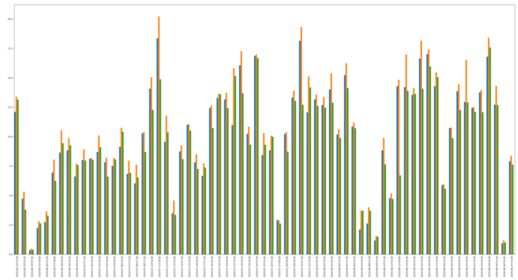


(c) 2022

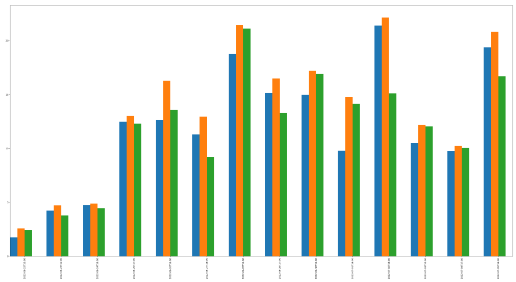
Figure 6.31: Before, after and daily ISI maximum value



(a) 2015

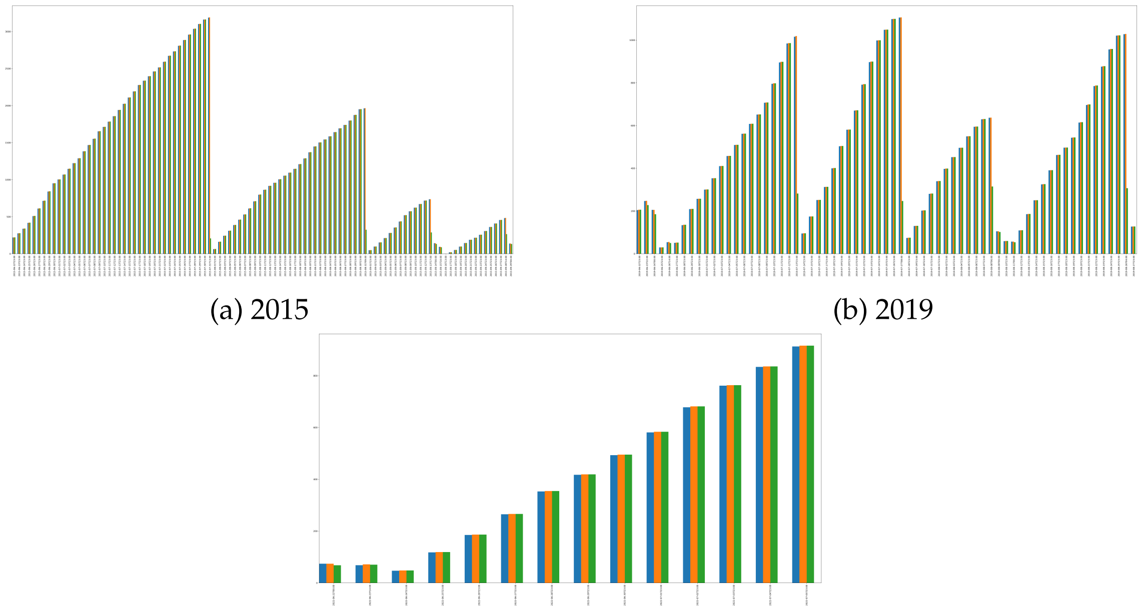


(b) 2019



(c) 2022

Figure 6.32: Before, after and daily BUI maximum value



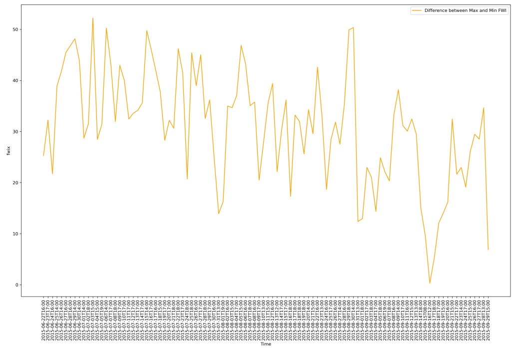
(a) 2015

(b) 2019

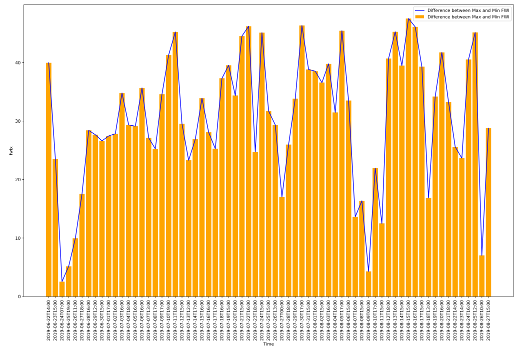
(c) 2022

6.7 Difference between the daily maximum and minimum values of the FWI variables

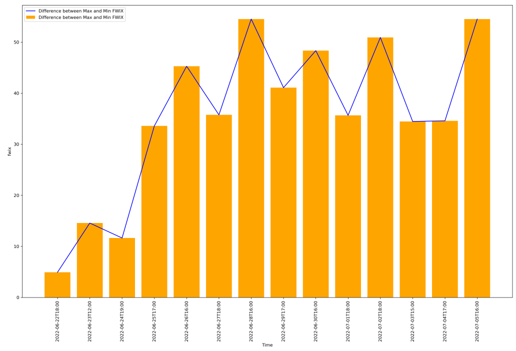
Figure 6.33: Daily difference of max and min FWI values



(a) 2015

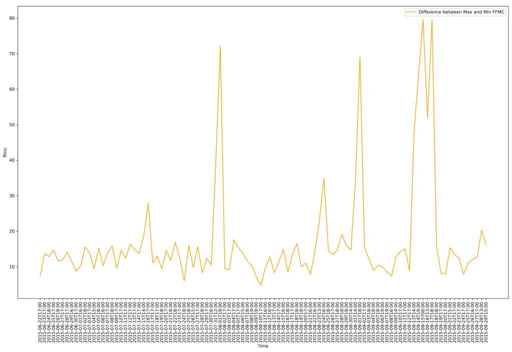


(b) 2019

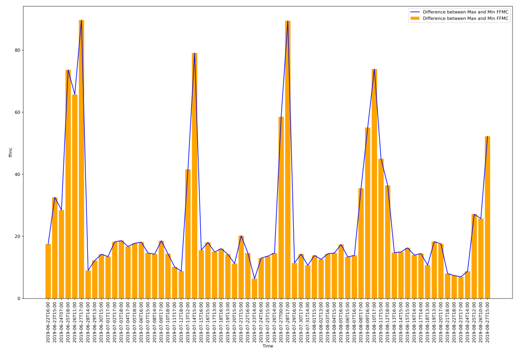


(c) 2022

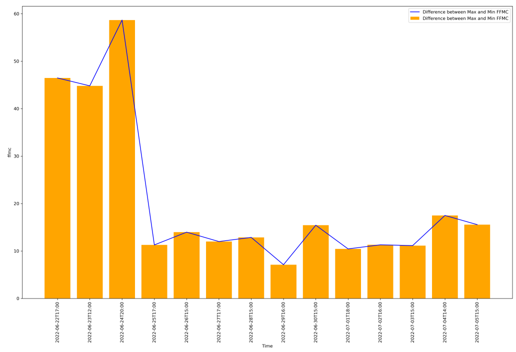
Figure 6.34: Daily difference of max and min FFMC values



(a) 2015

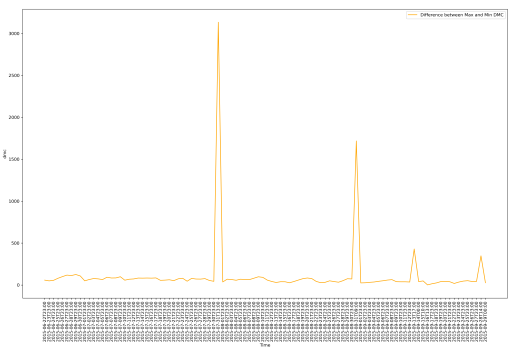


(b) 2019

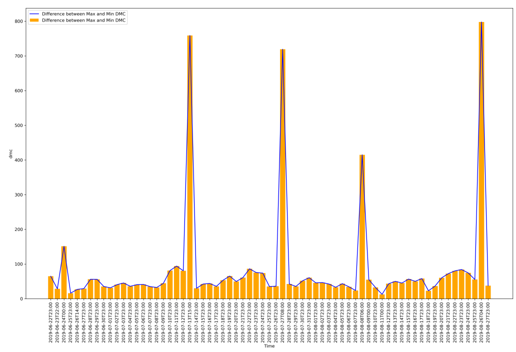


(c) 2022

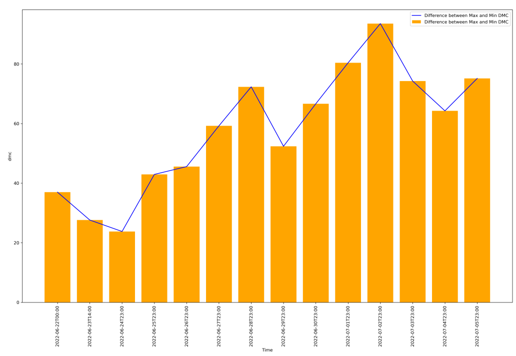
Figure 6.35: Daily difference of max and min DMC values



(a) 2015

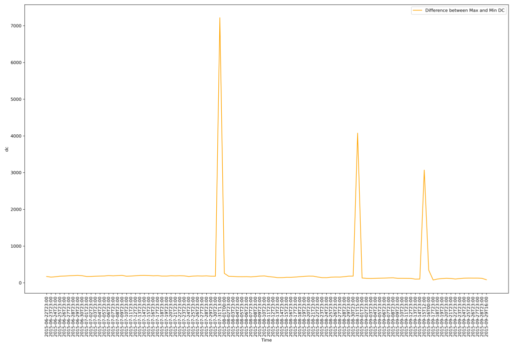


(b) 2019

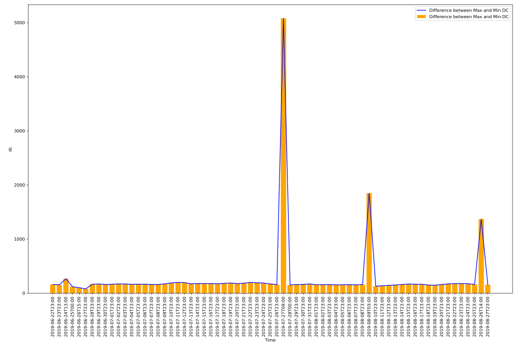


(c) 2022

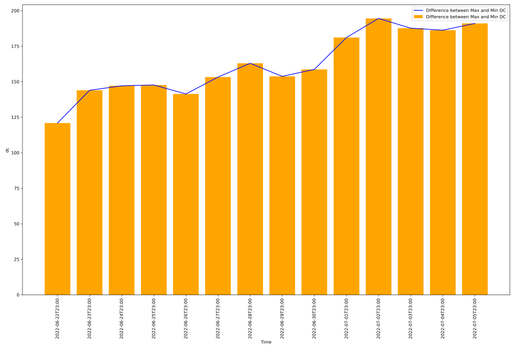
Figure 6.36: Daily difference of max and min DC values



(a) 2015

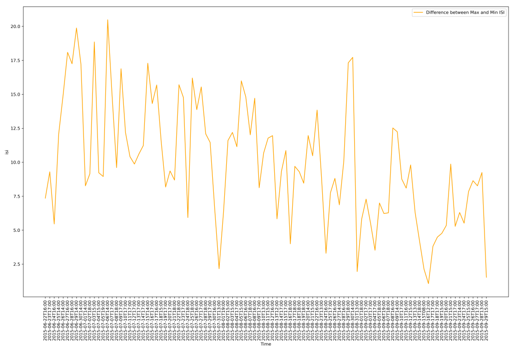


(b) 2019

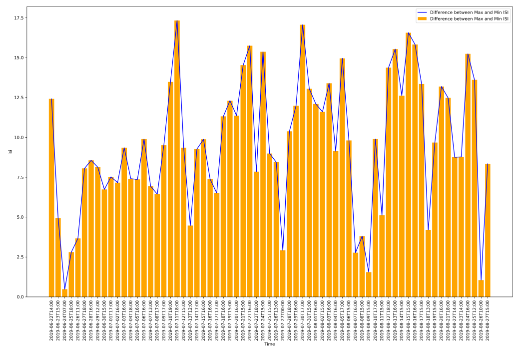


(c) 2022

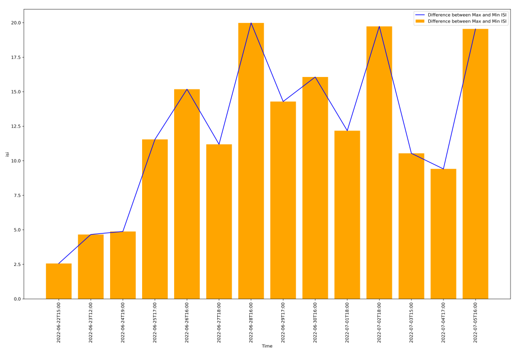
Figure 6.37: Daily difference of max and min ISI values



(a) 2015

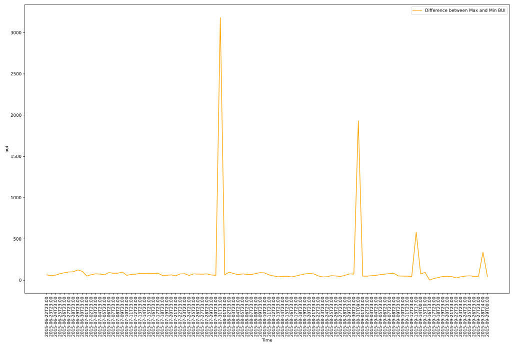


(b) 2019

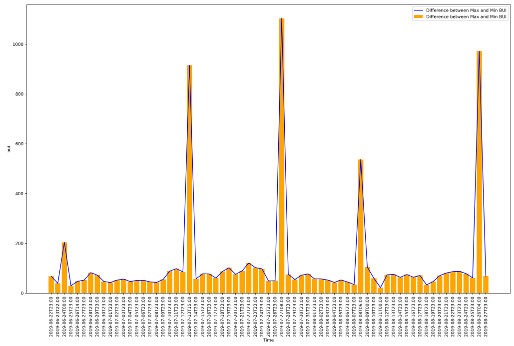


(c) 2022

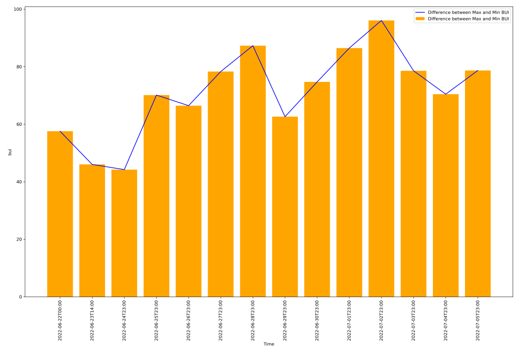
Figure 6.38: Daily difference of max and min BUI values



(a) 2015



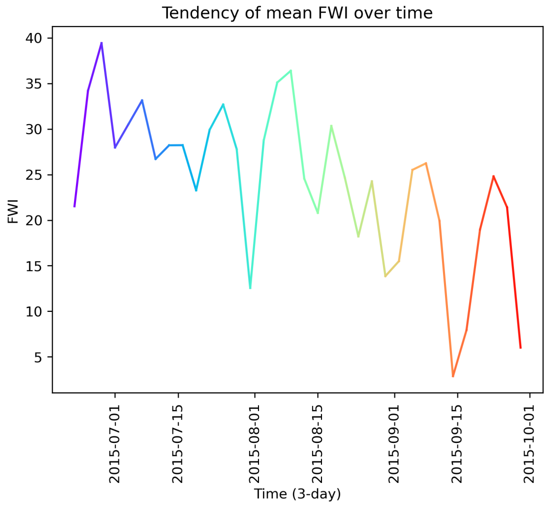
(b) 2019



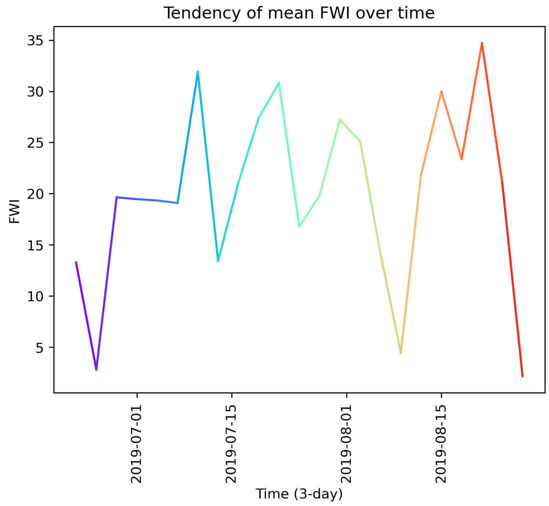
(c) 2022

6.8 3-day time frame mean tendency graphs of FWI variables

Figure 6.39: FWI mean tendency graph

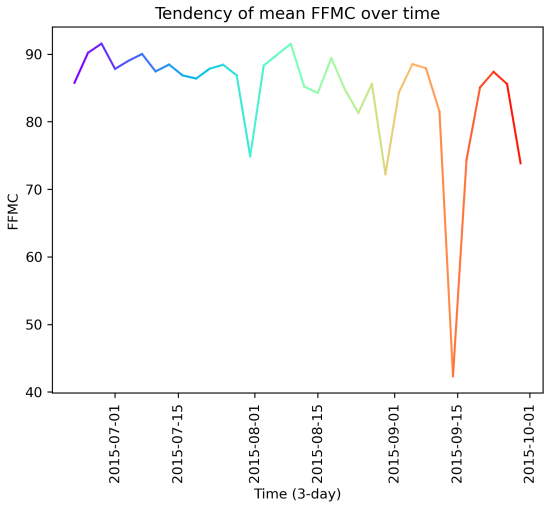


(a) 2015

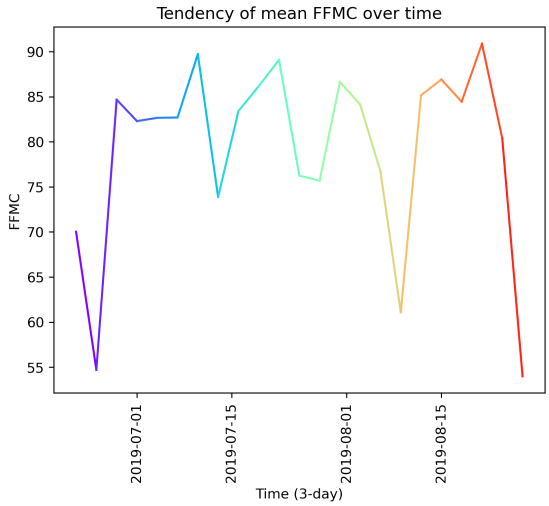


(b) 2019

Figure 6.40: FFMC mean tendency graph

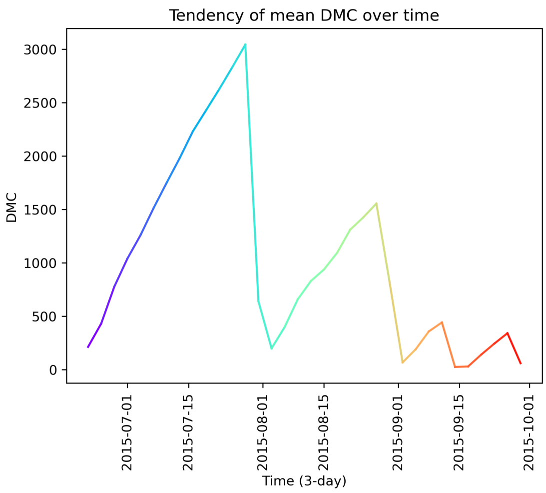


(a) 2015

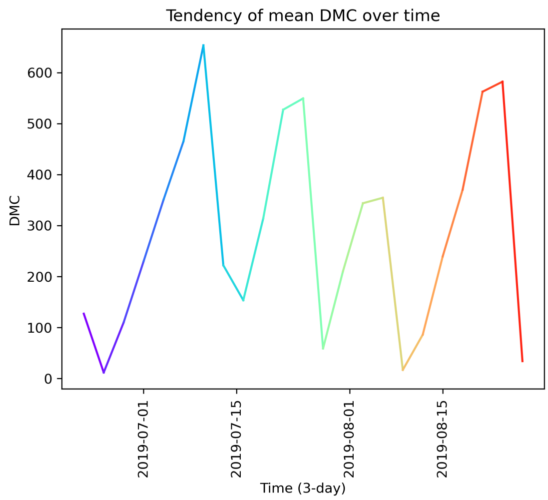


(b) 2019

Figure 6.41: DMC mean tendency graph

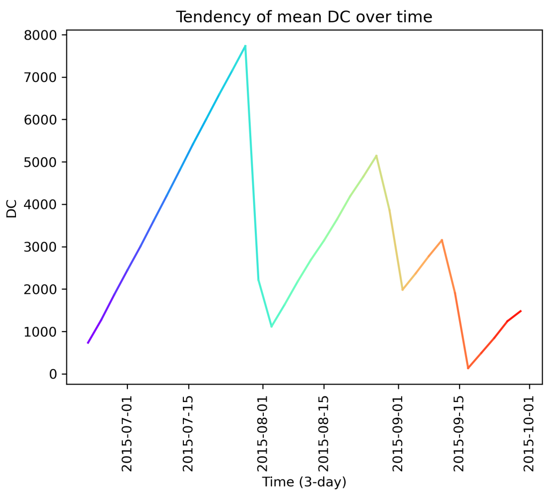


(a) 2015

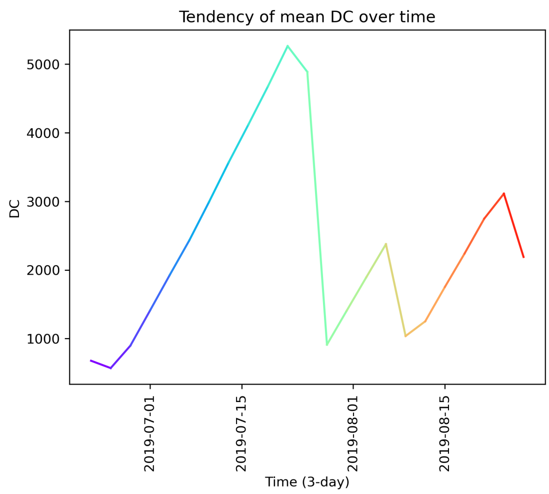


(b) 2019

Figure 6.42: DC mean tendency graph

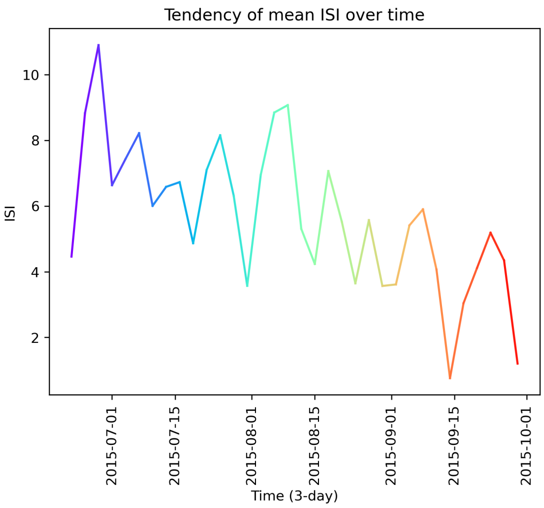


(a) 2015

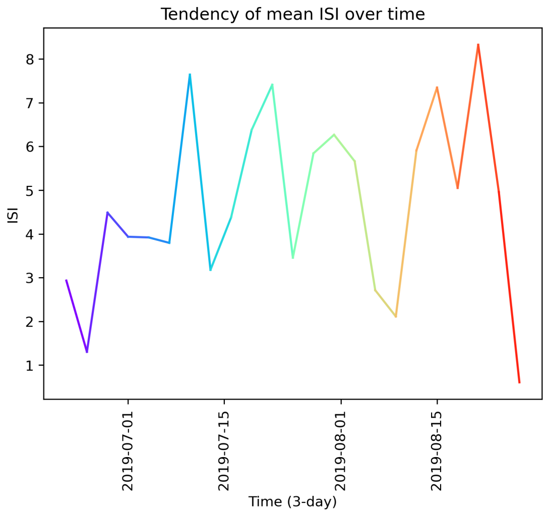


(b) 2019

Figure 6.43: ISI mean tendency graph

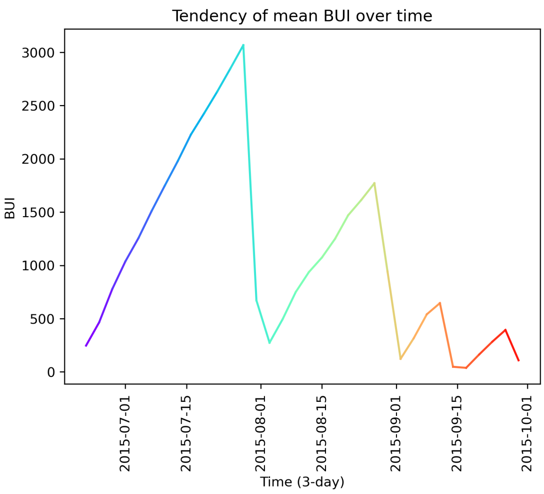


(a) 2015

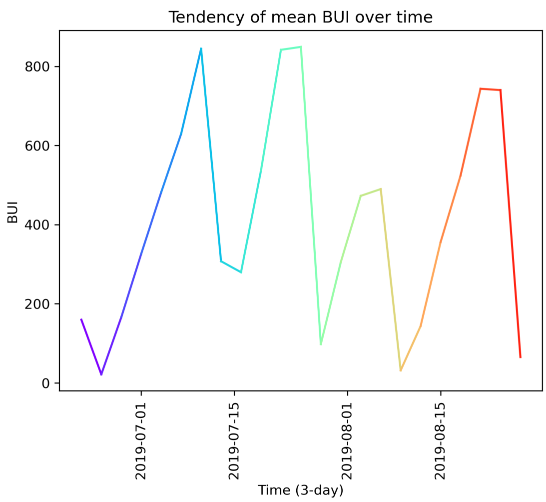


(b) 2019

Figure 6.44: BUI mean tendency graph



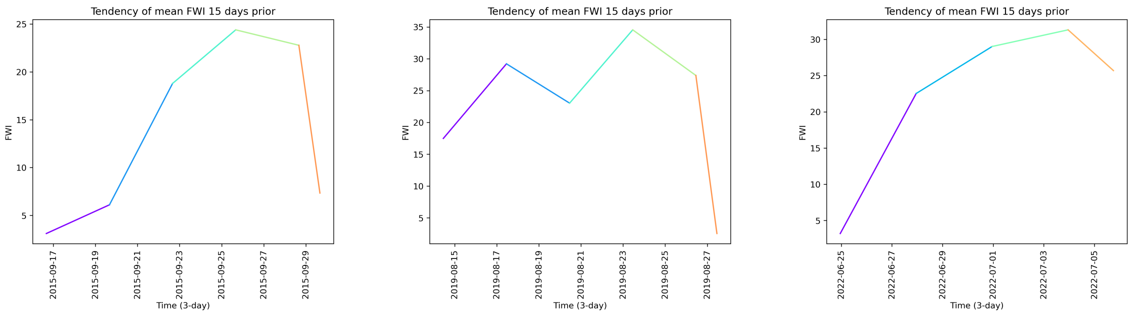
(a) 2015



(b) 2019

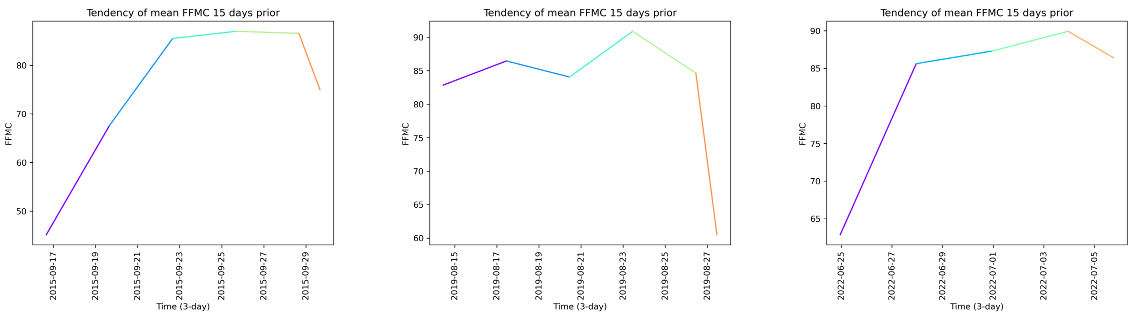
6.9 Comparison of mean FWI variables 15 days prior to the wildfire

Figure 6.45: FWI values 15 days prior to wildfire



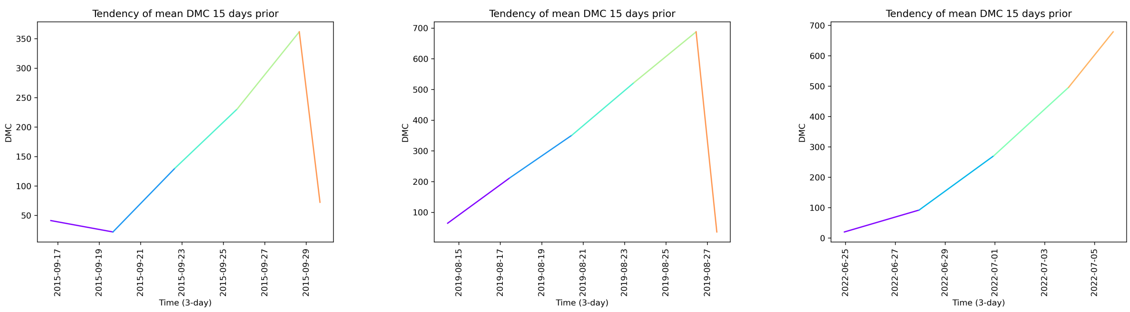
(a) 2015 (b) 2019 (c) 2022

Figure 6.46: FFMC values 15 days prior to wildfire



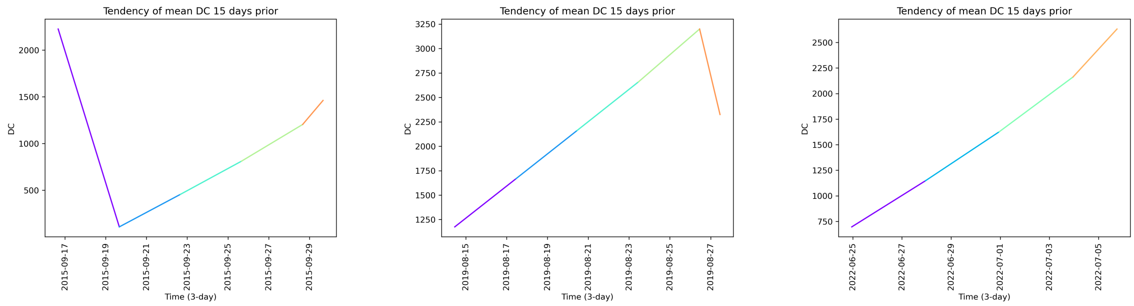
(a) 2015 (b) 2019 (c) 2022

Figure 6.47: DMC values 15 days prior to wildfire



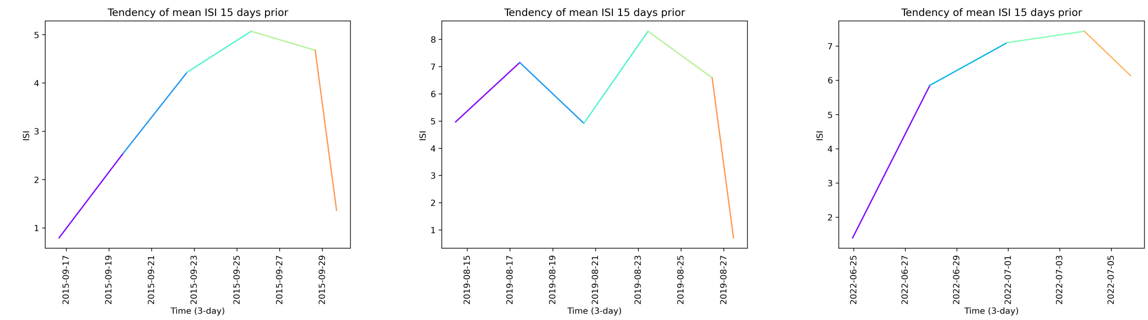
(a) 2015 (b) 2019 (c) 2022

Figure 6.48: DC values 15 days prior to wildfire



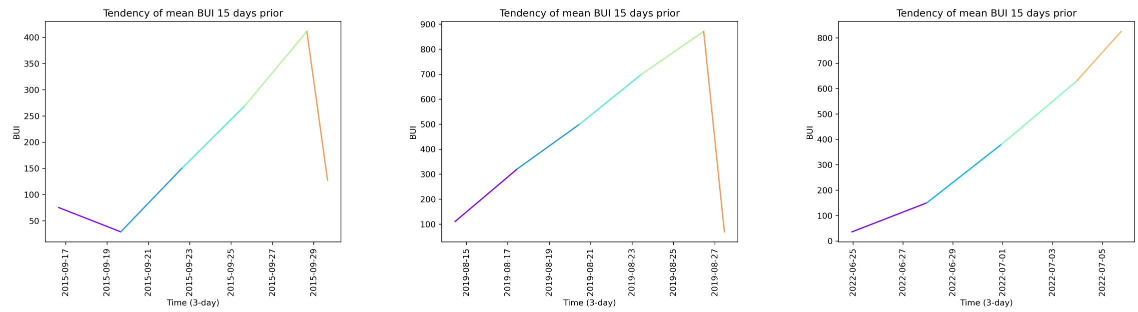
(a) 2015 (b) 2019 (c) 2022

Figure 6.49: IS15daysI values 15 days prior to wildfire



(a) 2015 (b) 2019 (c) 2022

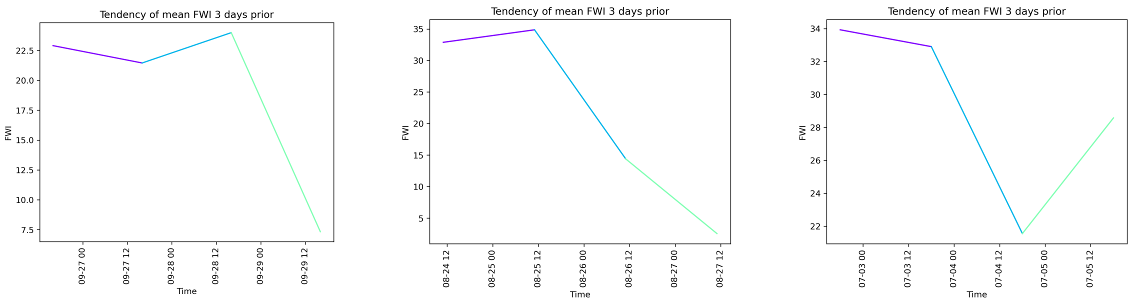
Figure 6.50: BUI values 15 days prior to wildfire



(a) 2015 (b) 2019 (c) 2022

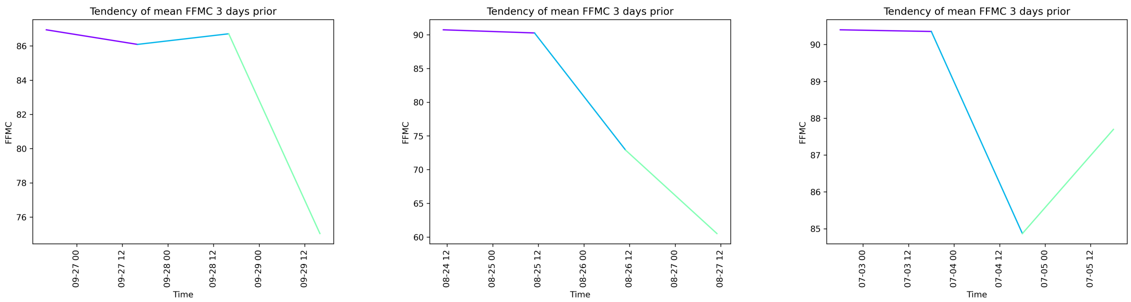
6.10 Comparison of mean FWI variables 3 days prior to the wildfire

Figure 6.51: FWI values 3 days prior to wildfire



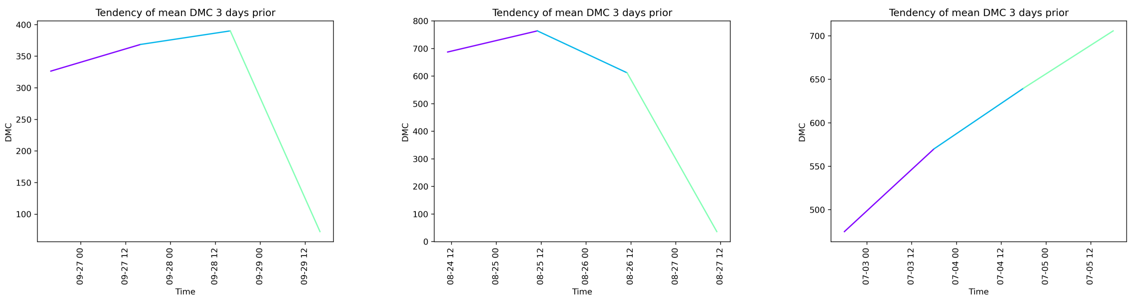
(a) 2015 (b) 2019 (c) 2022

Figure 6.52: FFMC values 3 days prior to wildfire



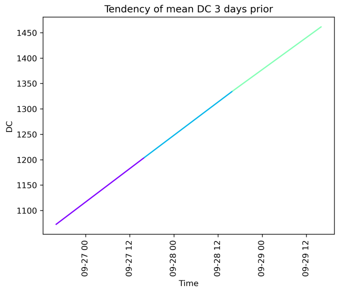
(a) 2015 (b) 2019 (c) 2022

Figure 6.53: DMC values 3 days prior to wildfire

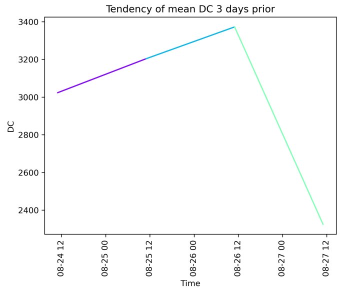


(a) 2015 (b) 2019 (c) 2022

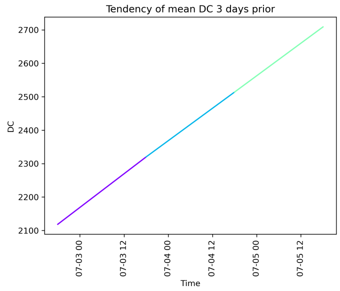
Figure 6.54: DC values 3 days prior to wildfire



(a) 2015

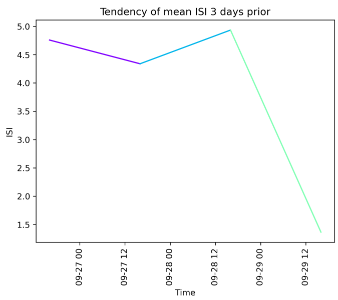


(b) 2019

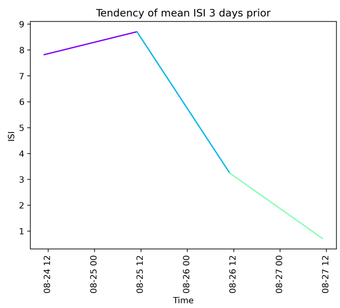


(c) 2022

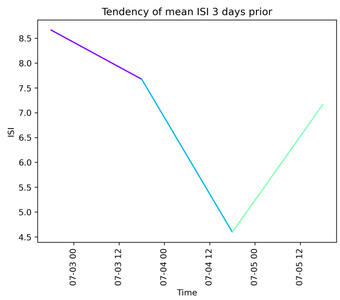
Figure 6.55: ISI values 3 days prior to wildfire



(a) 2015

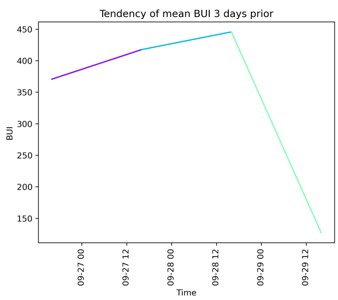


(b) 2019

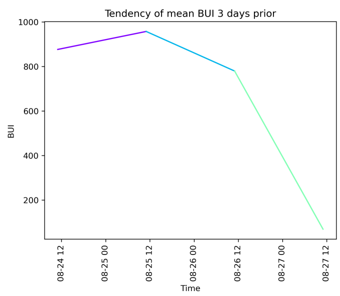


(c) 2022

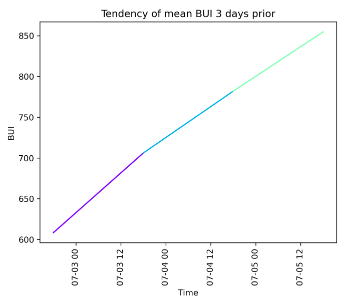
Figure 6.56: BUI values 3 days prior to wildfire



(a) 2015



(b) 2019



(c) 2022

Chapter 7

Conclusion

As the average temperature rises, so does the number of fires. It is therefore crucial to develop solutions that can help mitigate and control fires. Decision support systems play a central role in dealing with forest fires due to their early warning capacity and real-world impact. They improve the response to forest fires and ultimately help to protect the forests and the communities that depend on them.

This document outlines a system proposal comprised of multiple modules capable of analysing, evaluating, and collecting multiple data sources. Creating a data pipeline, classifying the severity of a wildfire occurrence, aggregating and fusing data, and visualising the findings.

Multiple topics regarding forest fires are analysed. There is a study of forest fire management, decision support systems, spatial and temporal prediction, and influencing factors. These topics are inherently connected to the understanding of wildfires.

Understanding what has been done before is also crucial to the underlying problem of wildfires. Therefore, multiple studies are shown and divided into occurrence, susceptibility, and risk prediction. In the literature presented, fire occurrence and susceptibility were calculated using only one machine learning model, while risk followed a hybrid approach and could contain several models.

Methodologies for aggregation, fusion, enhancement, and data visualisation tools were also presented. This was followed by an analysis and definition of the problem, as well as a risk analysis and evaluation of success.

References

Abdel-Basset, M. and Shawky, L. A. (2019). Flower pollination algorithm: a comprehensive review. *Artificial Intelligence Review*, 52(4):2533–2557.

Abid, F. (2021). A survey of machine learning algorithms based forest fires prediction and detection systems. *Fire Technology*, 57(2):559–590.

Abid, F. and Izeboudjen, N. (2020). Predicting forest fire in algeria using data mining techniques: Case study of the decision tree algorithm. In *Advanced Intelligent Systems for Sustainable Development (AI2SD'2019) Volume 4-Advanced Intelligent Systems for Applied Computing Sciences*, pages 363–370. Springer.

Al-Zebda, A. K., Al-Kahlout, M. M., Ghaly, A. M. A., and Mudawah, D. Z. (2021). Predicting forest fires using meteorological data: An ann approach.

Alkhatib, A. A. A. (2014). A review on forest fire detection techniques. *International Journal of Distributed Sensor Networks*, 10(3):597368.

Aminizade, M., Nekouei Moghaddam, M., Birami Jam, M., Shamsi, M., Majidi, N., Amanat, N., and Hoseini, S. H. (2017). The role of volunteer citizens in response to accidents and disasters. *Health Quarterly in Accidents and Disasters*, 2(3):107–124.

Arif, M., Alghamdi, K., Sahel, S., Alosaimi, S., Alsahaft, M., Alharthi, M., and Arif, M. (2021). Role of machine learning algorithms in forest fire management: A literature review. *J. Robot. Autom*, 5:212–226.

Artés, T., Oom, D., De Rigo, D., Durrant, T. H., Maianti, P., Libertà, G., and San-Miguel-Ayanz, J. (2019). A global wildfire dataset for the analysis of fire regimes and fire behaviour. *Scientific data*, 6(1):296.

Awan, A. A. (2022). A complete guide to data augmentation. `https://www.datacamp.com/tutorial/complete-guide-data-augmentation`. Accessed: 12/12/2023.

Bento, C. (2021). Multilayer perceptron explained with a real-life example and python code: Sentiment analysis. *Towards Data Science*.

Bilogur, A. and Contributors (2016). *Geoplot*. `https://residentmario.github.io/geoplot/index.html`.

Bountzouklis, C., Fox, D. M., and Di Bernardino, E. (2023). Predicting wildfire ignition causes in southern france using explainable artificial intelligence (xai) methods. *Environmental Research Letters*, 18(4):044038.

Cai, S., Gallina, B., Nyström, D., and Seceleanu, C. (2019). Data aggregation processes: A survey, a taxonomy, and design guidelines. *Computing*, 101(10):1397–1429.

Canada, N. R. (2021). Fire management. Accessed: 02-12-2023.

Cardoso, A. and Estima, J. (2023). Desenvolvimento de sistema inteligente para localização e monitorização de incêndios florestais. `http://estagios.dei.uc.pt/cursos/mei/ano-letivo-2023-2024/propostas-atribuidas-2023-2024/?idestagio=5329`. Online; Accessed: 19-11-2023.

Central de Dados Github Repository (2017). Incendios data repository. `https://github.com/centraldedados/incendios`.

centraldedados (2017). Incêndios. `http://centraldedados.pt/incendios/`. Accessed: 2024-01-29.

Chatzichristos, C., Van Eyndhoven, S., Kofidis, E., and Van Huffel, S. (2022). Chapter 10 - coupled tensor decompositions for data fusion. In Liu, Y., editor, *Tensors for Data Processing*, pages 341–370. Academic Press.

Cilli, R., Elia, M., D'Este, M., Giannico, V., Amoroso, N., Lombardi, A., Pantaleo, E., Monaco, A., Sanesi, G., Tangaro, S., Bellotti, R., and Lafortezza, R. (2022). Explainable artificial intelligence (xai) detects wildfire occurrence in the mediterranean countries of southern europe. *Scientific Reports*, 12(1):16349.

CISUC (2023). Fireloc: Where's the fire? identification, positioning, and monitoring forest fires with crowdsourced data. Accessed: 2023-12-18.

Clickworker (2023). Data enhancement - content marketing glossary. Accessed: 2023-12-18.

Community, T. G. (2023). Geopy.

Contributors, P. (2021). *Pydeck*. `https://deckgl.readthedocs.io/en/latest/`.

Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (2019). Copernicus Climate Change Service, Climate Data Store, (2019): Fire danger indices historical data from the Copernicus Emergency Management Service. `https://cds.climate.copernicus.eu/cdsapp#!/dataset/indices-era5-single-levels?tab=overview`. Accessed on 12-02-2024.

Cortez, P. and Morais, A. (2008). Forest Fires. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5D88D.

da Conservação da Natureza e das Florestas, I. I. (2023). 8.º relatório provisório de incêndios rurais 2023.

Dong, H., Wu, H., Sun, P., and Ding, Y. (2022). Wildfire prediction model based on spatial and temporal characteristics: A case study of a wildfire in portugal's montesinho natural park. *Sustainability*, 14(16).

Economic, W. (2023). What's driving the increase in forest fires? Online; Accessed: 18-11-2023.

Europeia, C. (2021). Relatório da comissão sobre os incêndios florestais: os efeitos das alterações climáticas são cada vez mais evidentes.

Frąckiewicz, M. (2023). From smoke signals to ai: The evolution of forest fire prediction techniques.

Gaikwad, A., Bhuta, N., Jadhav, T., Jangale, P., and Shinde, S. (2022). A review on forest fire prediction techniques. In *2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA*, pages 1–5.

Gao, C., Lin, H., and Hu, H. (2023). Forest-fire-risk prediction based on random forest and backpropagation neural network of heihe area in heilongjiang province, china. *Forests*, 14(2).

GBIF.Org User (2024). Occurrence download.

GDAL/OGR contributors (2024). *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation.

Gillies, S. and Contributors (2011). *Fiona*. `https://fiona.readthedocs.io/`.

Gillies, S. et al. (2013). Rasterio: geospatial raster i/o for Python programmers.

Goldammer, J. G., Frost, P. G., Jurvélius, M., Kamminga, E. M., Kruger, T., Ing-Moody, S., and Pogeyed, M. L. (2001). Community participation in integrated forest fire management: experiences from africa, asia and europe. `https://www.fao.org/3/AC798E/ac798e09.htm#fnB6`. Accessed: 02-12-2023.

Google (2024). Google maps. `https://www.google.com/maps`.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N. (2023). Era5 hourly data on single levels from 1940 to present.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

ICNF (2024). Iii.12 zonas de risco natural territórios ardidos (Área ardida entre 1975 e 2023). `https://geocatalogo.icnf.pt/catalogo_tema5.html`. Accessed: 2024-01-30.

Inc., P. T. (2015). Collaborative data science.

Ivanchuk, N. (2023). Wildfire prevention: How to prevent forest fires. [Online; accessed 14-November-2023].

Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., and Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28(4):478–505.

Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L. J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, and Leblanc, F. (2020). geopandas/geopandas: v0.8.1.

Kalogirou, S. A. (2009). Chapter eleven - designing and modeling solar energy systems. In Kalogirou, S. A., editor, *Solar Energy Engineering*, pages 553–664. Academic Press, Boston.

Lacroix, Z. (2003). Chapter 4 - issues to address while designing a biological information system. In Lacroix, Z. and Critchlow, T., editors, *Bioinformatics*, The Morgan Kaufmann Series in Multimedia Information and Systems, pages 75–108. Morgan Kaufmann, Burlington.

Lang, Y. and Moeini-Meybodi, H. (2023). Un/desa policy brief 111: Wildfires – a growing concern for sustainable development. United Nations Department of Economic and Social Affairs.

Li, T., Cui, L., Liu, L., Chen, Y., Liu, H., Song, X., and Xu, Z. (2023). Advances in the study of global forest wildfires. *Journal of Soils and Sediments*, pages 1–15.

Lin, H., Liu, X., Wang, X., and Liu, Y. (2018). A fuzzy inference and big data analysis algorithm for the prediction of forest fire based on rechargeable wireless sensor networks. *Sustainable Computing: Informatics and Systems*, 18:101–111.

Mekala, R., Srinath, S., Gokul, S., Balavigneshwar, E., and Muralidharan, R. (2023). Forest fire probability prediction based on humidity and temperature. In *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, pages 1–5.

Mohammed, Z., Hanae, C., and Larbi, S. (2020). Comparative study on machine learning algorithms for early fire forest detection system using geodata. *International Journal of Electrical and Computer Engineering (IJECE)*, 10(5):5507–5513.

mParticle (2021). Data enrichment or machine learning. `https://www.mparticle.com/blog/data-enrichment-or-machine-learning/`. Accessed: 12/12/2023.

Muñoz Sabater, J. (2019). Era5-land hourly data from 2001 to present.

Naderpour, M., Rizeei, H. M., and Ramezani, F. (2021). Forest fire risk prediction: A spatial deep neural network-based framework. *Remote Sensing*, 13(13).

Nemtinov, V., Kalach, A., Egorov, S., and Nemtinova, Y. (2021). Information support for decision-making in emergency situations. *Journal of Physics: Conference Series*, 1902(1):012080.

Novo, A. L. A., Fariñas-Álvarez, N., Martínez-Sánchez, J., González-Jorge, H., Fernández-Alonso, J. M., and Lorenzo, H. (2020). Mapping forest fire risk - a case study in galicia (spain). *Remote. Sens.*, 12:3705.

of Encyclopaedia Britannica, T. E. (2023). Wildfire. [Online; Accessed: 15-11-2023].

Papčo, M., Rodríguez-Martínez, I., Fumanal-Idocin, J., Altalhi, A. H., and Bustince, H. (2021). A fusion method for multi-valued data. *Information Fusion*, 71:1–10.

Perry, M. T. and Contributors (2015). *Rasterstats*. `https://pythonhosted.org/rasterstats/`.

Pourghasemi, H. R., Gayen, A., Lasaponara, R., and Tiefenbacher, J. P. (2020). Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling. *Environmental Research*, 184:109321.

Qu, J. and Cui, X. (2020). Automatic machine learning framework for forest fire forecasting. *Journal of Physics: Conference Series*, 1651(1):012116.

Resco de Dios, V. and Nolan, R. H. (2021). Some challenges for forest fire risk predictions in the 21st century. *Forests*, 12(4).

Sadatrazavi, A., Motlagh, M. S., Noorpoor, A., and Ehsani, A. H. (2022). Predicting wildfires occurrences using meteorological parameters. *International Journal of Environmental Research*, 16(6):106.

Sayad, Y. O., Mousannif, H., and Al Moatassime, H. (2019). Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal*, 104:130–146.

Schimanke, S., Ridal, M., Le Moigne, P., Berggren, L., Undén, P., Randriamampianina, R., Andrea, U., Bazile, E., Bertelsen, A., Brousseau, P., Dahlgren, P., Edvinsson, L., El Said, A., Glinton, M., Hopsch, S., Isaksson, L., Mladek, R., Olsson, E., Verrelle, A., and Wang, Z. (2021). Cerra sub-daily regional reanalysis data for europe on single levels from 1984 to present.

Schmitt, M. and Zhu, X. (2016). Data fusion and remote sensing – an ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine*, 4:6–23.

Service, U. F. (2023). Keetch-byram drought index (kbdi). `https://www.drought.gov/data-maps-tools/keetch-byram-drought-index`. Accessed: 04-12-2023.

Sharma, R., Rani, S., and Memon, I. (2020). A smart approach for fire prediction under uncertain conditions using machine learning. *Multimedia Tools and Applications*, 79(37):28155–28168.

Silva, C., Madeira, A., Cardoso, A., and Ribeiro, B. (2020). Fire and smoke recognition in crowdsourced images with yolo networks. In *Portuguese Conference on Pattern Recognition, RECPAD 2020*.

Silva, C. S. and Santos, M. J. P. (2012). Climate changes in the 20th and 21st centuries in mainland portugal. *Cadernos de Geografia*, 31:49–54.

SimonKettle (2017). Distance on a sphere: The haversine formula. `https://community.esri.com/t5/coordinate-reference-systems-blog/distance-on-a-sphere-the-haversine-formula/ba-p/902128`. Accessed: 2024-03-05.

Smith, D. H., Smith, D. H., Smith, D. H., Smith, D. H., Smith, D. H., Stebbins, R. A., Grotz, J., Grotz, J., Aguirre, B. E., Macias-Medrano, J., et al. (2016). Spontaneous volunteering in emergencies. *The Palgrave handbook of volunteering, civic participation, and nonprofit associations*, pages 311–329.

Society, N. G. (2023). The ecological benefits of fire. [Online; accessed 14-November-2023].

Story, R. (2013). *Folium*. `https://python-visualization.github.io/folium/`.

Surya, L. (2017). Risk analysis model that uses machine learning to predict the likelihood of a fire occurring at a given property. *International Journal of Creative Research Thoughts (IJCRT), ISSN*, pages 2320–2882.

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., and Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine*, 3(1):17.

Thakkar, R., Abhyankar, V., Reddy, P. D., and Prakash, S. (2022). Environmental fire hazard detection and prediction using random forest algorithm. In *2022 International Conference for Advancement in Technology (ICONAT)*, pages 1–4.

Tien Bui, D., Le, H. V., and Hoang, N.-D. (2018). Gis-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. *Ecological Informatics*, 48:104–116.

Uva, J. S., Onofre, R., Moreira, J., Faias, S. P., Barreiro, S., Santos, E., Capelo, J., Corte-Real, L., Martins, J., Ribeiro, J. R., Cancela, J., Rainha, M., Amaral, N., Santos, C., Perpétua, J., Pinho, J., Araújo, J. M., Reis, L., Canaveira, P., Paulino, J., Pina, A., Binev, Y., and Coelho, P. (2021). Forestry inventory 2015. `https://doi.org/10.15468/33hvm4`. Accessed via GBIF.org on 2024-04-04.

Viegas, D. X. (2018). Wildfires in portugal. *Fire Research*, 2(1).

Wang, S. S.-C., Qian, Y., Leung, L. R., and Zhang, Y. (2022). Interpreting machine learning prediction of fire emissions and comparison with firemip process-based models. *Atmospheric Chemistry and Physics*, 22(5):3445–3468.

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021.

Zacharakis, I. and Tsihrintzis, V. A. (2023). Environmental forest fire danger rating systems and indices around the globe: A review. *Land*, 12(1).

Zhang, G., Wang, M., and Liu, K. (2019). Forest fire susceptibility modeling using a convolutional neural network for yunnan province of china. *International Journal of Disaster Risk Science*, 10:386–403.

Zheng, Z., Gao, Y., Yang, Q., Zou, B., Xu, Y., Chen, Y., Yang, S., Wang, Y., and Wang, Z. (2020). Predicting forest fire risk based on mining rules with ant-miner algorithm in cloud-rich areas. *Ecological Indicators*, 118:106772.

Zippenfenig, P. (2023). Open-meteo.com weather api.