IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Prediction of Area Burned by Forest Fires (Using Machine Learning Algorithms)

[1]Javaid Mohi u din Bhat, [2]Ishfaq Ahmad Wani, [3]Mohammad Rayees Ganai

[1]Software Developer, [2]Student, [3]Software Developer

[1]University of Kashmir,

[1]Hazratbal Srinagar Kashmir

Abstract: Forest fires are a major environmental issue, creating economical and ecological damage while endangering human lives. Fast detection is a key element for controlling such phenomenon. To achieve this, one alternative is to use automatic tools based on local sensors, such as provided by meteorological stations. In effect, meteorological conditions (e.g. temperature, wind) are known to influence forest fires and several fire indexes, such as the forest Fire Weather Index (FWI), use such data. In this work, we explore a Data Mining (DM) approach to predict the burned area of forest fires. Two different DM techniques, e.g. Support Vector Machines (SVM) and one distinct feature selection setups (using spatial, temporal, FWI components and weather attributes), were tested on recent real-world data collected from the northeast region of Portugal. The best configuration uses a SVM and four meteorological inputs (i.e. temperature, relative humidity, rain and wind) and it is capable of predicting the burned area of small fires, which are more frequent. Such knowledge is particularly useful for improving firefighting resource management (e.g. prioritizing targets for air tankers and ground crews).

Keywords: Data Mining Application, Fire Science, Decision Tree, Support Vector Machines.

***Index Terms* - Component, formatting, style, styling, insert.**

I. Introduction

One major environmental concern is the occurrence of forest fires (also called wildfires), which affect forest preservation, create economical and ecological damage and cause human suffering. Such phenomenon is due to multiple causes (e.g. human negligence and lightnings) and despite an increasing of state expenses to control this disaster, each year millions of forest hectares (ha) are destroyed all around the world. In particular, Portugal is highly affected by forest fires . From 1980 to 2005, over 2.7 million ha of forest area (equivalent to the Albania land area) have been destroyed. The 2003 and 2005 fire seasons were especially dramatic, affecting 4.6% and 3.1% of the territory, with 21 and 18 human deaths.

Fast detection is a key element for a successful firefighting. Since traditional human surveillance is expensive and affected by subjective factors, there has been an emphasis to develop automatic solutions. These can be grouped into three major categories: satellite-based, infrared/smoke scanners and local sensors (e.g. meteorological). Satellites have acquisition costs, localization delays and the resolution is not adequate for all cases. Moreover, scanners have a high equipment and maintenance costs. Weather conditions, such as temperature and air humidity, are known to affect fire occurrence. Since automatic meteorological stations

are often available (e.g. Portugal has 162 official stations), such data can be collected in real-time, with low costs.
In the past, meteorological data has been incorporated into numerical indices, which are used for prevention (e.g. warning the public of a fire danger) and to support fire management decisions (e.g. level of readiness, prioritizing targets or

evaluating guidelines for safe firefighting). In particular, the Canadian forest Fire Weather Index (FWI) system was designed in the 1970s when computers were scarce, thus it required only simple calculations using look-up tables with readings from four meteorological observations (i.e. temperature, relative humidity, rain and wind) that could be manually collected in weather stations. Nevertheless, nowadays this index highly used not only in Canada but also in several countries around the world (e.g. Argentina or New Zealand). Even though Mediterranean climate differs from those in Canada, the FWI system was correlated with fire activity in southern Europe countries, including Portugal.

On the other hand, the interest in Data Mining (DM), also known as Knowledge Discovery in Databases (KDD), arose due to the advances of Information Technology, leading to an exponential growth of business, scientific and engineering databases. All this data holds valuable information, such as trends and patterns, which can be used to improve decision making. Yet, human experts are limited and may overlook important details. Moreover, classical statistical analysis breaks down when such vast and/or complex data is present. Hence, the alternative is to use automated DM tools to analyze the raw data and extract high-level information for the decision-maker.

Indeed, several DM techniques have been applied to the fire detection domain. For example, Vega-Garcia et al. adopted Neural Networks (NN) to predict human caused wildfire occurrence. Infrared scanners and NN were combined in to reduce forest fire false alarms with a 90% success. A spatial clustering (FASTCiD) was adopted by Hsu et al. to detect forest fire spots in satellite images. In 2005, satellite images from North America forest fires were fed into a Support Vector Machine (SVM), which obtained a 75% accuracy at finding smoke at the 1.1-km pixel level. Stojanova et al. have applied Logistic Regression, Random Forest (RF) and Decision Trees (DT) to detect fire occurrence in the Slovenian forests, using both satellite-based and meteorological data. The best model was obtained by a bagging DT, with an overall 80% accuracy.

In contrast with these previous works, we present a novel DM forest fire approach, where the emphasis is the use of real-time and non-costly meteorological data. We will use recent real-world data, collected from the northeast region of Portugal, with the aim of predicting the burned area (or size) of forest fires. Several experiments were carried out by considering three DM techniques (i.e. multiple regression, DT, and SVM) and four feature selection setups (i.e. using spatial, temporal, the FWI system and meteorological data). The proposed solution includes only four weather variables (i.e. rain, wind, temperature and humidity) in conjunction with a SVM and it is capable of predicting the burned area of small fires, which constitute the majority of the fire occurrences. Such knowledge is particularly useful for fire management decision support (e.g. resource planning).

II. How climate change is affecting forest fires around the world

This year has seen unprecedented wildfires cause havoc across the world. Australia recently battled its largest bushfire on record, while parts of the Arctic, the Amazon and central Asia have also experienced unusually severe blazes.

Since this article was first published, the western US has also faced intense fires, with the state of California experiencing its worst fire season since modern records began. The states of Oregon and Washington have also seen a spike in large wildfires in 2020.

It follows on from "the year rainforests burned" in 2019. Last year saw the Amazon face its third-largest fire on record, while intense blazes also raged in Indonesia, North America and Siberia, among other regions.

A rapid analysis released this year found that climate change made the conditions for Australia's unprecedented 2019-20 bushfires at least 30% more likely. Further analysis – visualized below in an interactive map – has shown that, globally, climate change is driving an increase in the weather conditions that can stoke wildfires.

But despite a growing field of evidence suggesting that climate change is making the conditions for fire more likely, research finds that the total area burned by wildfires each year decreased by up to a quarter in the past two decades.

Understanding this paradox requires scientists to assess a vast range of influential factors, including climate change, human land-use and political and social motivations.

III. When and where are most of the world's wildfires?

At any given time, some part of the world is on fire. The map below, from the Global Forest Watch, shows where in the world has had a fire in the past 24 hours.

A "wildfire" can be defined as any type of uncontrolled fire that is spreading across wildland, including pastureland, forests, grasslands and peatlands. (Sometimes, fires are started intentionally and in a controlled manner, including during "prescribed burning".)

Globally, wildfires have many impacts on humans, wildlife and the economy. Wildfires are a major driver of greenhouse gas emissions and are also responsible for 5-8% of the 3.3 million annual premature deaths from poor air quality, research suggests.

Some wildfires are started naturally, chiefly by lightning. The rest are started by humans, either accidentally or by arson.

Across the world, it is estimated that just 4% of fires start naturally. However, the proportion of human-started versus lightning-started fires varies widely from region to region. For example, in the US, 84% of fires are started by humans. However, in Canada, the majority (55%) of wildfires are started by lightning.

Many regions experience distinct wildfire seasons, driven by rainy and dry periods and human practices, such as agricultural burning. However, other regions have a risk of fire year-round.

North America, the Amazon, southern Africa and parts of Australia tend to see an uptick in fires from around August to late November. These months coincide with the height of the dry season in southern Africa and the Amazon.

By contrast, central Africa and parts of southeast Asia see most of their fires in December through to March. This coincides with the peak of the dry season in Africa's Sahel region, which is north of the equator.

The year 2019 saw several large and – in some cases – unprecedented wildfires.

The UK had a record-breaking year in terms of total area burned by fire, with most its fires occurring in February to April. In June to July, more than 100 intense fires broke out across the Arctic Circle, mostly in Alaska and Siberia. August saw the Amazon battle its third-largest fires on record, while September saw large fires in North America and Indonesia. Towards the end of the year, Australia began to face its largest bushfires on record.

Australia's unprecedented fire season continued into 2020, coming to a close at the end of March. The spring saw the return of fires to rural parts of the UK – likely boosted by an unusually hot winter, which left vegetation dry. The end of March also saw large fires break out in southwestern China, which killed at least 19 people and forced almost 25,000 more to evacuate.

From March to July this year, unprecedented heat in the Arctic fanned large "zombie fires" in Siberia. The fires ripped across vast stretches of permafrost, threatening the release of millions of tonnes of long-held carbon.

Analysis by Dr Mark Parrington, a senior wildfire scientist at the EU's Copernicus Atmosphere Monitoring Service, has found that the Arctic fires seen in June 2020 were larger and more intense than those seen in June 2019. However, both the 2019 and 2020 Arctic fires are far above the average level seen in the region from 2003-18.

https://twitter.com/m_parrington/status/1278243686225190913

Reacting to the analysis, fellow wildfire researcher Dr Thomas Smith, from the London School of Economics, posted:

"Will 2019 & 2020 be considered extremes? Or are we seeing the beginning of a new regime?"

June 2020 also saw the return of large fires in the Amazon. This was several months before the region's usual "fire season", which typically runs from August to to October, says Dr Ane Alencar, director of science at the Amazon Environmental Research Institute (IPAM) in Brazil. In June, she told a press briefing:

"I think it is very important and very worrisome to think about what is going to happen during this fire season."

Fires in the Amazon remained unusually severe in July and hit a new high in August. In September, Unearthed reported that fires in Brazil's Pantanal, the world's largest wetland, had reached their highest level since records began in 1998.

At the same time, the western US entered into a record-breaking fire season. Since September, fires in California have been at their highest level on record – and the entire west coast is currently on track to see more land burned by fire in 2020 than in any other year since modern records began in the 1980s.

The US fires have killed at least 30 people and left millions suffering from unsafe air pollution from smoke across the west coast. The fires have also wiped out populations of endangered species, including half of Washington's pygmy rabbit population, according to the New York Times.

Though fires happen all over the world, the largest and fastest-spreading fires mostly occur in sparsely populated grasslands in Australia, Africa and central Asia, according to NASA's fire datasets.

Across all of the world's grasslands, Africa sees a high number of large fires, explains Thailynn Munroe, a fire research analyst at the Global Forest Watch, an open-source forest monitoring service.

In fact, some researchers estimate that up to 70% of the world's fires occur on the African continent. Munroe explains to Carbon Brief:

"Most of the fires that are started in Africa are for land-clearing and agriculture. A lot of it is either clearing for pastureland in sub-Saharan Africa or for agriculture in central Africa. I think that's why we don't see a lot of fires in Africa in the news because it's more just part of the way of life there."

The timing and whereabouts of fire is also influenced by the world's major climate systems, such as El Niño, which periodically affects weather in many world regions, including southeast Asia and South America. Munroe explains:

"Fires in southeast Asia are greatly impacted by El Niño. In 2015-16, there was a really strong El Niño event which brought hot and dry weather. That was one of the reasons why Indonesia had such an intense fire season in that period."

In 2015, Indonesia's wildfires spiked, causing greenhouse gas release on the same scale as Brazil's total annual emissions. The smoke from the fires led to 19 deaths and caused up to half a million people to suffer from respiratory illness, the Guardian reported. Fires in Indonesia have been worsened by the practice of draining peatlands.

IV. HOW IS CLIMATE CHANGE AFFECTING WILDFIRE RISK?

There are several ways in which climate change can raise the risk of wildfires – and the importance of each of these factors varies from region to region, says Dr Cristina Santin, a wildfires researcher from Swansea University.

However, in general, one of the most important ways that climate change can increase the risk of severe fires is by causing vegetation to dry out, she says.

When temperatures are warmer than average, rates of evaporation increase, causing moisture to be drawn out from plants on the land. This drying can create "tinderbox conditions" – meaning that, if a fire is sparked, it can spread very quickly over large areas.

Similar conditions can also be created by long periods of drought, Santin tells Carbon Brief:

"In the fire community, we call vegetation that is available to burn 'fuel'. If you have a forest, not all vegetation is considered fuel because, under normal circumstances, it's not going to burn. But if you have a huge drought or heatwave, a lot of that vegetation will be very dry and, therefore, it will become fuel."

Such conditions occurred during the 2018 northern hemisphere heatwave, which saw all-time temperature records broken across Europe, North America and Asia.
In the Attica region of Greece, wildfires ripped across large swathes of dry land at lightning speed, causing people to rush to nearby beaches and into the sea. (A study covered by Carbon Brief found that the 2018 northern-hemisphere heatwave would have been "impossible" without human-caused climate change.)
Sustained hot temperatures were also a major driver of Australia's unprecedented 2019-20 bushfires, says Dr Friederike Otto, acting director of the Environmental Change Institute at the University of Oxford.
Otto co-authored an analysis finding that temperatures during the bushfires were 1C to 2C hotter than they would have been in the early 20th century. She tells Carbon Brief:
"We found that climate change made the bushfires at least 30% more likely – and that is a conservative estimate. When we looked at temperatures alone, we found there has been a big increase in extreme temperatures in that region because of climate change."
Warmer than average temperatures are also likely to be the primary driver of this year's record-breaking fire season in the western US, the Atlantic reports.
As well as making fires more severe, warming temperatures are also making fire seasons longer in some regions, explains Dr Megan Kirchmeier-Young, a researcher of climate extremes at the Government of Canada. (Fire seasons are a stretch of a time when a particular region is most likely to see large and intense blazes.) She tells Carbon Brief:
"Climate change is affecting wildfires in two main ways. The first is an increase in the risk or the likelihood of wildfire. The second is longer fire seasons – and this is mostly coming from warming temperatures."
A study published in 2015 found that, globally, the number of days where wildfires are likely to burn has risen as a result of climate change.
The influence of climate change on fire seasonality is especially pronounced in regions that have seasonal snow cover, including parts of North America and northern Europe, she explains:
"In some regions, especially in Canada, we also consider when we lose winter snow cover. With warming temperatures, that's going to be happening earlier in the year and it will also be later in the year before we see cooler temperatures and the return of snow cover. So a longer fire season means more time of the year where you could have fires."

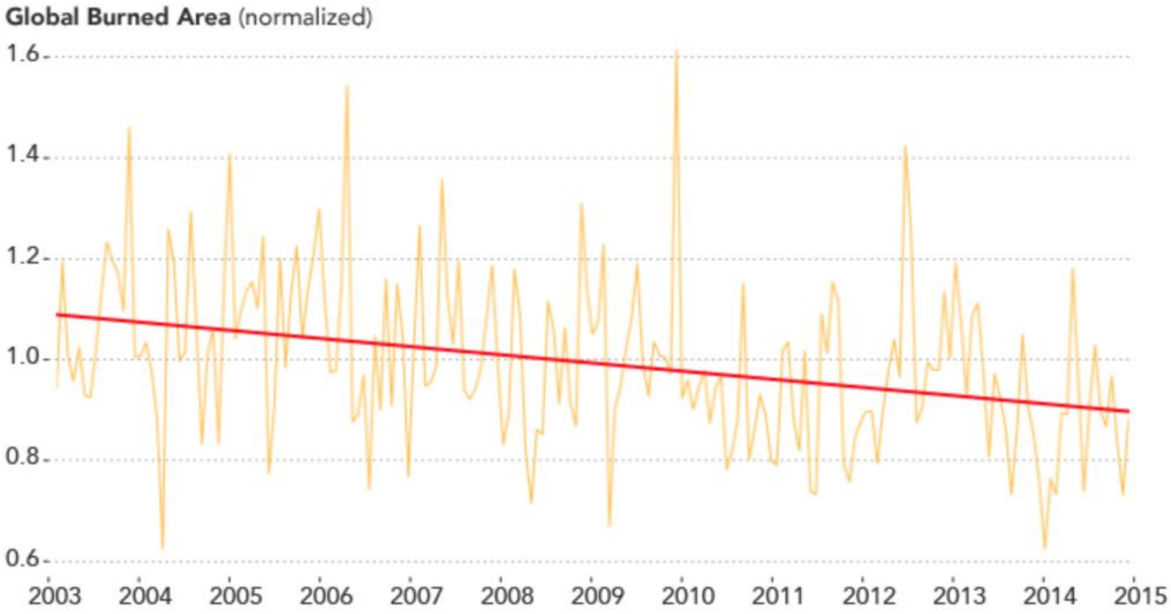
In the Amazon, human-caused climate change and shifts to how people use the land have worked in tandem to greatly increase the risk of wildfires, says Alencar. In June, she told a press briefing:
"Unlike some forests in California, Florida, or in Australia, the Amazon doesn't burn naturally. The natural fire regime in the Amazon is said to be somewhere between 500 to 1,000 years. Our results have demonstrated that in some places in the Amazon, actually, that fire regime has changed to 12 years. So, an area that should burn naturally every 500 to 1,000 years is burning every 12."
This year, Science Brief – a UK-based web platform run by a team of scientists – released a review of 73 scientific studies finding that climate change is increasing the risk of wildfires at a global level.
For the review, the scientists analysed all research papers published since 2013 that investigate the link between climate change and wildfires, either in a certain region or from a global perspective.
The researchers assessed how each study compared to the statement: "Climate change increases the risk of wildfires".
Carbon Brief's interactive map below includes each of the 73 studies, which are displayed according to what part of the world they focus on.
Studies that fully support the statement that climate change increases wildfire risk are represented with a dark red icon, studies that mostly support the statement are represented in dark orange and studies that inform the statement but do not fully support it are represented in light orange.

V. Are wildfires increasing across the globe?

With climate change raising the risk of hot and dry weather in many parts of the world, it may seem prudent to assume that the global area burned by wildfires each year is increasing.

However, several research papers looking into wildfires at a global level have come to the opposite conclusion. A paper published in the journal Science found that, globally, the area burned by wildfires decreased by 25% between 2003 and 2015.

The animation below, which uses data from the study, illustrates the pattern of how global area burned by wildfires changed over this time period.



Global burned area chart 2003-2015. Source: NASA Earth Observatory.

The chart shows how the global burnt area zigzags throughout a year-long period. This is likely down to the influence of dry seasons and seasonal human practices, says Munroe. (For more information see: "When and where are most of the world's wildfires?")

However, despite peaks and troughs, an overall downward trend is seen until 2015. (And unpublished data from Global Forest Watch suggests that this downward trend continued up until 2020, Munroe says.)

There are several reasons why global burnt area could be decreasing at a global level despite the effect of warming, explains Santin, who co-authored a research paper looking into this paradox in 2016.

According to NASA scientists, the Amazon's severe 2019 fire season was "more consistent with land clearing than with regional drought".

"Slash-and-burn" clearing has also been linked to damaging peat fires in Indonesia.

Fires in the country have been exacerbated further by the practice of peatland draining. In order to grow palm oil and other crops, such as timber, peatlands are often drained of their natural moisture – leaving them dry and more likely to catch alight.

The influence of human land-use activity on global burnt area makes it an imperfect metric of how climate change is truly affecting wildfires, Prof Anthony Westerling a fire and drought researcher at the University of California, Merced, told Carbon Brief in 2018.

Another way that scientists could study the influence of climate change is by tracking fire severity.

Research published in 2017 analysed all "extreme wildfire events" from 2002-13 and found that 96% occurred during periods of hot and dry weather. ("Extreme wildfires" were deemed to be fires that caused serious economic or social damage.)

However, there is little research into how fire severity has changed at a global scale in recent decades. (A string of research papers have linked increasing fire severity in North America to climate change.)

VI. How will wildfires change in the future?

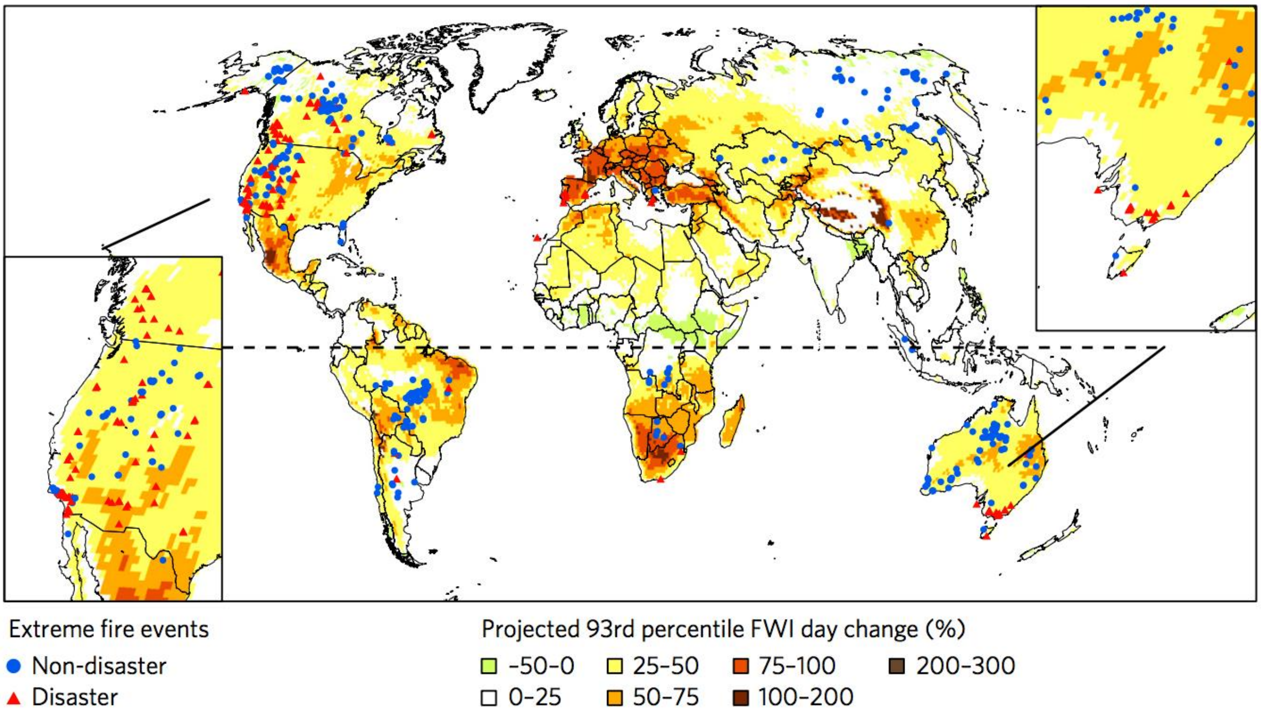
Climate change will continue to drive temperature rise and more unpredictable rainfall in many parts of the world, meaning that the number of days with "fire weather" – conditions in which fires are likely to burn – is expected to increase in coming decades, says Kirchmeier-Young:

"As we continue to see increasing temperatures, we will continue to see an increase in the likelihood of wildfires. So, more days with increased fire risk, longer fire seasons."

A study covered by Carbon Brief found that, by the middle of the century, there could be a 35% increase in the days with a high danger of fire across the world, on average – if little action is taken to tackle climate change.

The regions likely to see the highest increase in days with extreme fire weather in this very high emissions scenario ("RCP8.5") include the western US, southeastern Australia, the Mediterranean and southern Africa, according to the study.

This is highlighted on the map below, which shows the projected change in the number of days with "high" fire danger from 2000-14 to the middle of the century (2014-70). The regions with the largest increase are shaded orange and red, while the areas with decreasing risk are shown in green



Other studies have investigated how the risk of fire weather is likely to change in different world regions if the world does take action to tackle climate change.

For example, a study published in 2020 found that, in California, "climate change will further amplify the number of days with extreme fire weather by the end of this century". However if efforts are taken to limit global warming to below 2C, which is the goal of the Paris Agreement, this "would substantially curb that increase", the authors say.

In the Mediterranean, limiting global warming to 1.5C, which is the aspirational target of the Paris Agreement, could halve the total area burned by wildfires in the summer, when compared to a scenario where warming reaches 3C, according to a second study.

In Australia, days with fire weather akin to that seen during the 2019-2020 bushfires could become at least four times more likely under 2C of global warming, according to Otto's recent analysis.

A landmark special report in 2018 from the Intergovernmental Panel on Climate Change (IPCC) concluded (pdf) that limiting global warming to 1.5C rather 2C would "reduce" the average wildfire risk worldwide.

However, though the risk of wildfires is likely to heighten in coming decades, it is not yet clear whether the area burned by wildfires will increase correspondingly, says Santin:

"It's a complicated issue because it's not only climate change controlling future fire risk, it's also the interaction of humans and climate change."

It is possible that efforts by humans to suppress wildfires could stem increases in the area burned by fire, despite the increased risk posed by climate change, she says.

VII. TURKEY WILDFIRES: DESPAIR AND QUESTIONS AS FORESTS BURN

Devastating wildfires have tore through forests and villages, killing at least eight people and burned through huge tracts of land.

Manavgat, Turkey – Turkey's southern coastline is burning. On the wooded hills of Antalya's Manavgat district, plumes of thick smoke appear in the sky one after another as each time a forest fire is brought under control, another seems to ignite.

A blood-red sun shines through the sallow haze and as visibility clears, the charred, skeletal remains of what were forests and villages are revealed. This, many believe, is just the latest sign the world is entering an era of climate crisis, and Turkey is not prepared for it.

Over the past six days, 132 destructive blazes have raged through southern and other parts of Turkey, killing eight people and burning at least 118,789 hectares of land, according to the European Forest Fire Information System.

While controversy is rife, with many in Turkey believing the fires are the result of "sabotage" – a theory encouraged by many politicians – they coincide both with months of severe drought and extreme temperatures.

Antalya, a tourist hotspot that averages near the mid-30s Celsius (95 Fahrenheit) at this time of year, has seen highs of more than 40C this week. On July 20, Turkey recorded its hottest ever temperature at 49.1 (120.38 Fahrenheit) degrees in the southeast.

Many of the blazes tore through forests near beach destinations popular with local and European tourists, such as Bodrum and Marmaris, with people fleeing in cars, small boats and in some cases luxury yachts. Soaring temperatures have also seen wildfires break out across much of southern Europe, including Greece, Spain and Italy.

Manavgat is among the most fire-affected places in Turkey, and while seasonal fires are normal and even healthy for the local ecosystem, environmental groups say they have never been seen on this scale. With the landscape parched and strong winds – in particular, one that blows from the northeast known as "poyraz" in Turkey – the authorities are struggling to move fast enough to control things.

The small village of Sirtkoy, where the main income comes from growing aromatic bay leaves used in cooking, caught fire in the early hours of Sunday morning. Within an hour, the local school was gutted and many of the houses were reduced to rubble.

"All of this area was fine yesterday," said resident Mustafa, who did not want to give his last name, pointing to a blackened, still-smouldering pile of stones that had been his friend's house. "The fires came this morning at 5am and this burned down. At 6am, the fire was done, but at 9am the wind came back again and so did the fire."

When Al Jazeera visited Sirtkoy on Sunday afternoon, planes, firefighters and forestry workers battled tirelessly to control flames that, like a trick candle, kept reigniting. With pungent smoke turning the whites of their eyes red, many of the men had little more to protect their lungs than disposable surgical masks.

Villagers doused exterior walls and possessions with bottles of water, trying desperately to keep the flames away, while some asked why the authorities were yet to take responsibility for a lack of preparation for a disaster that many could see coming.

"No one has taken responsibility for these fires and we have nothing after today," said resident Hatice Cinar, the sound of burned trees collapsing ringing from the forest below.

Cinar said the village is not suitable for growing vegetables or keeping animals so the only way of life they know is growing bay leaves – she had 500 trees, which she had hoped would give her 18-year-old son a future, but they have all been destroyed.

VIII. FOREST FIRE DATA

The forest Fire Weather Index (FWI) is the Canadian system for rating fire danger and it includes six components (Figure 1) : Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), Buildup Index (BUI) and FWI. The first three are related to fuel codes: the FFMC denotes the moisture content surface litter and influences ignition and fire spread, while the DMC and DC represent the moisture content of shallow and deep organic layers, which affect fire intensity.

The ISI is a score that correlates with fire velocity spread, while BUI represents the amount of available fuel. The FWI index is an indicator of fire intensity and it combines the two previous components. Although different scales are used for each of the FWI elements, high values suggest more severe burning conditions. Also, the fuel moisture codes require a memory (time lag) of past weather conditions: 16 hours for FFMC, 12 days for DMC and 52 days for DC.

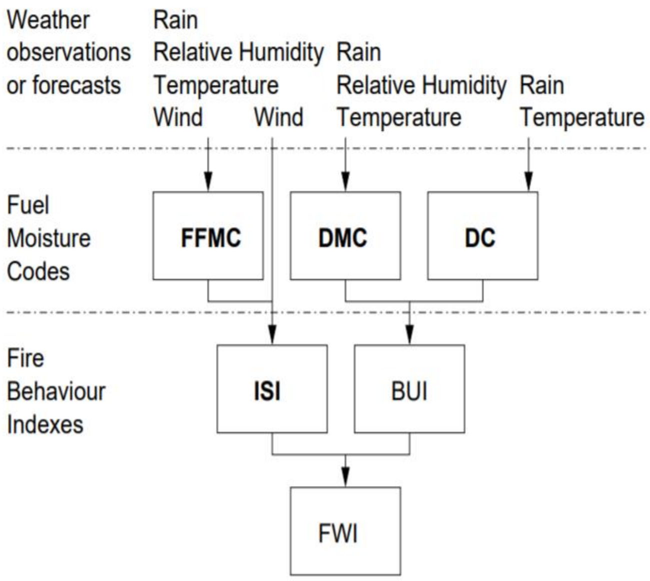


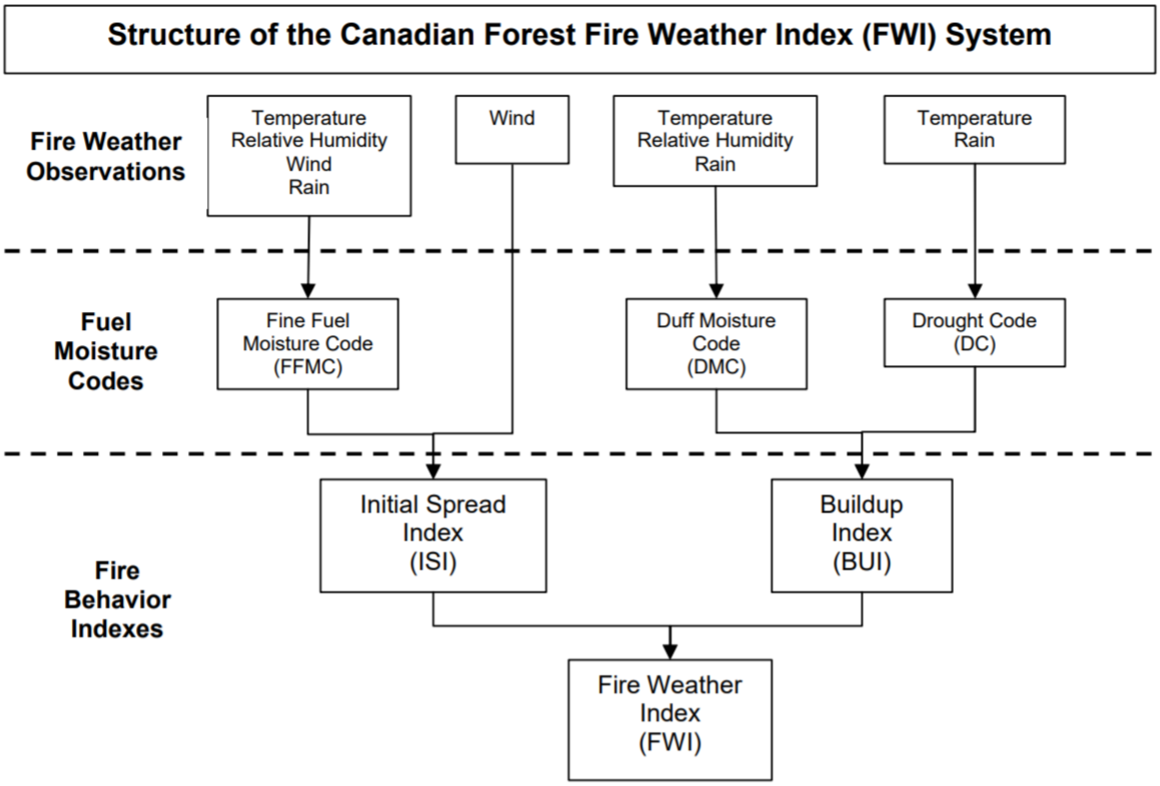
Fig. 1. The Fire Weather Index structure (adapted from [24])

IX. INTERPRETING THE CANADIAN FOREST FIRE WEATHER INDEX (FWI) SYSTEM

Fire danger is defined by the Canadian Committee on Forest Fire Management (Merrill and Alexander 1987) as: A general term used to express an assessment of both fixed and variable factors of the fire environment which determine the ease of ignition, rate of spread, difficulty of control and fire impact. The Canadian Forest Fire Danger Rating System (CFFDRS) is the national system for rating fire danger in Canada. The Canadian Forest Fire Weather Index (FWI) System is a sub-system of the CFFDRS and has been in its present form since 1970, with the fourth version of the tables for the FWI System now being used (Canadian Forestry Service 1984; Van Wagner 1987). The purpose of the FWI System is to account for the effects of weather on forest fuels and forest fires. Other factors affecting fire danger (i.e., fuels, topography) are dealt with elsewhere in the CFFDRS.

The FWI System is comprised of six components (see Fig. 1): three fuel moisture codes and three fire behavior indexes. Each component has its own scale of relative values. Even though the scales for the six components are different, all are structured so that a high value indicates more severe burning conditions. The FWI System uses temperature, relatively humidity, wind speed, and 24-hr precipitation values measured at noon Local Standard Time (LST). These values are used to predict the peak burning conditions that will occur during the heat of the day, near

1600 hr LST, assuming that the measured weather parameters follow a normal diurnal pattern (Turner and Lawson 1978; Van Wagner 1987). 1 A presentation made at the Fourth Central Region Fire Weather Committee Scientific and Technical Seminar, April 2, 1987, Winnipeg, Manitoba. 2 Fire Research Officer, Saskatchewan District Office Canadian Forestry Service, Western and Northern Region 101-15 Street East, Prince Albert, Sask., S6V 1G1.



Fuel Moisture Codes

The FWI System evaluates fuel moisture content and relative fire behavior using the past and present effect of weather on forest floor fuels. The three moisture codes represent the fuel moisture content of three classes of forest floor fuels in the “standard” mature pine stand (Fig. 2). The moisture codes calculate the net effect of a daily drying and wetting phase, similar to a bookkeeping system of moisture losses and additions.

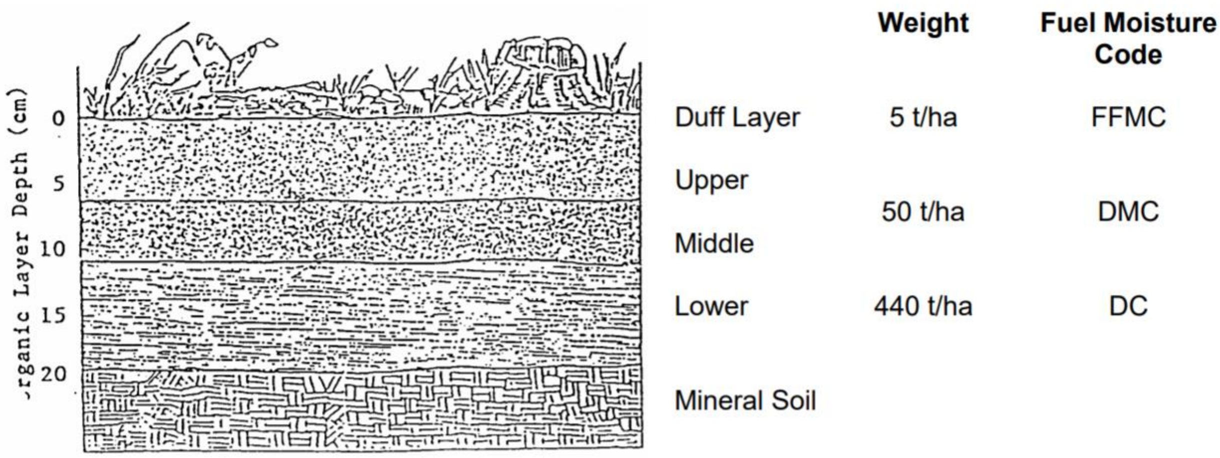


Figure 2. Representation of forest floor fuels by Fuel Moisture Codes of the FWI System

Fine Fuel Moisture Code (FFMC) The FFMC is a numerical rating of the moisture content of litter and other cured fine fuels (needles, mosses, twigs less than I cm in diameter). The FFMC is representative of the top litter layer less than 1 inch deep (1-2 cm deep), and has a typical fuel loading of about 2.2 tons per acre (5 tonnes per hectare). FFMC fuels are affected by temperature, wind speed, relative humidity, and rain. However, to account for the interception of rain by the forest canopy, the wetting phase of the FFMC is not initiated if the 24-hr rainfall is 0.03 inches (0.5 mm) or less. The rate at which fuels lose moisture is measured in terms of timelag, similar to the 'half-life' decay rate of radioactive material.

Timelag is the time required for fuel to lose two-thirds of its free moisture content with a noon temperature reading of 70°F (21°C), relative humidity of 45%, and a wind speed of 10 mph (13 km/h) (Lawson 1977). The timelag for FFMC fuels is two-thirds of a day. FFMC values change rapidly because of a high surface area to volume ratio, and direct exposure to changing environmental conditions. This characteristic of rapidly changing moisture content causes the FFMC to have a short-term memory and only reflects the weather conditions that have occurred over the past three days. The FFMC can be adjusted for times other than 1600 h LST (Van Wagner 1972, 1977; Alexander 1982a; Alexander et al. 1984) to account for changing moisture content of the fine fuels throughout the day or to allow for an irregular diurnal pattern of temperature or humidity. Because fires usually start and spread in fine fuels, the FFHC is used to indicate ease of ignition, or ignition probability (Fig. 3). The FFMC scale ranges from 0-99 and is the only component of the FWI System which does not have an open-ended scale.

Generally, fires begin to ignite at FFMC values near 70, and the maximum probable value that will ever be achieved is 96. At the high end of the scale, a general rule of thumb is that the fuel moisture content is 101 minus the 'FFMC value. Of importance is the fact that fire starts increase exponentially with an increase in FFMC values at the high end of the scale. In the boreal forest, a high potential for fire starts exists once the FFMC reaches 86-89.

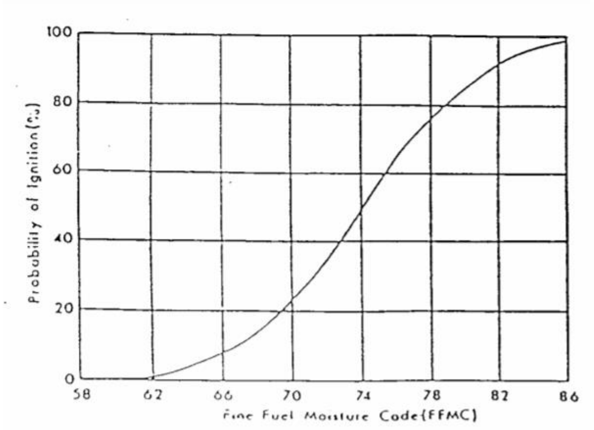


Figure 3. Ignitability of 'shaded' slash pine needle litter under 'no wind' conditions as a function of the Fine Fuel Moisture Code (adapted from Blackmarr 1972 by M.E. Alexander based on Van Wagner 1987).

Duff Moisture Code (DMC) The DMC indicates the moisture content of loosely-compacted organic layers of moderate depth. It is representative of the duff layer that is 2-4 inches (5-10 cm) deep, and has a fuel loading of about 22 tons per acre (50 t/ha).

DMC fuels are affected by rain, temperature and relative humidity. Because these fuels are below the forest floor surface, wind speed does not affect the fuel moisture content.

A 24-hr rainfall of less than 0.06 inches (1.5 mm) has no effect on the DMC because of interception by the forest canopy And the fine fuel layer.

The DMC fuels have a slower drying rate than the FFMC fuels, with a timelag of 12 days. Due to the slower drying rate, the length of daily drying time is important. Therefore, a seasonal day-length factor has been incorporated into the drying phase of the DMC.

Although the DMC has an open-ended scale, the highest probable value is in the range of 150. The DMC is often used to assist in predicting the probability of lightning fire start? (Fig. 4) Since lightning, strikes usually result in fires smoldering in the duff layer.

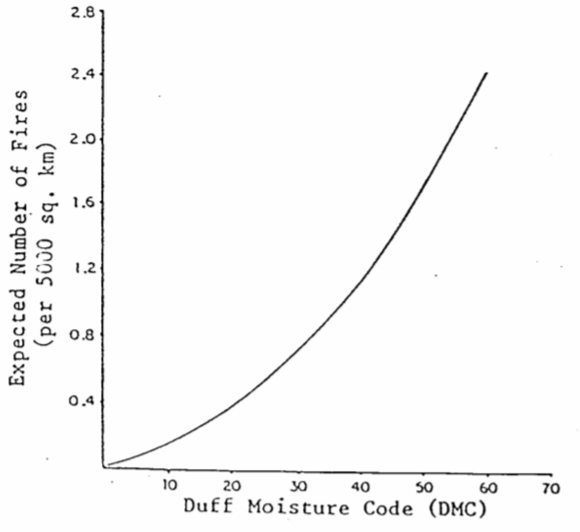


Figure 4. Typical relationship between DMC and lightning fire starts (adapted from Martell 1976)

Drought Code (DC) The third moisture code is the DC, and it is an indicator of moisture, content in deep, compact organic layers. This code represents the fuel layer approximately 4 to 8 inches (10-20 cm) deep, having a fuel loading of about 200 tons/acre (440 t/ha).

Temperature and rain affect the DC, although wind speed and relative humidity do not because of the depth of this fuel layer. A 24-hr rainfall greater than 0.11 inches (2.8 mm) is required to affect the moisture content due to interception by upper fuel layers and the forest canopy.

The DC fuels have a very slow drying rate, with a timelag of 52 days. Therefore, a seasonal daylength factor is also incorporated in the drying phase.

The DC is indicative. Of long-term moisture conditions and can be used in estimating mop-up difficulty due to deep burning fires (Table 1). The DC scale is open-ended, although the maximum probable value is about 800.

Because of the slow drying rate of DC fuels, the amount of over winter precipitation is critical to calculating spring starting values. If there has not been sufficient over winter precipitation to recharge moisture levels in the deep organic layers, then an upward adjustment of the DC in the spring must be done to reflect the drier conditions (Turner and Lawson 1978; Alexander l982b, 1983).

Table 1. Mop-up recommendations as determined by the Drought Code (adapted after Muraro and Lawson 1970; Canadian Forestry Service, 1971).

DC	INTERPRETATION
< 300	Moisture will increase with depth. Usual attention to mop-up and patrol, with closer attention to critical perimeters as a DC value of 300 is approached.
300 - 500	Moisture content may decrease with depth. Extensive mop-up of edges should be initiated as control problems could be posed by critical edges.
> 500	Moisture content will most likely decrease with depth. Extensive mop-up and patrol of all edges is required.

TABLE 2. Summary of Fuel Moisture Code Features.

ITEM	FFMC	DMC	DC
Fuel Association	Litter and other cured fine fuels	Loosely-compacted organic layers of moderate depth	Deep, compact organic layers
Fire Potential Indicator	Ease of Ignition	Probability of lightning fires; fuel consumption in moderate duff	Mop-up difficulty; fuel consumption of deep organic material
Depth (cm; inches)	1-2 cm (.4" to .8")	5-10 cm (2"-4")	10-20 cm (4"-8")
Fuel Loading (t/ha; t/ac)	5 t/ha; 2.2 t/ac	50 t/ha; 22 t/ac	440 t/ha; 196 t/ac
Required Weather Inputs:			
Dry-Bulb Temperature	X	X	X
Relative Humidity	X	X	
Windspeed	X		
Rain	X	X	X
24 hr Rainfall Threshold (mm; inches)	0.5 mm-0.03"	1.4 mm-0.06"	2.8 mm-0.11"
Timelag Constant	16 hrs	12 days	52 days
Value Range	0-99	0-350 [1]	0-1200 [1]
Maximum Probable Value	96	150	800
Spring Starting Value	85	6	15 [2]

An open-end scale; the upper value is shown for convenience of comparing the relative range of scales.

This value may be adjusted upwards to account for lack of sufficient over winter precipitation.

X. Fire Behavior Indices

Initial Spread Index (ISI) The ISI combines the FFMC and wind speed to indicate the expected rate of fire spread (Fig. 5). Generally, a 10 mph (13 km/h) increase in wind speed will double the ISI value. The ISI is accepted as a good indicator of fire spread in open light fuel stands with wind speeds up to 25 mph (40 km/h.

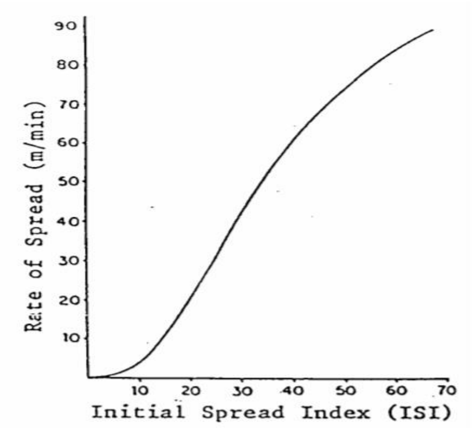


Figure 5, .Rate of spread for the mature jack or lodgepole pine fuel type on level terrain as a function of ISI (from Alexander, Lawson, Stocks and Van Wagner 1984).

Buildup Index (BUI) The BUI is a weighted combination of the DMC and DC to indicate the total amount of fuel available for combustion by a moving flame front (Fig. 6). The DMC has the most influence on the BUI value. For example, a DMC value of zero always results in a BUI value of zero regardless of what the DC value is. The DC has strongest influence on the BUI at high DMC values, and the greatest effect that the DC can have is to make the BUI value equal to twice the DMC value. This weighting procedure makes the BUI an upper organic layer moisture monitor with a deep duff indicator built in. The BUI is often used for pre-suppression planning purposes.

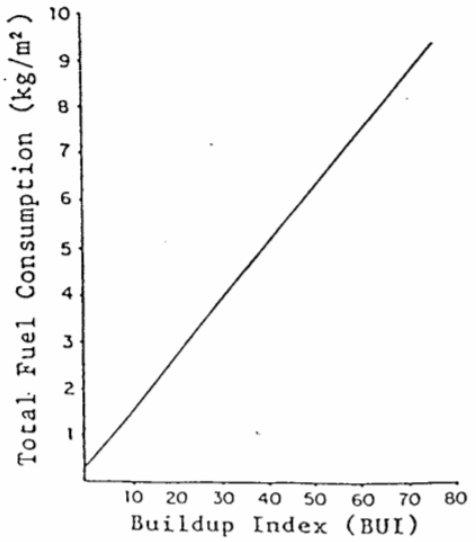


Figure 6. Relationship between total fuel consumption and BUI in jack pine slash (from Stocks and Walker 1972).

Fire Weather Index (FWI) The FWI is a combination of ISI and BUI, and is a numerical rating of the potential frontal fire intensity (Fig. 7). In effect, it indicates fire intensity by combining the rate of fire spread with the amount of fuel being consumed. Frontal fire intensity is useful for determining fire suppression requirements, as shown in Alexander and De Groot (1988). As well, the FWI is used for general public information about fire danger conditions.

The Canadian Forest Fire Weather Index (FWI) System consists of six components that account for the effects of fuel moisture and weather conditions on fire behavior. The first three components are fuel moisture codes, which are numeric ratings of the moisture content of the forest floor and other dead organic matter. Their values rise as the moisture content decreases. There is one fuel moisture code for each of three layers of fuel: litter and other fine fuels; loosely compacted organic layers of moderate depth; and deep, compact organic layers.

The remaining three components are fire behavior indices, which represent the rate of fire spread, the fuel available for combustion, and the frontal fire intensity; these three values rise as the fire danger increases.

Calculation of the components is based on consecutive daily observations of temperature, relative humidity,wind speed, and 24-hour precipitation. The six standard components provide numeric ratings of relative potential for wildland fire.

Operational Application

The FWI System provides relative numerical ratings of fire potential over a large area - represented .by an individual fire weather station site. Understanding the limits of such a system will ensure its proper application. For instance, to account for isolated rainfall at a weather station, the fire manager must also calculate a second set of FWI System values using no-rain to represent areas which did not receive any precipitation (the calculation using the actual rainfall at the ' weather station is used for the following days calculation). A recalculation of the FWI System would also have to be done if normal diurnal conditions did not occur between noon and the peak burning period. For example, this would typically be done after a frontal passage and would only be valid for that afternoon (but not used for the following day's calculation).

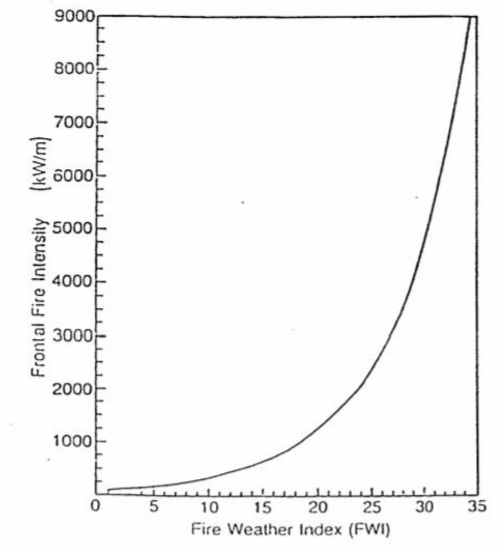


Figure 7. Frontal fire intensity in mature jack pine as a function of the FWI (from Alexander and De Groot 1988).

An understanding of the sensitivity of the FWI System can only be gained by daily observation of the component values and changing weather conditions. By comparing fire activity (fire starts, rate of spread, difficulty of control, etc.) to the values produced by the FWI System, fire managers will gain an expertise in interpreting the FWI System.

XI. The Fire Weather Index Structure

This study will consider forest fire data from the Montesinho natural park, from the Tr´as-os-Montes northeast region of Portugal (Figure 2). This park contains a high flora and fauna diversity. Inserted within a supra-Mediterranean climate, the average annual temperature is within the range 8 to 12◦C. The data used in the experiments was collected from January 2000 to December 2003 and it was built using two sources. The first database was collected by the inspector that was responsible for the Montesinho fire occurrences. At a daily basis, every time a forest fire occurred, several features were registered, such as the time, date, spatial location within a 9×9 grid (x and y axis of Figure 2), the type of vegetation involved, the six components of the FWI system and the total burned area. The second database was collected by the Braganc¸a Polytechnic Institute, containing several weather observations (e.g. wind speed) that were recorded with a 30 minute period by a meteorological station located in the center of the Montesinho park. The two databases were stored in tens of individual spreadsheets, under distinct formats, and a substantial manual effort was performed to integrate them into a single dataset with a total of 517 entries.

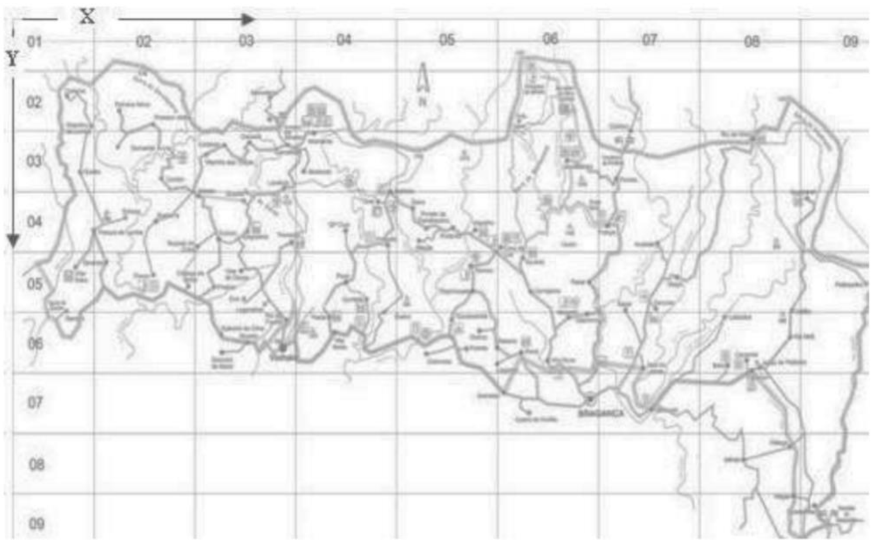


Fig. 2. The map of the Montesinho natural park

Table 1 shows a description of the selected data features. The first four rows denote the spatial and temporal attributes. Only two geographic features were included, the X and Y axis values where the fire occurred, since the type of vegetation presented a low quality (i.e. more than 80% of the values were missing). After consulting the Montesinho fire inspector, we selected the month and day of the week temporal variables. Average monthly weather conditions are quite distinct, while the day of the week could also influence forest fires (e.g. work days vs weekend) since most fires have a human cause. Next come the four FWI components that are affected directly by the weather conditions (Figure 1, in bold). The BUI and FWI were discarded since they are dependent of the previous values. From the meteorological station database, we selected the four weather attributes used by the FWI system. In contrast with the time lags used by FWI, in this case the values denote instant records, as given by the station sensors when the fire was detected. The exception is the rain variable, which denotes the accumulated precipitation within the previous 30 minutes.

Table 1. The preprocessed dataset attributes

Attribute	Description
X	x-axis coordinate (from 1 to 9)
Y	y-axis coordinate (from 1 to 9)
month	Month of the year (January to December)
day	Day of the week (Monday to Sunday)
FFMC	FFMC code
DMC	DMC code
DC	DC code
ISI	ISI index
temp	Outside temperature (in °C)
RH	Outside relative humidity (in %)
wind	Outside wind speed (in km/h)
rain	Outside rain (in mm/m^2)
area	Total burned area (in ha)

The burned area is shown in Figure 3, denoting a positive skew, with the majority of the fires presenting a small size. It should be noted that this skewed trait is also present in other countries, such as Canada. Regarding the present dataset, there are 247 samples with a zero value. As previously stated, all entries denote fire occurrences and zero value means that an area lower than 1ha/100 = 100m2 was burned. To reduce skewness and improve symmetry, the logarithm function $y = \ln(x + 1)$, which is a common transformation that tends to improve regression results for right-skewed targets, was applied to the area attribute (Figure 3). The final transformed variable will be the output target of this work.

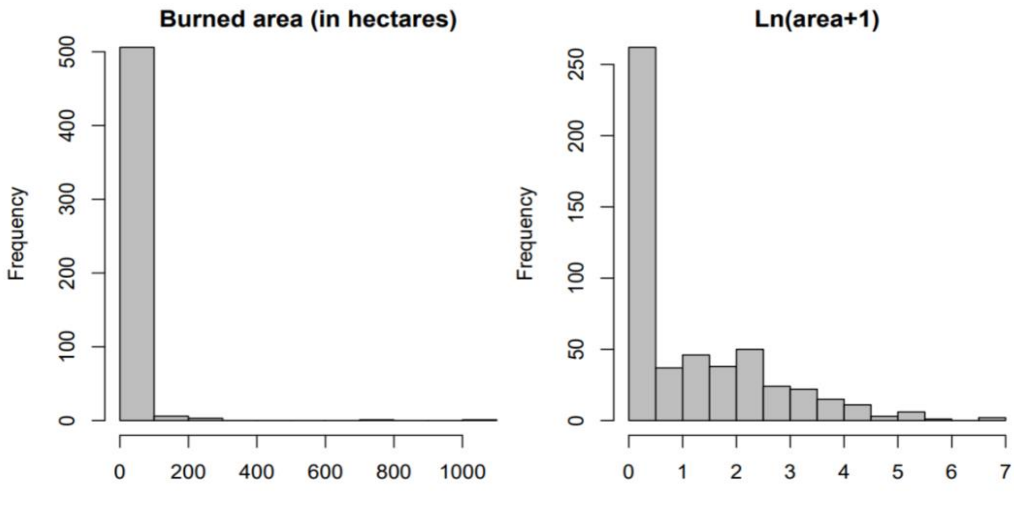


Fig. 3. The histogram for the burned area (left) and respective logarithm transform (right)

XII. Data Mining Models

Several DM algorithms, each one with its own purposes and capabilities, have been proposed for regression tasks. This work will consider three DM models. The Multiple Regression (MR) model is easy to interpret and this classical approach has been the widely used. Yet, it can only learn linear mappings. To solve this drawback, one alternative is to use methods based on tree structures, such as Decision trees (DT) ,or nonlinear functions, such as Support Vector Machines (SVM).

The DT is a branching structure that represents a set of rules, distinguishing values in a hierarchical form.

1. Decision Tree Regressor Decision tree are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split. An example of a decision tree can be explained using above binary tree. Let’s say you want to predict whether a person is fit given their information like age, eating habit, and physical activity, etc. The decision nodes here are questions like ‘What’s the age?’, ‘Does he exercise?’, ‘Does he eat a lot of pizzas’? And the leaves, which are outcomes like either ‘fit’, or ‘unfit’. In this case this was a binary classification problem (a yes no type problem). There are two main types of Decision Trees:

1.1 Classification trees (Yes/No types)

What we’ve seen above is an example of classification tree, where the outcome was a variable like ‘fit’ or ‘unfit’. Here the decision variable is Categorical.

1.2 Regression trees (Continuous data types)

Here the decision or the outcome variable is **Continuous**, e.g. a number like 123. **Working** Now that we know what a Decision Tree is, we’ll see how it works internally. There are many algorithms out there which construct Decision Trees, but one of the best is called as **ID3 Algorithm**. ID3 Stands for **Iterative Dichotomiser 3**. Before discussing the ID3 algorithm, we’ll go through few definitions.

Entropy: Entropy, also called as Shannon Entropy is denoted by H(S) for a finite set S, is the measure of the amount of uncertainty or randomness in data. Intuitively, it tells us about the predictability of a certain event. Example, consider a coin toss whose probability of heads is 0.5 and probability of tails is 0.5. Here the entropy is the highest possible, since there’s no way of determining what the outcome might be.

Information Gain: Information gain is also called as Kullback-Leibler divergence denoted by IG(S,A) for a set S is the effective change in entropy after deciding on a particular attribute A. It measures the relative change in entropy with respect to the independent variables. Alternatively, where IG(S, A) is the information gain by applying feature A. H(S) is the Entropy of the entire set, while the second term calculates the Entropy after applying the feature A, where P(x) is the probability of event x.

2. Support Vector Machines (Regressor)

We fit a Support Vector Machine for each feature selection setups for forest fires data. We fit the Support Vector Machine (SVM) using svm function in e1071 library in R.

In SVM regression, the input x *2* $\mathbf{R}^{A}$ is transformed into a high m-dimensional feature space, by using a nonlinear mapping. Then, the SVM finds the best linear seperating hyperplane in the feature space:

$$\hat{y} = b + \sum_{i=1}^{m} w_i \phi_i(x)$$

where φi(x) represents a nonlinear transformation, according to the kernel function

$$K(x, x') = \sum_{i=1}^{m} \phi_i(x)\phi_i(x').$$

We used the popular Radial Basis Function kernel, which presents less hyperparameters and numerical difficulties than other kernels (e.g. polynomial or sigmoid.

SVM present theoretical advantages over NN, such as the absence of local minima in the model optimization phase. In SVM regression, the input x ∈ ℜA is transformed into a high m-dimensional feature space, by using a nonlinear mapping. Then, the SVM finds the best linear separating hyperplane in the feature space.

Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line).

XIII. Objective/Aim

This project aims at developing and implementing algorithms of machine learning on a dataset which is of Northern Portugal. The dataset is a collection of forest fire attributes in which area, the output attribute, implies the area (in ha) burned due to fire. The dataset contains a total of 13 attributes aka conditions for recording the total area (output of dataset) burned due to forest fire. There ere several attributes like the coordinates of the park where the burned area id recorded, month and day on which the recordings are taken, FFMC index of FWI system, DMC index of FWI system, Relative Humidity (RH), temp, wind, outside rain. This project aims at applying various machine learning algorithms to find the accuracy of each of them regarding the prediction of how much area the fire will burn.

XIV. Innovativeness and Usefulness

Training a machine by providing a training data set for the purpose of analysing the data and the finally make decisions according the different machine learning algorithms is the sole purpose of this project. The innovative power of machine learning algorithms and usefulness of the data represents the overall accuracy of the project in providing good results. Data is fed to the machine as a training set which is usually of the ratio 70 out of 70:30 whole set. This 70% of the data is anaysed by the machine using various machine learning algorithms and yielding results which are then recorded, processed, analysed. The machine on the basis of this 70% training dataset makes decisions on the test data set provided to the machine at the end to calculate the final accuracy of the whole project. Now machine applies the same rules earlier applied to the training dataset to the test set and yields results accordingly. Different algorithms are used to find the accuracy of each algorithm. The algorithm which has the maximum accuracy is preferred and further decisions are made as per that algorithm.

XV. Performance of the Algorithms

Among all the algorithms we apply over the dataset for yielding and calculating the accuracy of the results, the one with the maximum accuracy will be selected as the idle one. We are given to implement different algorithms individually over the dataset and the assemble the results of the algorithm evaluations. For example if we are a team of four members, each of us will implement different machine

learning algorithms over the dataset and we will the calculate the accuracy at the individual level. After calculating the accuracy at individual level, we will then directly compare our results with each other and see who among the team mates has got the maximum accuracy among all. Simultaneously we will then see which algorithm the team member was provided with to work on the dataset and has got the maxmum result. Then that algorithm's performance will be evaluated on the basis of how much lines of code the algorithm demands in order to facilitate the whole operation. Thus performance of the algorithm will finally be decided upon the factors of how much accuracy the algorithm has provided over the given dataset and how many lines of code does the algorithm require in order to implement the same procedure of finding the accuracy in programming language environment

Experimental Results

Before fitting the models, some preprocessing was required by the DT and SVM models.

Data Transformation

```
In [131]: df.dtypes
Out[131]: X          int64
          Y          int64
          month     object
          day       object
          FFMC     float64
          DMC      float64
          DC       float64
          ISI      float64
          temp     float64
          RH         int64
          wind     float64
          rain     float64
          area     float64
          dtype: object
```

The above image depicts that the attributes viz Month and Day were object types because they contain string values. Since it is impossible to process non numeric data , we transformed the attributes containing string values into numeric values.

The nominal variables (i.e. discrete with more than two non-ordered values), such as the month and day, were transformed . Also, for SVM methods, all attributes were standardized to a zero mean and one standard deviation. Next, the regression models were

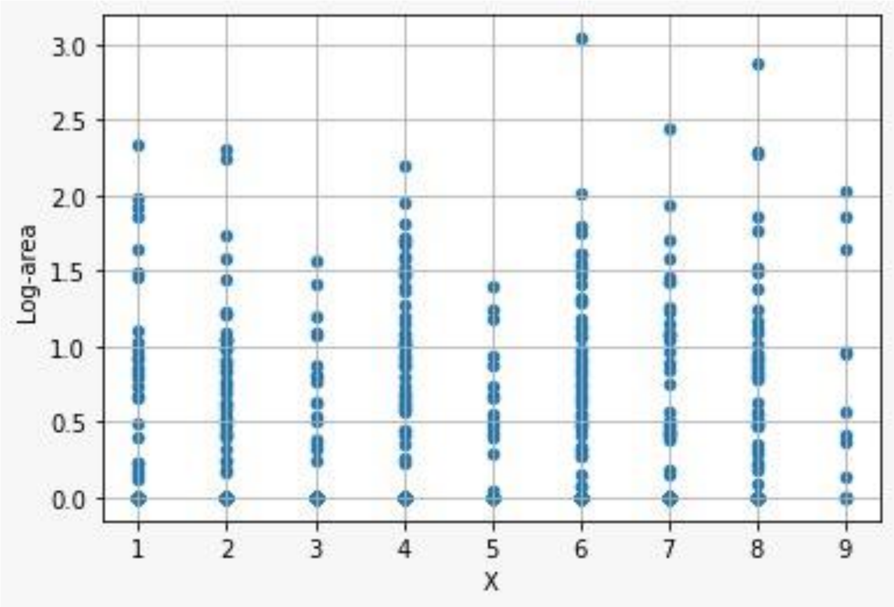
```
In [143]: #change month and day to numerrical values
          df = df.replace({'month':{'jan':1, 'feb':2,'mar':3,'apr':4,'may':5,'jun':6,'jul':7,'aug':8,'sep':9,'oct':10,'nov':11,'dec':12}})
          df = df.replace({'day':{'mon':1, 'tue':2,'wed':3,'thu':4,'fri':5,'sat':6,'sun':7}})
          df.dtypes
Out[143]: X          int64
          Y          int64
          month      int64
          day        int64
          FFMC     float64
          DMC      float64
          DC       float64
          ISI      float64
          temp     float64
          RH         int64
          wind     float64
          rain     float64
          area     float64
          dtype: object
```

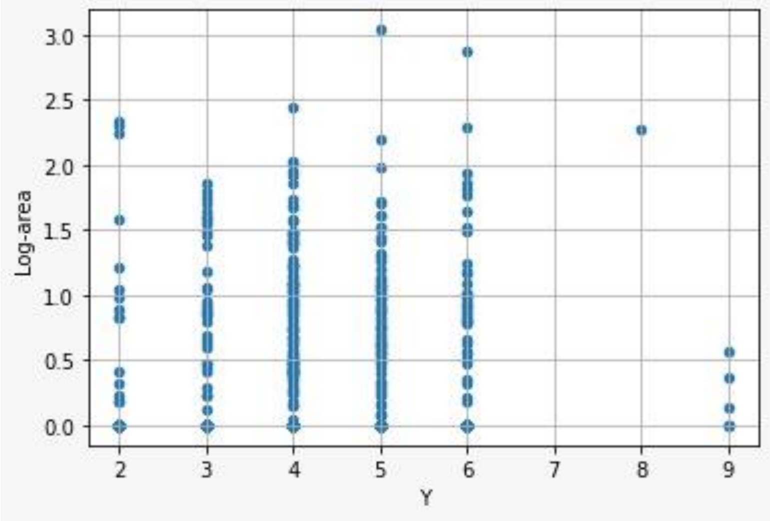
fitted.

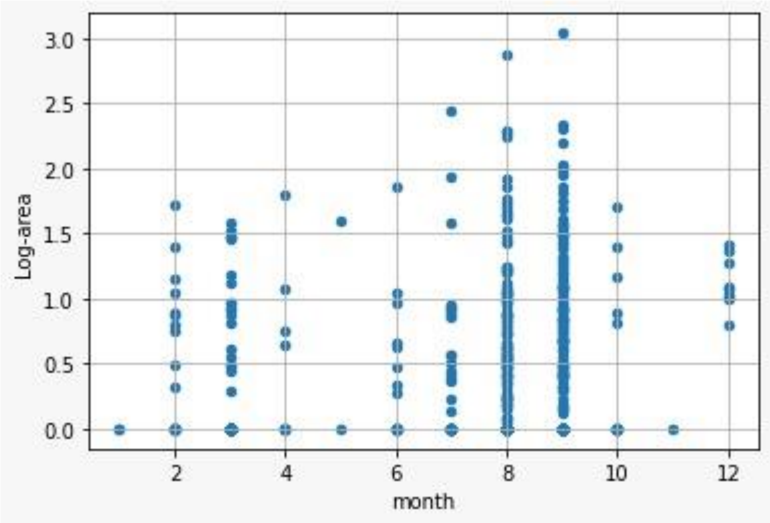
The DT node split was adjusted for the reduction of the sum of squares. After fitting the DM models, the outputs were postprocessed using the inverse of the logarithm transform. In few cases, this transformation may lead to negative numbers and such negative outputs were set to zero.

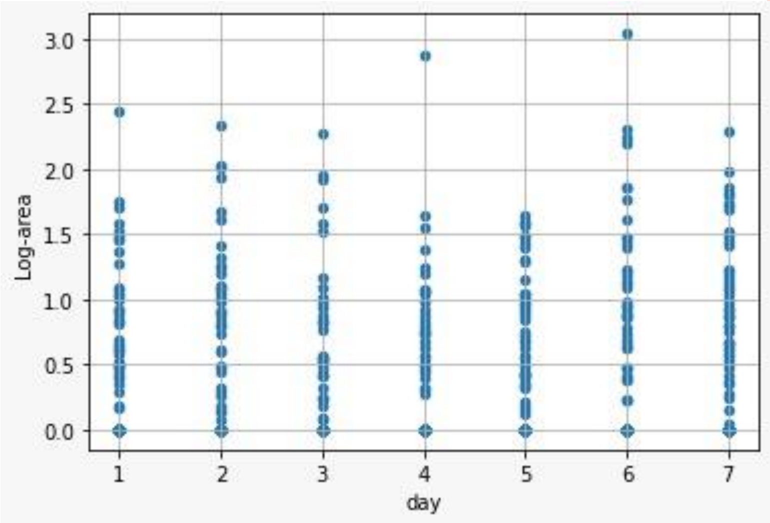
Data Analysis

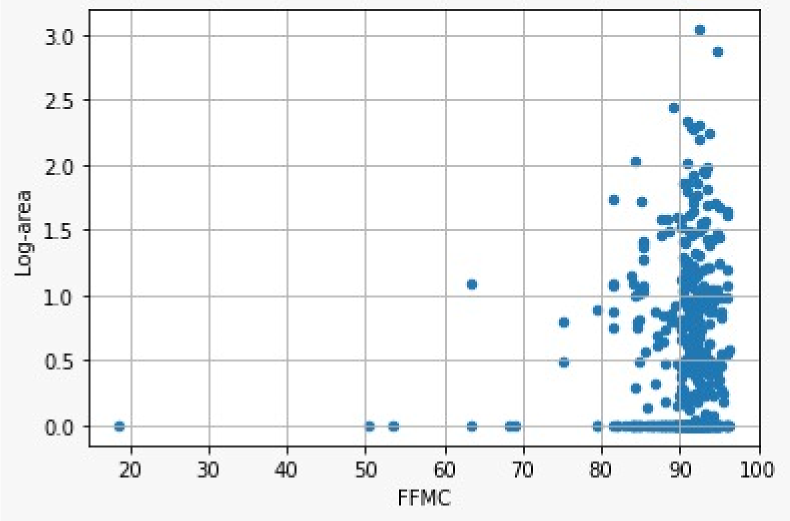
On analysing the data we found the following results

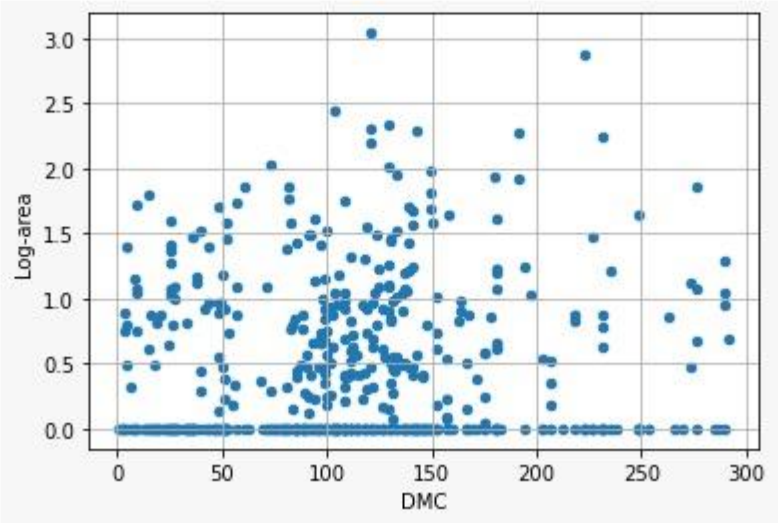


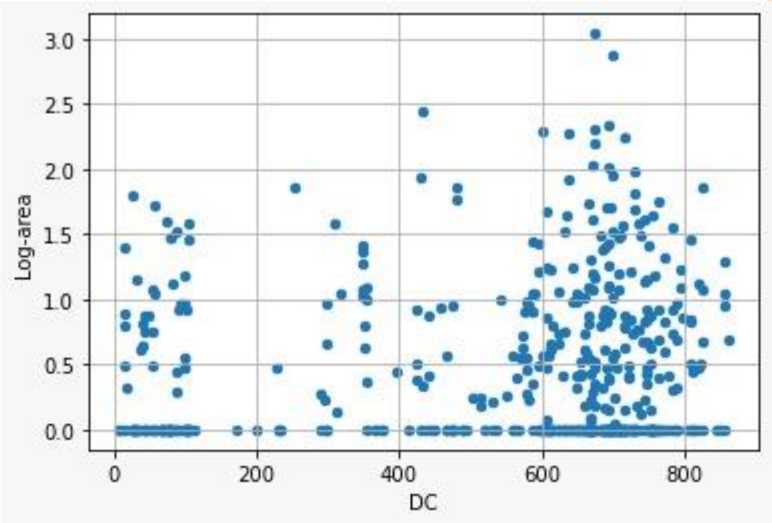

Log-area
Y

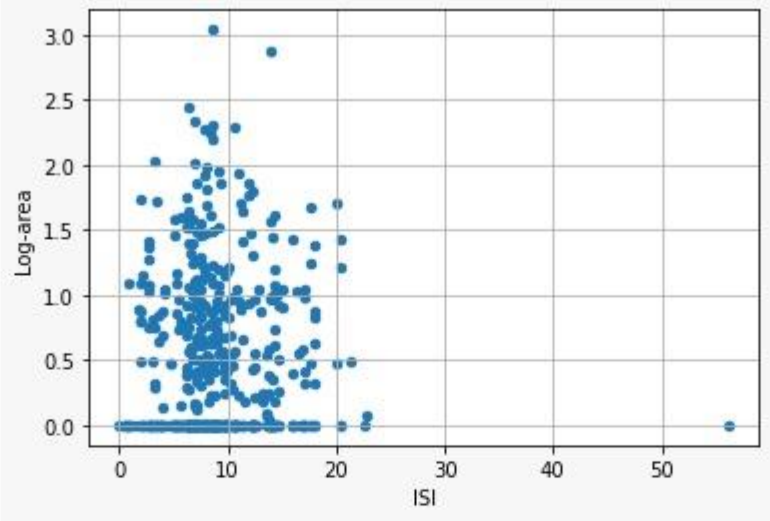

Log-area
month

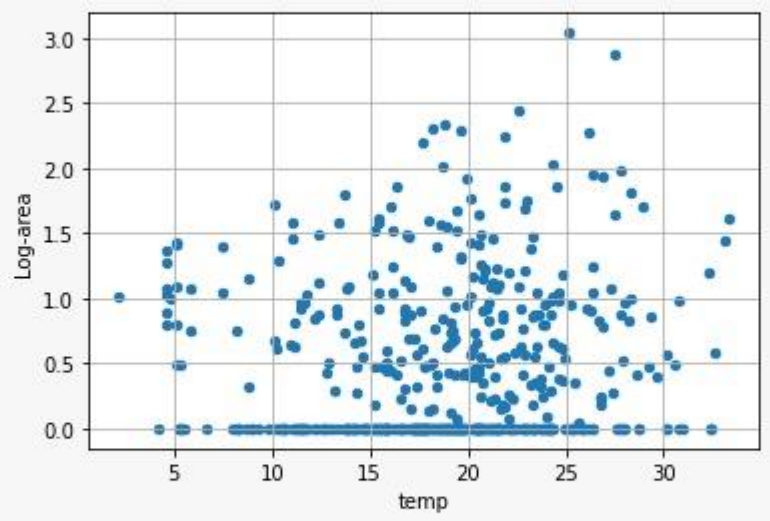

Log-area
day

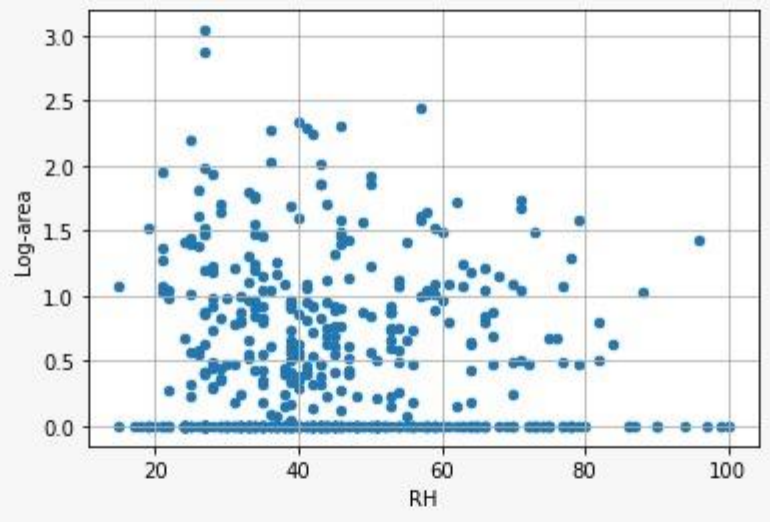

Log-area
FFMC
0.0
0.5
1.0
1.5
2.0
2.5
3.0
20
30
40
50
60
70
80
90
100

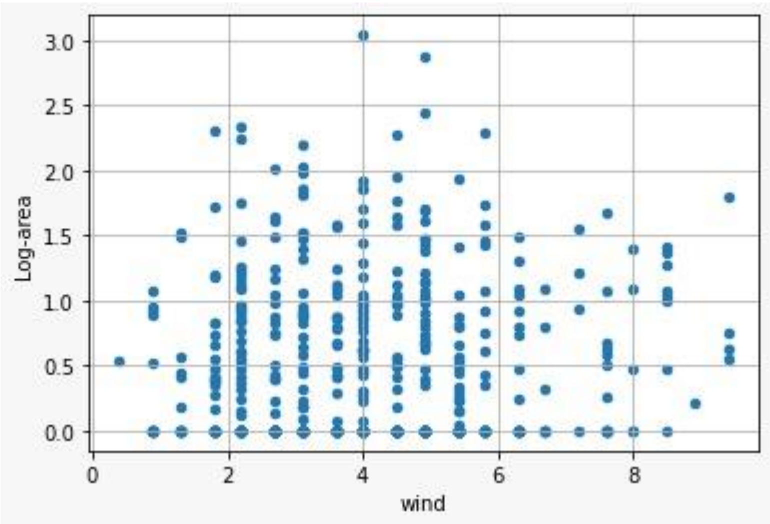

Log-area
DMC
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
50
100
150
200
250
300

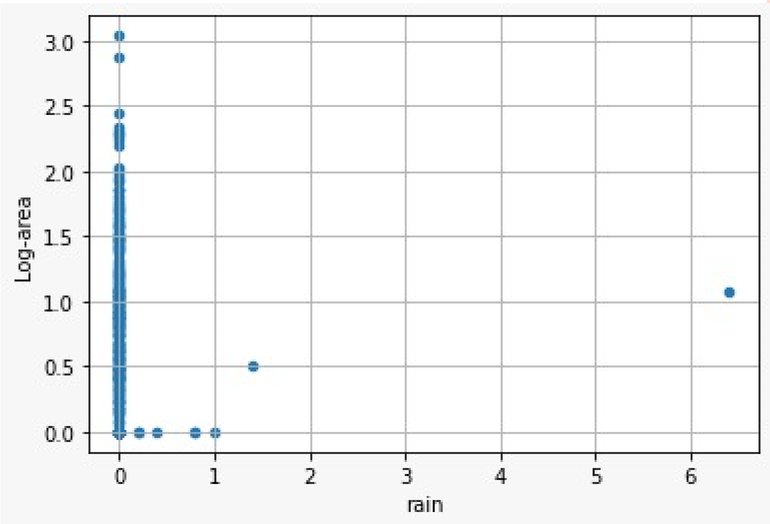

Log-area
DC
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
200
400
600
800


Log-area
ISI
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
10
20
30
40
50


Log-area
temp
0.0
0.5
1.0
1.5
2.0
2.5
3.0
5
10
15
20
25
30







We used Scatter plots to check how much it affects the decision attribute (area burned) as is clear from the plots above.

Comparing results of different algorithms

Results of different algorithms were compared based on RMSE

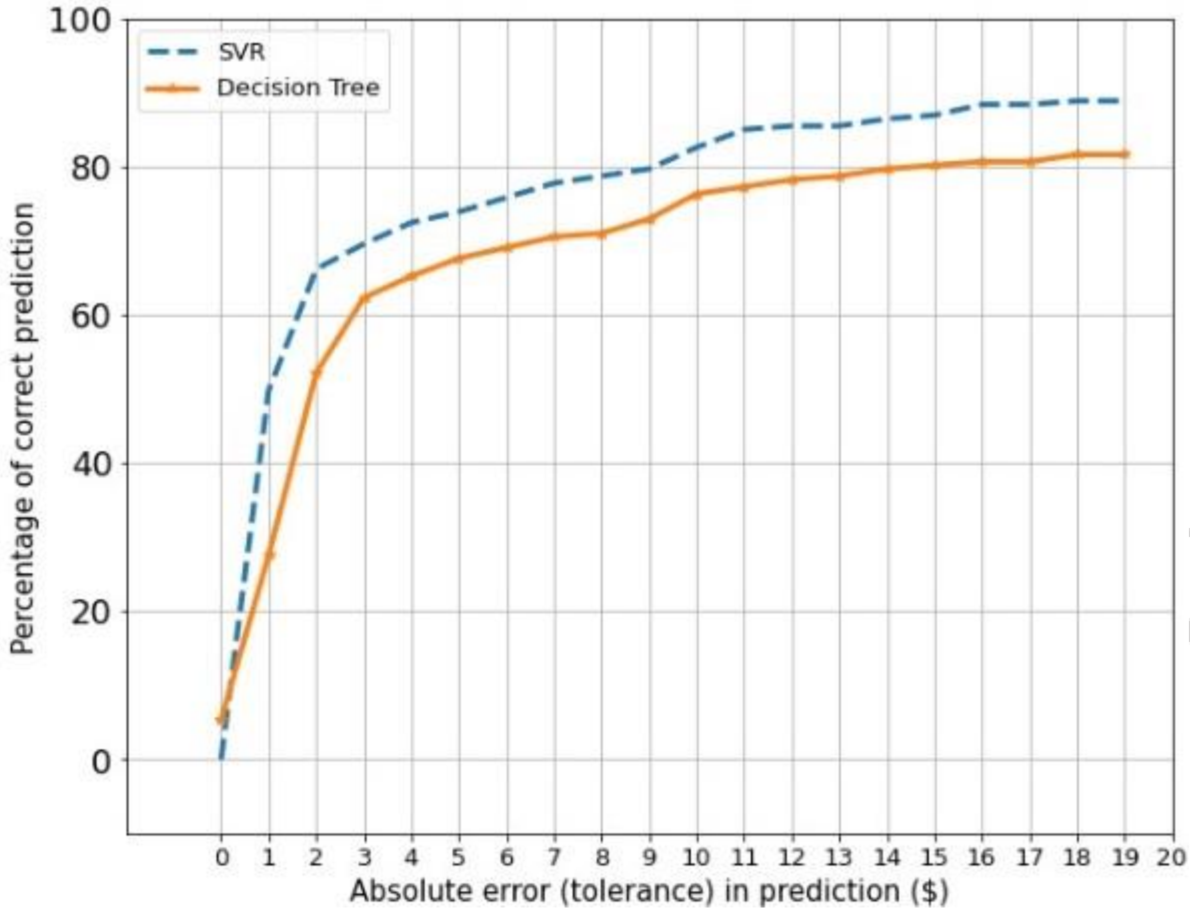
SVR has less RMSE than DT

Algorithm	RMSE
SVR	0.673
DT	1.023

Regression Error Characteristic (REC) Curve

REC curves were introduced in ML (Bi and Bennett, 2003) to compare the predictive ability of regression models.

A more detailed analysis to the quality of the predictive errors is given by using REC curves. From the REC analysis, the SVM is clearly the best solution, with the highest area.



Close to zero SVM has better percentage of prediction It is important to predict small fires i.e fires at early stage to avoid massive damage to forests and its wildlife

XVI. Conclusions

Forest fires cause a significant environmental damage while threatening human lives. In the last two decades, a substantial effort was made to build automatic detection tools that could assist Fire Management Systems (FMS). The three major trends are the us e of satellite data, infrared/smoke scanners and local sensors (e.g. meteorological). In this work, we propose a Data Mining (DM) approach that uses meteorological data, as detected by local sensors in weather stations, and that is known to influence forest fire s. The advantage is that such data can be collected in real-time and with very low costs, when compared with the satellite and scan ner approaches. Recent real-world data, from the northeast region of Portugal, was used in the experiments.

The database included spatial, temporal, components from the Canadian Fire Weather Index (FWI) and four weather conditions. T his problem was modeled as a regression task, where the aim was the prediction of the burned area. Two different DM algorithms, including Support Vector Machines (SVM), and one feature selections (using distinct combinations of spatial, temporal, FWI elem ents and meteorological variables) were tested. The proposed solution, which is based in a SVM and requires only four direct weat her inputs (i.e. temperature, rain, relative humidity and wind speed) is capable of predicting small fires, which constitute the major ity of the fire occurrences. The drawback is the lower predictive accuracy for large fires. To our knowledge, this is the first time th e burn area is predicted using only meteorological based data and further exploratory research is required.

As argued in, predicting the size of forest fires is a challenging task. To improve it, we believe that additional information (not ava ilable in this study) is required, such as the type of vegetation and firefighting intervention (e.g. time elapsed and firefighting strate gy).

Nevertheless, the proposed model is still useful to improve firefighting resource management. For instance, when small fires are p redicted then air tankers could be spared and small ground crews could be sent. Such management would be particularly advantag eous in dramatic fire seasons, when simultaneous fires occur at distinct locations. This study was based on an off-line learning, sin ce the DM techniques were applied after the data was collected. However, this work opens room for the development of automatic tools for fire management support. Indeed, in the future we intend to test the proposed approach by using an on-line learning envir onment as part of a FMS. This will allow us to obtain after some time a valuable feedback from the firefighting managers, in term s of trust and acceptance of this alternative solution.

Another interesting possibility would be the use of weather forecasts, in order to build proactive responses. Since the FWI system i s widely used around the world, further research is need to confirm if direct weather conditions are preferable than accumulated va lues, as suggested by this study. Finally, since large fires are rare events, outlier detection techniques will also be addressed.

XVII. References

1. B. Arrue, A. Ollero, and J. Matinez de Dios. An Intelligent System for False Alarm Reduction in Infrared Forest-Fire Detection. IEEE Intelligent Systems, 15(3):64–73, 2000.

2. J. Bi and K. Bennett. Regression Error Characteristic curves. In Proceedings of 20th International Conference on Machine Lear ning (ICML), pages 43–50, Washington DC, USA, 2003.

3. L. Breiman. Random Forests. Machine Learning, 45(1):5–32, 2001.

4. L. Breiman, J. Friedman, R. Ohlsen, and C. Stone. Classification and Regression Trees. Wadsworth, Monterey, CA, 1984.

5. V. Cherkassy and Y. Ma. Practical Selection of SVM Parameters and Noise Estimation for SVM Regression. Neural Networks, 17(1):113–126, 2004. 6. P. Cortez. RMiner: Data Mining with Neural Networks and Support Vector Machines using R. In R. Raje sh (Ed.), Introduction to Advanced Scientific Softwares and Toolboxes, In Press.

6 J. Bi and K. P. Bennett. *Regression error characteristic curves*. In Proceedings of the 20th International Conference on Machine Learning, 2003.

7. Chih-Chung Chang and Chih-Jen Lin. *a Library for Support Vector Machines*. http://www.csie.ntu.edu.tw/∼cjlin/papers/libsvm .pdf, 2009.

8.Paulo Cortez and Anibal Morais. *A Data Mining Approach to Predict Forest Fires using Meteorological Data*. Cambridge Univ ersity Press, Department of Information Systems R and D Algoritimi Centre, University of Minho, Portugal, 2007.