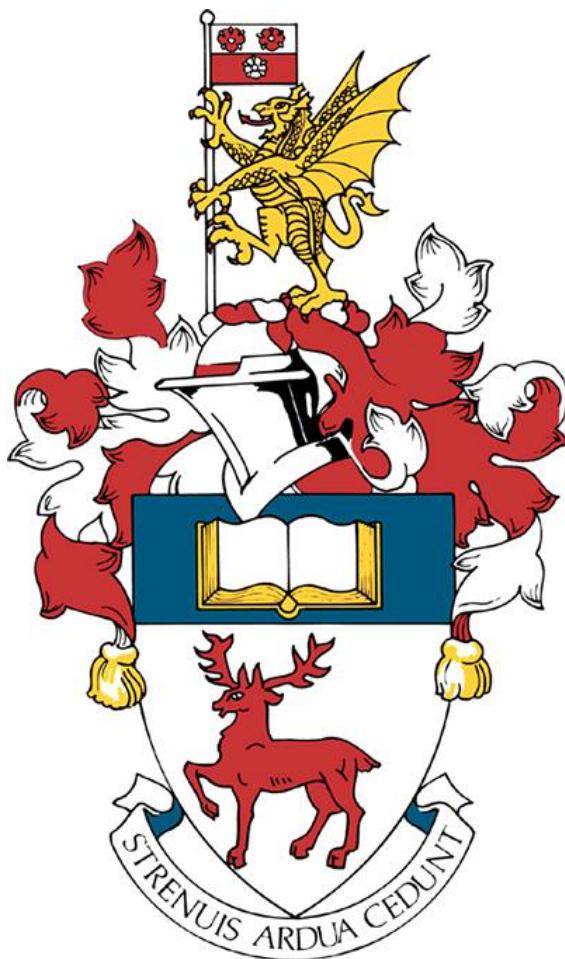


An investigation into wildfires in California in the month of August 2020



**Ben Steele
29 April 2021
Supervisor: Dr Davide Lasagna**

Word count: 9930

This report is submitted in partial fulfillment of the requirements for the Bachelor of Engineering, Aeronautics and Astronautics, Faculty of Engineering and Physical Sciences, University of Southampton.

Declaration

I, Ben Steele, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a degree at this University;
2. Where any part of this thesis has previously been submitted for any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission;

Acknowledgments

I would like to thank Dr Davide Lasagna, my project supervisor. His advice and generosity with his time are much appreciated.

Abstract

This report is an investigation into the wildfire siege that occurred in the state of California, USA in August 2020. This report has two main aims. Firstly, to quantify the extent to which this time period saw a higher incidence and/or larger area of wildfires than previous years, and secondly to investigate any possible correlation between fires in this time period and other environmental factors, such as air temperature, wind speed, and lightning strikes. Hopefully this project will shed some light on the events in California in August 2020 and help to provide an understanding of why they occurred. NASA Thermal Anomalies and Fire data is used to detect fire events. This data is analysed numerically in Python to calculate total fire area, total number of distinct fires, how long individual fires burn for, and how the fires interact with each other. This data is also displayed graphically to show how the wildfires overlap with areas that experienced higher than average temperatures, little or no precipitation, and high numbers of lightning strikes. This report shows that 2020 was a year of extremes. California experienced record breaking temperatures, a highly concentrated surge of lightning strikes, and huge wildfires. All of these emphasise the urgent need for governments and individuals to implement measures to protect the climate.

Contents

1	Introduction	1
2	Background	3
2.1	Relevant information	3
2.2	Literature review	5
3	Aims and objectives	10
4	Data sources and analysis	12
4.1	Data sources	12
4.1.1	NASA data	12
4.1.2	NOAA data	15
4.1.3	Lightning strike data	16
4.1.4	GIS data	18
4.2	Processing and analysis methodology	18
4.2.1	MOD14A1 data	19
4.2.2	MOD14A2 and MCD64A1 data	23
4.2.3	NOAA data	23
4.2.4	Lightning strike data	23
5	Results	25
5.1	Analysis of MODIS data	26
5.2	Analysis of NOAA data	34
5.3	Analysis of lightning strike data	37
5.4	Correlations	38
5.5	PCA of wildfire growth and meteorological factors	44
6	Conclusions and future works	46
Bibliography		49

1. Introduction

This report is an investigation into the trends in wildfires in California over the last six years. 2020 was a year of wildfire records for the state of California. The largest Californian fire in recorded history, the August Complex, started in August 2020. Of the six largest wildfires ever in California, five of them started in summer 2020 [1]. The wildfires that started in August continued to burn for weeks afterwards, and often reached their peak area in September. Wildfires in this period will be referred to in this report as August wildfires, as this is when they started. This does not, however, mean that their lifetime was limited to the month of August.



Figure 1.1: A photograph of a wildfire burning in California [2].

The inspiration for this report came from a number of different news reports that were published about wildfires in 2020 claiming that the year had been particularly bad in a number of different locations, notably California. Figure 1.1 is a photograph taken of a

wildfire in California in 2020. This image is shocking, and serves as a visceral reminder of the destructive force of wildfires.

Wildfires have a huge impact on many aspects of human life. They can disrupt transportation and communication infrastructure by destroying roads and telephone and internet cables. They can cut people off from utilities such as gas, electricity and water supplies, as well as health services. They can destroy crops, livestock, and property, leaving families with no possessions, no means for income, and no access to clean water or electricity [3]. Wildfires also cause harmful health effects. They release carbon monoxide, carbon dioxide and fine particulates into the atmosphere. These pollutants can cause respiratory and cardiovascular health problems. They also tragically can cause loss of life to both humans and wildlife. In 2020 there were 33 deaths as a direct consequence of wildfires in California [1]. Wildfires also accelerate global warming, by releasing greenhouse gases which cause the planet to heat up further, forming a positive feedback loop [3]. For all of these reasons and more, society has lots to benefit from an improved understanding of how wildfires occur. The wildfires in California alone cost the state over US\$ 2 billion in fire suppression [4] and US\$ 10 billion in damages [5]. Between 21 August and 1 October, approximately 2 million Californians were evacuated from areas in the path of oncoming wildfires [6]. Across the year as a whole 10,488 structures were damaged or destroyed [1].

A number of data sources are used throughout this report. MODIS data from the NASA Terra satellite is gathered to analyse the fires directly. The satellite detects fires on the ground at a 1km^2 resolution. Additional meteorological data is collected from the National Oceanic and Atmospheric Administration (NOAA). This contains data on precipitation, evaporation, and air temperatures across the state. Lightning strike data is collected from Vaisala, a third party company which supplies NOAA. The meteorological and lightning strike data are gathered for the purpose of comparing and correlating with the lightning strike data to identify any possible trends. These data will be analysed numerically in Python and used to quantify the number of wildfires experienced by the state of California. The data will also be visualised with the Geographical Information System (GIS). The analysis performed within this report will highlight possible causes for the large number of wildfires in California in 2020.

There is little doubt that the fire siege experienced in California in 2020 was due ultimately to climate change [7] which threatens all life on earth, and requires our immediate and devoted attention. Scientific research and engineering innovation are crucial to combat the affects that humans have had on the climate.

This document is organised as follows: Chapter 2 contains review of the relevant literature and information available regarding wildfires in California in 2020. Chapter 3 states the aims and objectives of this report. Chapter 4 discusses the selection, collection and the methods used to analyse each of the data series. Chapter 5 presents the results of this analysis. Chapter 6 concludes this report and identifies possible future works to be carried out. All data used in this report (save for the proprietary lightning strike data), and the Python code used to process it are available at: <https://sotonac.sharepoint.com/teams/AninvestigationintowildfiresinCaliforniainthemonthofAugust2020>.

2. Background

the chapter contains a review of the relevant literature and information available regarding wildfires in California in 2020. Chapter 2.1 contains background information from sources such as news reports and the California Fire Department which highlight the impacts of the fire siege in August. Chapter 2.2 contains government reports and data which provide more detail on the extent of, and damage caused by, the wildfires. This review of the relevant information and available literature determined the data that needed to be collected in order to achieve the aims and objectives of this report.

2.1 Relevant information



Figure 2.1: A GIS map of the United Kingdom with the August Complex wildfire overlaid in red across the south of England.

California made the news across the globe in 2020 as it experienced devastating wildfires over the summer months. For the first time in modern history, a wildfire in California burned an area exceeding one million acres (4047km^2), making it the first ever ‘gigafire’ [8]. This fire was the August Complex, shown in Figure 2.1 overlaid on the United Kingdom. The fire burnt an area greater than Hampshire.

Rank	Fire	Start Date	Area (km ²)
1	August Complex	August 2020	4179
2	Mendocino Complex	July 2018	1858
3	SCU Lightning Complex	August 2020	1605
4	Creek Fire	September 2020	1528
5	LNU Lightning Complex	August 2020	1470
6	North Complex	August 2020	1290

Table 2.1: The largest wildfires in recorded history in California, ranked by total area covered [1].

The California state fire service — CalFire — has an extensive database of wildfires within the state. From Table 2.1 of the six largest Californian wildfires ever recorded, four of them, including the largest one the state has seen, started in August 2020, and another started in September 2020. The database lists the cause of each fire as determined by CalFire, however, at the time of writing this report, the causes of the fires in Table 2.1 are still under review. The fact that so many large wildfires were started in such a short period of time is noteworthy, and invites further investigation.



Figure 2.2: Dry and cracked land emblematic of the drought in California [9].

CalFire identify the time period from 15 August through to the end of the month as a ‘fire siege’ [6], meaning they experienced an abnormally large number of new fires started. During this period the state of California was experiencing record breaking temperatures, low precipitation, and varying levels of drought. These factors indicated that the area was primed for ignition [6]. Figure 2.2 shows the kind of conditions the state was suffering from in August 2020.



Figure 2.3: Lightning activity over the Bay Bridge, San Francisco, California [10].

However, these factors are not necessarily causes themselves. They are merely factors which can exacerbate the situation and allow wildfires to start more easily, grow more quickly, and spread over a larger area. A CalFire report on the fire siege notes that the state suffered from over 15,000 lightning strikes in the space of just a few days. The report suggests that these lightning strikes may have been the initiator of the wildfires, which coupled with the 'extremely dry conditions' allowed fires to start and build quickly [6]. Figure 2.3 shows some lightning activity around the time of the fire siege in California.

Tropical storm Fausto is considered to be responsible for causing this surge in lightning strikes. The storm brought in a plume of moisture which fuelled thunderstorms in the area, and was characterised by low rainfall, high winds, high temperatures, and a high frequency of dry lightning strikes — lightning strikes not accompanied by rainfall. These are all prime conditions for wildfires, which could potentially explain the wildfire statistics for this time period [11]. A NASA blog [12] confirms that tropical storm Fausto was indeed devoid of any heavy rainfall, allowing its lightning strikes to hit the ground and start fires unabated by any precipitation.

2.2 Literature review

It was suggested in the news report discussed in the Chapter 2.1 that there were many environmental factors which contributed to the fire siege in August. The state was reported to be suffering from an extremely hot and dry August. The US Government have a few different data sources which can be used to evaluate the claims previously stated.

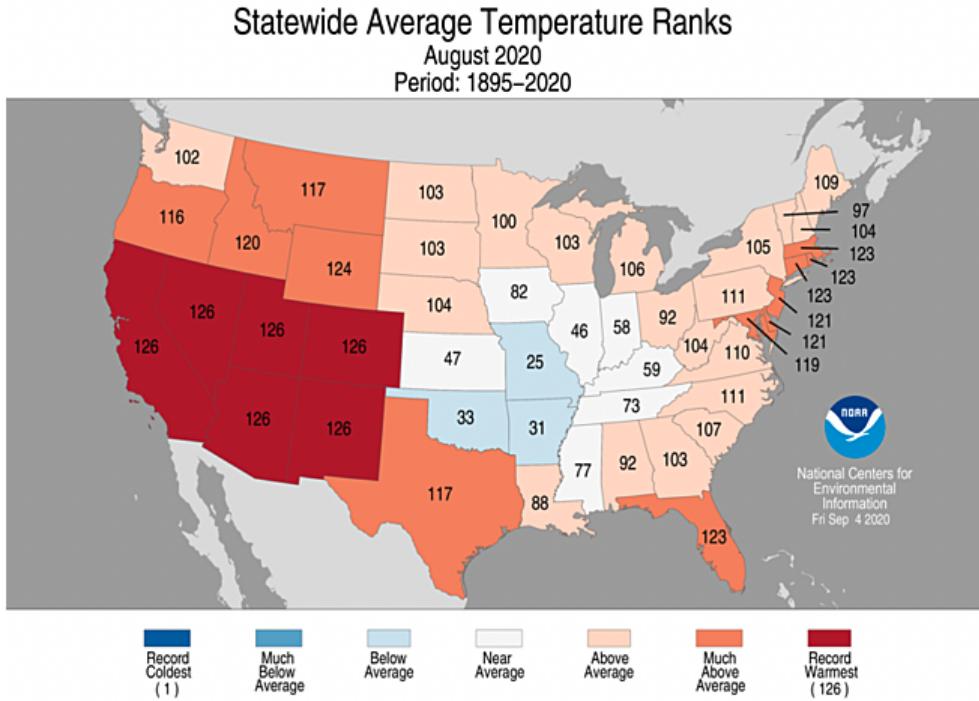


Figure 2.4: NOAA statewide average temperature rankings for August 2020 for the contiguous United States (CONUS) [13].

Figure 2.4 shows that many of the western states experienced their hottest August for over 100 years. The trends seen in Figure 2.4 are also seen in a similar manner for maximum and minimum temperatures. The fact that California experienced these record breaking temperatures at the same time that it experienced record breaking wildfires provides a good basis for an investigation.

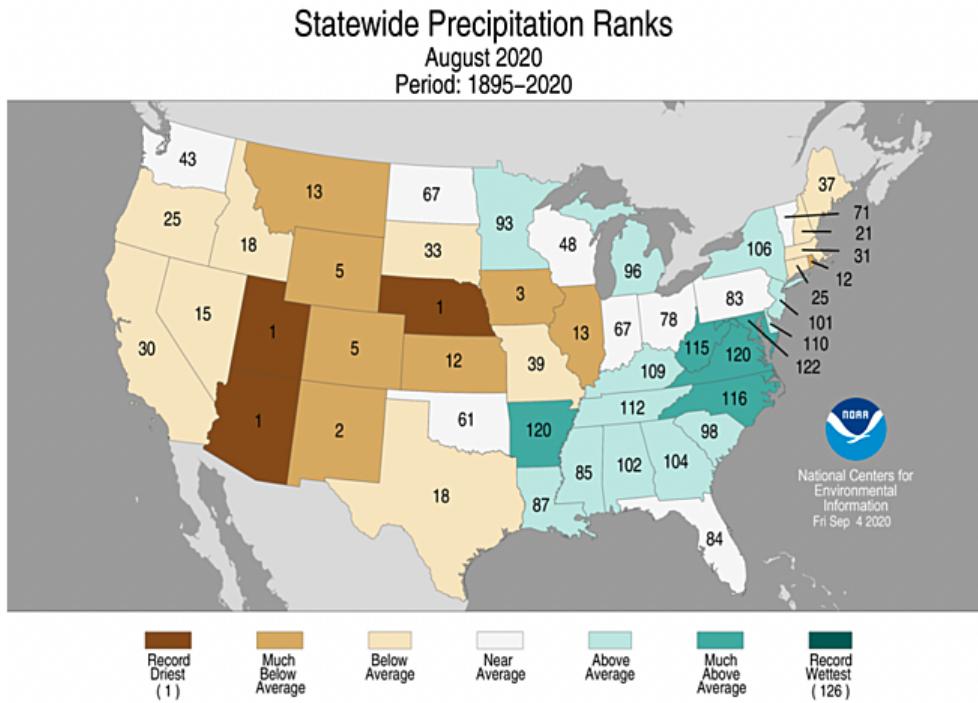


Figure 2.5: NOAA statewide precipitation rankings for August 2020 for CONUS [13].

A similar map of precipitation data in Figure 2.5 shows that the state experienced lower than average precipitation in August, with the month ranking in the bottom quartile for the 126 years on record.

Low precipitation creates a dry environment in which fires can start easily and spread quickly. Prolonged dry periods can lead the area to suffer drought. These factors can affect both how easy it is for a wildfire to start and how hard it is for firefighters to combat it. Low rain can lead to little reservoir water available to use to combat wildfires.

Low precipitation affects the growth of wildfires in many ways. Water that has fallen as precipitation can stay on the ground for different lengths of time based on the type of landscape it falls on and the vegetation present. Air temperature, pressure and humidity can all affect how quickly water evaporates from the ground and transpires from plants, altering how dry the potential wildfire fuel is.

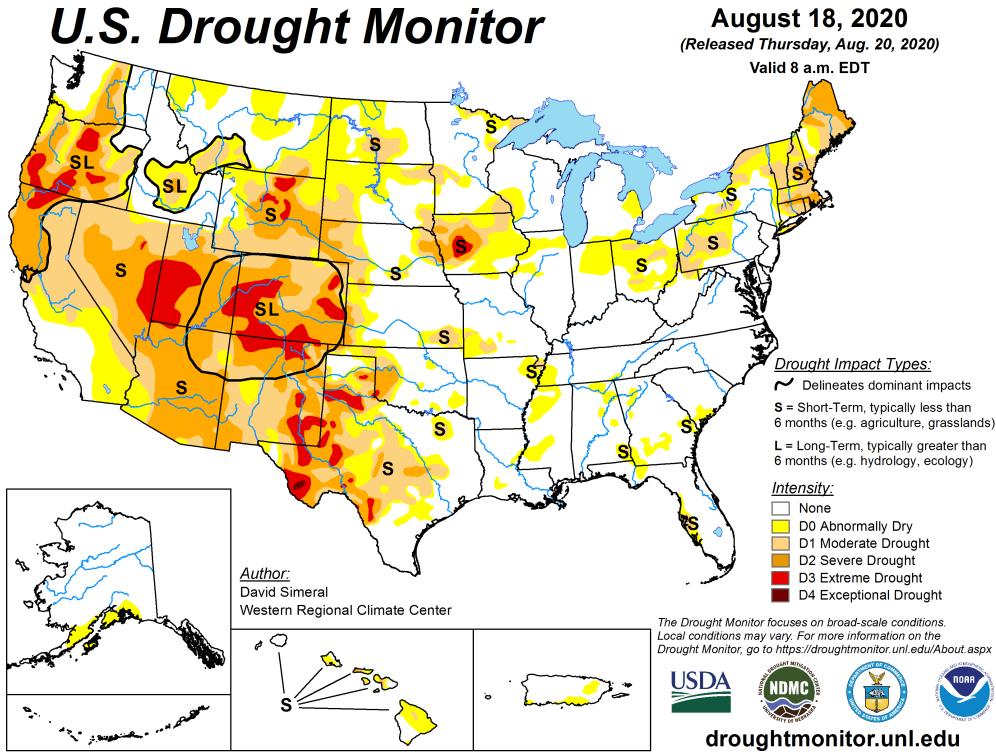


Figure 2.6: US Drought monitor map of drought classifications for United States territories for the week of 18 August 2020 [14].

As California was experiencing little precipitation in August, the state was simultaneously suffering from various levels of drought. As Figure 2.6 shows, the northern parts of the state were classified as in a state of either moderate, severe, or extreme drought.

The National Interagency Fire Center (NIFC) has data available for the causes of wildfires. These data highlight the differences between those caused by lightning strike and those caused by human activity. Humans are responsible for roughly 10 times as many wildfires as lightning strikes, however they both account for roughly equal land areas [4]. This could be due to a number of different reasons. It is possible that wildfires that start as a result of human activity are more quickly reported, and are therefore extinguished before they have time to encompass a large area. They may also start in more easily accessible areas, allowing firefighters to combat the fires more effectively. They may start in areas with less flammable material to fuel them therefore posing an inherent limit on the size they can become. These data serve to show that while the majority of wildfires are started by human activity, lightning strike initiated wildfires have the potential to grow to a much larger size, and therefore cause more damage to the environment on a per-fire basis. Human activity as a primary cause of ignition of wildfires is beyond the scope of this report, and will therefore not be considered in the following evaluation and analysis.

The literature and information examined within this chapter shows that August 2020 in California is a time period which requires further investigation. In the following chapters of this report satellite data is used to understand the extent of the wildfires which burned in this period. To fully understand the events that occurred in California, and to possibly identify times in the future where they may occur again, other meteorological factors, such as air temperatures, precipitation, and lightning strikes are examined in the context of the wildfires to determine how these factors influenced the wildfire season in California in 2020.

3. Aims and objectives

This chapter states the aims and objectives of this report. The aim of this report is to provide insight into the events that occurred in California around August 2020 and to identify possible phenomena that caused them. This will be achieved through numerical analysis and comparison of a number of different data sets regarding both wildfires themselves, and other relevant meteorological data. The objectives of this report are presented below:

- Gather relevant wildfire data for the years 2015-2020 inclusively where possible.
- Analyse the MODIS data to determine the total area of fire on each day.
- Calculate the total number of fires burning each day.
- Calculate the mean area of an individual fire on a given day.
- Create an algorithm which groups fires together based on proximity
- Ascertain the lifetime — how long a fire burns for — for each fire group on each day.
- Create an algorithm that classifies fires as either simple or complex fire groups
- Investigate how fires interact with each to produce a network plot of fire activity across California.
- Gather state and county boundary and land cover to aid in the presentation of the GIS maps.
- Display the wildfire data graphically using GIS to show how wildfires are distributed across the state of California.
- Gather additional relevant meteorological data for 2015-2020 inclusively where possible to analyse how 2020 compares to the five previous years and to investigate correlation with the wildfire data. The data includes:
 - maximum daily air temperatures
 - minimum daily air temperatures

- average daily air temperatures
 - precipitation
 - evaporation
 - lightning strikes
- Analyse lightning strike data to separate cloud-to-cloud lightning strikes from cloud-to-ground strikes.
 - Display the meteorological data in GIS on top of wildfire and land cover data to highlight any spatial correlations.
 - Combine the meteorological factors into one single GIS map to determine how they influence where wildfires burn.
 - Perform statistical analysis, namely principal component analysis, to determine which meteorological factors potentially influence the growth of wildfires.

This project will quantify the number and area of wildfires in California in 2020, compare this to previous years as well as other meteorological data to examine the claims that 2020 was a year which saw significantly more wildfires than in recent years.

4. Data sources and analysis

This chapter discusses the different data sources and analysis methods used in this report. Chapter 4.1 describes the different data sources used for the analysis in this report. It explains how the data is collected, the instrumentation used, and the organisations which produce the data. Chapter 4.2 describes and explains the different numerical analysis methods used to accomplish the various aims as described in Chapter 3.

4.1 Data sources

The different data sources and their respective forms and collection methods are discussed in this chapter.

4.1.1 NASA data

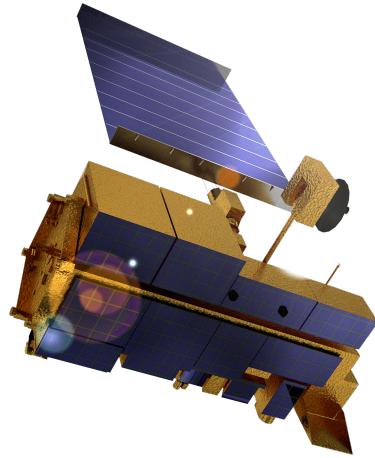


Figure 4.1: NASA Terra satellite computer image render [15].

The primary data source for this report is the NASA Terra satellite shown in Figure 4.1. The Terra satellite is the ‘EOS [earth observation system] flagship’ [16]. It was the first satellite to look at Earth system science. NASA says the satellite ‘explores the connections

between Earth's atmosphere, land, snow and ice, ocean, and energy balance to understand Earth's climate and climate change and to map the impact of human activity and natural disasters on communities and ecosystems' [16]. This description shows that the satellite is a perfect candidate for gathering data on wildfires.

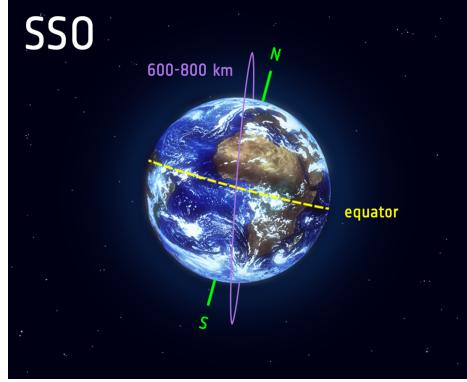


Figure 4.2: Diagram of a sun-synchronous orbit [17].

Terra is placed in a sun-synchronous orbit (SSO). Figure 4.2 shows the anatomy of an SSO. This orbit has unique benefits which make it ideal for earth observation missions. It is a near polar orbit, which allows the satellite to pass directly over a large percentage of the surface of the Earth. The inclination of the orbit causes it to precess exactly once a year. This means that the satellite passes over each point on the Earth's surface at the same local solar time, meaning it passes over every point on the Earth's surface at exactly the same time of day, making any data gathered by the satellite directly comparable to previous data collected at the same location.

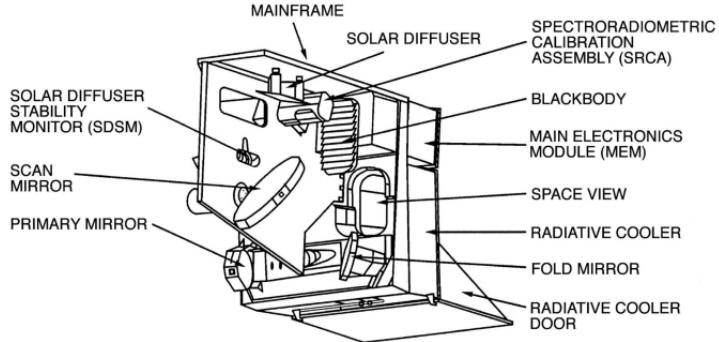


Figure 4.3: Diagram of the design of the MODIS instrument on board the NASA Terra satellite [18].

The Terra satellite is fitted with the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument shown in Figure 4.3. MODIS captures data in 36 discrete spectral

bands, from $0.4\mu\text{m}$ to $14.4\mu\text{m}$, which includes the visible and infrared bands of the electromagnetic spectrum. MODIS operates at three resolutions, 250m, 500m, and 1km. All the data used in this report are from the 1km resolution. The sensor rotates $\pm 55^\circ$, which at the Terra satellite's altitude of 705km, gives a swath width of 2,330km [19]. This swath width, combined with Terra's SSO allows the instrument to image the entire surface of the Earth every one to two days.

NASA supply three data products from the Terra satellite which are used in this report. These are MOD14A1, MOD14A2, and MCD64A1. The MOD14 data are part of the Thermal Anomalies and Fire data series. MOD14A1 is a single day product, whereas MOD14A2 is an eight day summation. The data is sampled in 1km^2 tiles which are assigned a value based on a key, corresponding to land, water, fire etc. These tiles make up a layer called the 'FireMask'. The series also comes with a quality assurance (QA) layer. MCD64A1 is part of the Burned Area product series, a monthly measurement of the total land area burned during that month. Like the MOD14 data, it is sampled in 1km^2 tiles, however the values assigned to these are more complex. The MCD64A1 data comes with more layers, including start date and end date of the burning in the square, as well as a QA layer. All three of these products are used in the analysis in this report.

The MODIS instrument measures whether a fire is present in an approximately 1km^2 grid and flags it accordingly. The accuracy of this observation is recorded as low, nominal or high confidence. The minimum size of fire the instrument can detect depends on a number of factors, such as 'scan angle, biome, sun position, land surface temperature, cloud cover, amount of smoke and wind direction etc.' [20]. NASA state that the instrument routinely detects fires 1000m^2 in size, and under 'very good observing conditions (e.g. near nadir, little or no smoke, relatively homogeneous land surface, etc.)' the instrument can detect fires 100m^2 in size [20].

The MODIS data are available to download directly from the NASA Earthdata Search website in hierarchical data format — Earth Observation System (HDF-EOS). From this site, the required data sets and temporal range are selected. To cover the entire land area of California two data sets are required, h08.v04 and h08.v05. The MOD14A1 data is sourced from this site. The site provides a UNIX script for downloading this data, which was used for this project as a large amount of data was required. The MOD14A2 and MCD64A1 data are sourced from the LPDAAC *AppEEARS* program as georeferenced tif (GeoTiff) files. This program transforms the data from the native sinusoidal projection, to EPSG:4326 — WGS 84, which is a standard geographical projection. The fact that the data is available in this form makes it ideal for use with GIS to graphically represent the land area consumed by wildfires in California.

4	4	4	...	4	3	3	Value	Description
4	4	4	...	3	3	3		3 Non-fire water pixel
4	4	4	...	3	3	3		4 Cloud (land or water)
			...					5 Non-fire land pixel
4	4	4	...	5	5	5		6 Unknown (land or water)
4	4	4	...	5	5	5		7 Fire (low confidence, land or water)
4	4	4	...	5	5	5		8 Fire (nominal confidence, land or water)
								9 Fire (high confidence, land or water)

Table 4.1: A sample of MOD14A1 data as shown in tabular format when access with the PyHDF library, on the right is the pixel key which explains the meaning of each pixel value.

Table 4.1 shows a sample of some MOD14A1 data (MOD14A2 data is organised in the exact same way, the only difference being the temporal range). The data is arranged as a 1200x1200 grid, with each value point representing a 1km² grid square.

For the purposes of this report, fire pixels of all confidence levels will be considered to represent areas that contain active fires. This seems to be a fair assumption to make, as it considers a worst case scenario for the total area of fire. It is kept constant across all samples. The factors which have the most effect on the accuracy of this assumption, latitude, longitude, and viewing angle, are also all kept constant as the data is collected from the same satellite of the same location, giving each image the same detection accuracy. Typically the highest confidence levels are seen in the centre of a fire, with the lower confidence levels seen around the edges. This is due to the fact that as the area of fire within each grid square decreases, so too does the probability of the fire being accurately detected [21].

4.1.2 NOAA data

The National Oceanic and Atmospheric Administration (NOAA) have a wide range of available meteorological data which will be useful for the purposes of this report. The data is recorded at multiple different weather stations across the US and is available as daily summaries which are used in this report. Data is collected for the years 2015-2020 inclusively, both for the state of California as a whole, and for a few individual counties in the state.

Data product code	Product description
TAVG	Average temperature
TMAX	Maximum temperature
TMIN	Minimum temperature
EVAP	Evaporation of water from evaporation pan
PRCP	Precipitation

Table 4.2: NOAA Data products and product codes sourced for this report.

The NOAA data is available to download from the NOAA Climate Data Online Search, which allows for selection of specific data products, temporal ranges and locations. This data is available in comma separated value (CSV) files. The different data products used in this report are listed in Table 4.2.

STATION	NAME	DATE	EVAP	PRCP	TAVG	TMAX	TMIN
USR0000CTHO	THOMES CREEK CALIFORNIA, CA US	2020-07-01			27.9	35.0	20.6
USR0000CTHO	THOMES CREEK CALIFORNIA, CA US	2020-07-02			27.4	33.9	20.6
USR0000CTHO	THOMES CREEK CALIFORNIA, CA US	2020-07-03			25.6	32.2	17.2
USC00041253	CACHUMA LAKE, CA US	2020-07-01	4.8	0.0		30.6	10.6
USC00041253	CACHUMA LAKE, CA US	2020-07-02	5.6	0.0		27.8	10.0
USC00041253	CACHUMA LAKE, CA US	2020-07-03	9.4	0.0		30.0	8.9

Table 4.3: A sample of the NOAA meteorological data in a table generated from the native CSV format. The blank spaces in the table are due to the fact that not all stations record all the data products.

A sample of some NOAA data is shown in Table 4.3. The data is listed by the weather station at which it was recorded (stations have a NOAA code in the station column and a location name in the name column). This data is collected for all of the available weather stations across California, and the data is aggregated and processed according to the methodology in the following section.

4.1.3 Lightning strike data

NOAA do not directly collect any data regarding lightning strikes themselves, rather, they source their data from Vaisala, a third party company. The data for this report is courtesy of Vaisala, through their research data program. The program supplies free data to researchers, for lightning strikes over a maximum of a five year period, and an area not exceeding 700,000km². Therefore, data was collected for the years 2016-2020 inclusively, for an area covering most of California. Vaisala collect data through their National Lightning

Detection Network (NLDN) which records the date, time, and latitude and longitude of the lightning strike. The data also includes the charge of the lightning strike and a tag to distinguish between in-cloud and cloud-to-ground lightning.

Timestamp	Latitude	Longitude	Charge	Tag
2016-01-03 02:48:14.646	34.0868	-122.5904	+23.2	G
2016-01-03 17:57:37.503	33.0695	-122.7170	+16.2	C
2016-01-03 17:57:37.504	33.0736	-122.7192	+13.9	C
2016-01-03 17:57:37.544	33.0546	-122.7587	+46.8	G
2016-01-03 18:06:15.501	33.1843	-122.6568	+64.3	G

Table 4.4: A sample of the lightning strike data provided by Vaisala (column headings have been added for clarity and are not present in the native ASCII data). The tag column indicates whether the lightning strike was cloud-to-cloud (C) or cloud-to-ground (G).

Table 4.4 shows a table created from a sample of the ASCII data provided by Vaisala for lightning strikes in the state of California. The lightning data is supplied in American Standard Code for Information Interchange (ASCII) format.

4.1.4 GIS data



Figure 4.4: The base GIS map onto which all the wildfire and meteorological is overlaid. The map contains the California shapefile, land cover data (areas of darker green represent vegetation cover), and large place names.

Some additional data is sourced to supplement the GIS maps created for this report. Shapefiles for the state of California and its counties are sourced for display purposes and to allow for clipping of the wildfire and meteorological data. This data is downloaded from the respective state and county government data websites. Large cities are also added to the map to give context to the locations of the fires. Land cover data is sourced to aid in the analysis of the maps, showing areas of vegetation — potential wildfire fuel. The land cover data used in this report is sourced from the United States Geological Survey (USGS). The additional data gathered for the GIS maps is supplied as georeferenced shapefiles requiring no further processing. Figure 4.4 shows the base layers used in all GIS maps in Chapter 5.

4.2 Processing and analysis methodology

This section details the methods used to analyse each data set. The different analytical processes are discussed in the order in which the results are presented in Chapter 5.

4.2.1 MOD14A1 data

The MOD14A1 data is loaded into Python using the PyHDF library. Each file contains eight 'FireMask' layers, one for each day, and each one is selected and loaded into a Numpy array. For each day there are two arrays, from the two sample data sets that are needed to include the whole land area of California. These arrays are vertically stacked to form a single array that captures the entire area of California. From this some basic analysis can be carried out immediately.

These arrays can be loaded as images using Matplotlib, giving a visual insight into the state of wildfires on each day. Figure 4.5 shows one of these images. However, the data in this form are not particularly suited for identifying the locations of the fires as they are shown in the native sinusoidal projection of the MODIS instrument. The HDF files do not come with any projection or latitude/longitude information, so these images cannot be transformed easily in Python. For visual analysis of wildfires the MOD14A2 data is used as it is available as a perfectly georeferenced GeoTiff.

By simply counting the instances of the different fire pixels in the array, the area of land burning can be calculated. In the case of the MODIS data this is very straightforward as each pixel represents an area of 1km^2 . For the analysis performed in this report the arrays are transformed such that fire pixels of all confidence levels are represented by a '1', and all other pixels are represented by a '0'. As discussed in Chapter 4.1.1 the accuracy of this simplification varies based on a number of factors, but it is the most appropriate choice for this report.

Following on from this, a function is written to calculate the cumulative area consumed in fire throughout the year. This function is written such that a fire pixel is only counted if on the previous day the same grid square did not contain a fire. This prevents fires that exist for multiple days in the same location from being counted multiple times.

The next level of analysis performed on these arrays is to identify the number of distinct fires present in each day. This is done with the use of a flood fill algorithm.

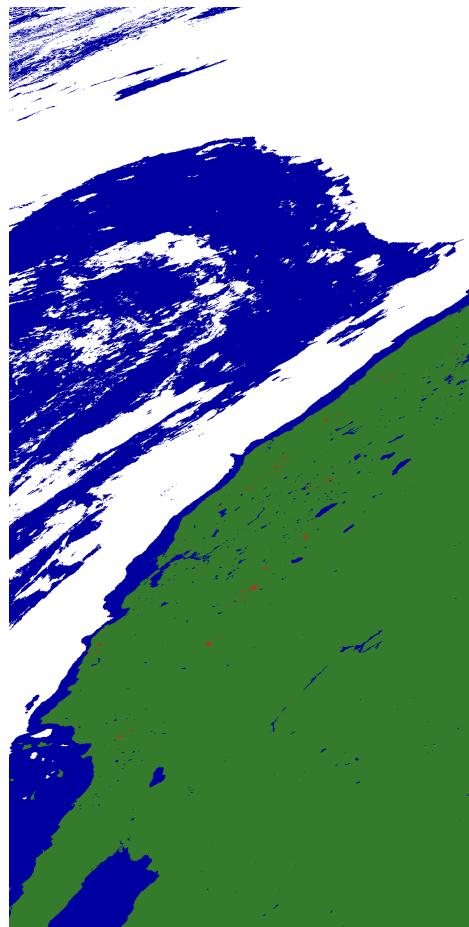


Figure 4.5: A 1200x2400 pixel image of California produced from the MOD14A1 data, wildfires burning are shown in red.

Algorithm 1Flood fill

```
if fire detected at  $x, y$  then
    if  $x > 0$  then
        run flood fill at  $x - 1, y$ 
    end if
    if  $x < \max(x)$  then
        run flood fill at  $x + 1, y$ 
    end if
    if  $y > 0$  then
        run flood fill at  $x, y - 1$ 
    end if
    if  $y < \max(y)$  then
        run flood fill at  $x, y + 1$ 
    end if
end if
```

Algorithm 1 shows the basic principle of the flood fill algorithm. Each grid square is checked to determine whether it contains a fire and, if it does, the surrounding grid squares are also checked. Any grid square checked by this function is marked as such, so that it is not counted more than once. This function is then run for every location in the array, unless it has already been checked. This returns a list of sets which define each distinct fire in each day. From this it is easy to extract the data concerning the number of distinct fires in each day. Once this data is known, it can be combined with the data previously mentioned to calculate the average area of a fire each day. From the fire set data, the area of individual fires can also be determined.

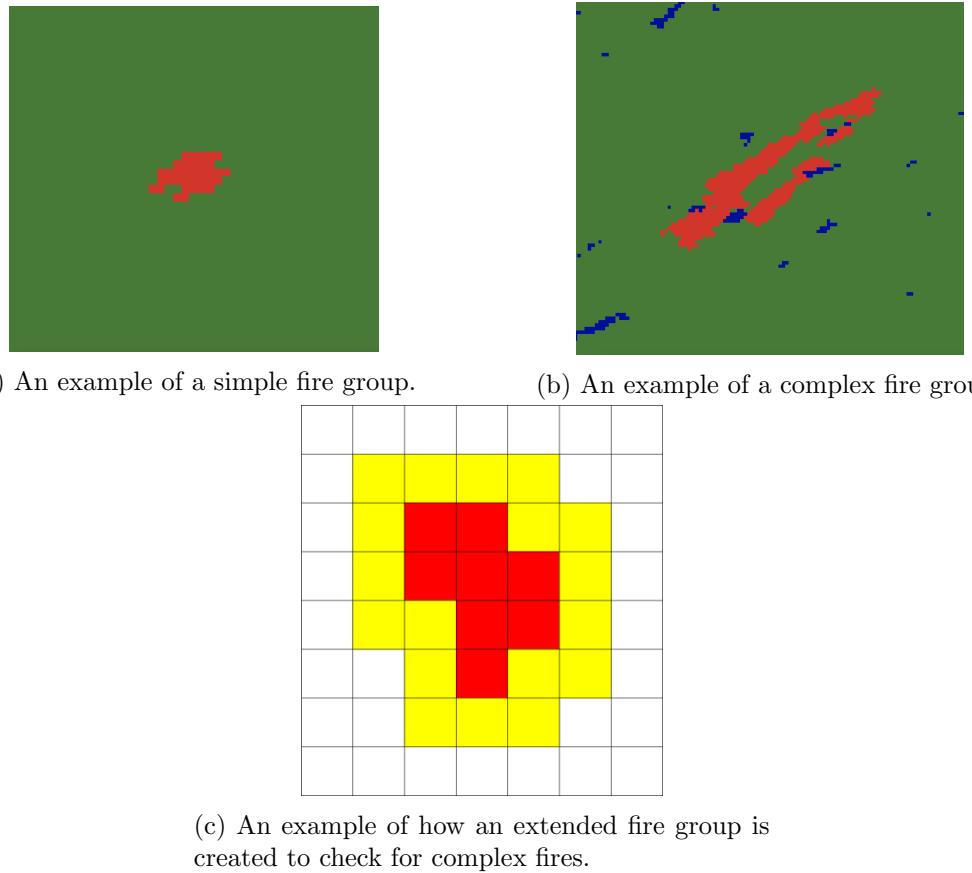


Figure 4.6: Examples images of simple and complex fire groups taken from a Matplotlib image of the MOD14A1 data for 2020 in California. Also shown is a representation of the grouping algorithm used to detect complex fires. Border points (yellow) are added to the existing fire set (red). This forms a new expanded group for each fire.

A distinction used throughout this report is that of simple and complex fires, examples of which are shown in Figures 4.6a and 4.6b. Derived from the classifications used by CalFire, a complex fire is one which consists of multiple fires in close proximity. With the fire set data compiled, these distinct fires can be condensed into complex fires; groups of fires close together in space and time. The function which generates these groups does so by grouping together all fires that are within one grid square (approximately 1km) of each other. Figure 4.6c shows a graphical representation of how this algorithm works. This generates a new list of sets which defines all fire groups present in a day. The term fire group refers to any group of fire pixels which are either in contact with each other (simple), or separated by a single pixel (complex). This group can be subdivided to obtain lists of simple fires (groups of length one), and complex fires (groups of length greater than one). As with the fire set data, average and individual fire area can also be calculated for the fire groups.

Algorithm 2Fire group lifetimes

```

i ← 1
while i = 1 do
    for testset in firegroupsetlist do
        setlist ← []
        for firegroupset in next day do
            if testset ∩ firegroupset ≠ ∅ then
                setlist append firegroupset
                firegroupset ← marked as checked
            end if
        end for
    end for
    if setlist = [] then
        i ← 0
    else
        firegroupsetlist ← setlist
        next day ← next day + 1
    end if
end while

```

These fire groups can then be analysed in two main ways. Firstly, the lifetime of each fire group can be calculated. This is achieved through the use of a fire group lifetime tree algorithm, shown in Algorithm 2. Each fire group is compared to each fire group in the next day to determine if they overlap in geospatial location, indicating that they are the same fire. The function sets up the fire group sets as an array, with the days vertically stacked. A fire is selected from the current day, and checked against all the fires in the next day. The fires that match are then themselves tested against all the fires in the next day. This is done recursively, continuing for as long as the fire continues to overlap with new fires. This function then returns the start and end date of each fire, determining the lifetime of each fire group. This algorithm only calculates lifetimes of fire groups which overlap in geospatial location on consecutive days. It is of course possible that a fire could move significantly in the space of 24 hours such that it is not flagged as the same fire group by this algorithm, but this will likely only occur very rarely, making this algorithm sufficient for this report.

The lifetime analysis determines which fires split and which merge, to obtain a list of each fire group throughout the year. This analysis can also be used to plot a cumulative count of the number of distinct fire groups throughout the year.

Secondly, the fire group sets can be analysed to determine all of the fires with which they overlap in the following day, which identifies how fires merge and split. A function is written which returns a list of all the fires in the next day that overlap with the fire group being checked in the current day. This is run iteratively over all the fires, to collect the data of all fire group connections.

4.2.2 MOD14A2 and MCD64A1 data

The MOD14A2 data is used for visual analysis in QGIS. It is loaded as a raster layer and clipped with the California state shapefile. Then the values of the array are assigned colours, with fire pixels red, and all other pixels transparent. For the subsequent county analysis, these steps are repeated with the raster layer being clipped by the respective county shapefile.

The same process is followed for the MCD64A1 data. Both of these data are used to produce a mask which, when overlaid over the state of California, shows the land area covered in, or burned by, wildfires.

The masks produced for the MOD14A2 are saved as GeoTiff files for the various counties, which can be loaded into Python. When loaded into Python as an array, the number of fire pixels, and therefore fire area, can be counted in the same way as described in Chapter 4.2.1, to produce a weekly fire area count for each county.

4.2.3 NOAA data

The NOAA data are loaded into a dataframe in Python using the Pandas library. From this dataframe a new dataframe is produced, which contains the mean values for each parameter for each day, discounting any stations where that particular parameter is not recorded.

The NOAA data are also imported directly into QGIS and displayed as delimited text layers. The CSV files contain the latitude and longitude of each station at which the data were recorded, meaning that the various meteorological data can be plotted across the whole of California, displayed as either colour coded points or as discretely banded heat maps. Different layers are imported to show each of the different data sourced from NOAA. It is not necessary to clip these layers for county level analysis, as county data sets can be downloaded directly from NOAA.

4.2.4 Lightning strike data

The ASCII file contains all lightning strike data within a $700,000\text{km}^2$ area over California, for the years 2016 to 2020 inclusively. This is split into five separate Pandas dataframes for each year, and is filtered to only include cloud to ground strokes. These dataframes are then saved as CSV files and subsequently loaded into QGIS, where they can be clipped to remove all strokes that occur over areas of water in California. These clipped layers are saved again as CSV files, which can then be used for analysis in both Python and QGIS. In QGIS these layers show how the location of the lightning strokes compares to the location of the wildfires.

In Python, a simple count is performed to determine the number of strikes each day. From this data weekly lightning strike counts are produced for the weeks 27/07/20 - 04/09/20. Weekly data is also collated for the maximum, minimum, and average temperature NOAA data previously discussed. Weekly fire growth counts are also produced

from the MOD14A2 data. All of this data is put into a data frame to be used for Principal Component Analysis (PCA). The weekly increase in fire area is identified as the target variable, with the component variables set as the number of weekly ground strikes and the three temperature measures. PCA is performed on this data set to determine if there is any clustering between days that experienced large fire growth and days that did not, based on these variables.

5. Results

This chapter explains the findings of the analysis of the data collected and processed as described in Chapter 4. Chapter 5.1 presents the results of the analysis of the MODIS data. The analysis compares 2020 to the previous five years to quantify the extent to which wildfires in California in 2020 were worse than in previous years. Chapter 5.2 presents the analysis of the meteorological data from NOAA. These data, similarly to the MODIS data, are analysed over a period of six years to identify any measures that were particularly bad for 2020, and to quantify them in the context of previous years. Chapter 5.3 presents the lightning strike data analysis in much the same way as the NOAA data, comparing 2020 to previous years. All of these data are then displayed in GIS in Chapter 5.4 to produce wildfire maps for August 2020. These maps show the locations of the wildfires, as well as the NOAA and lightning strike data. Chapter 5.5 presents the results of the PCA of the effects of lightning strikes and of temperature to explain the contribution of these different factors to weekly wildfire growth.

5.1 Analysis of MODIS data

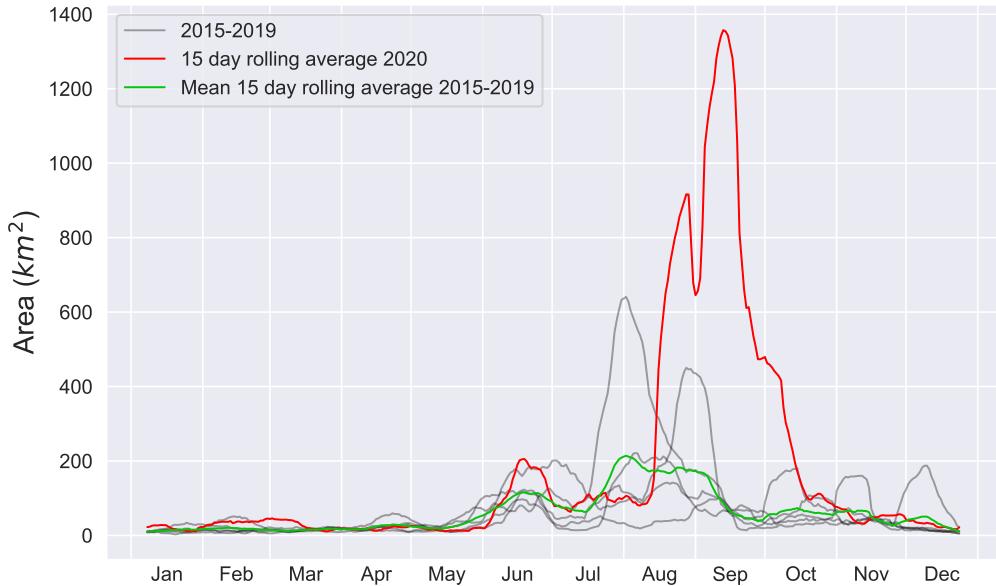


Figure 5.1: 15 day rolling average of fire area for 2020 and the mean 15 day rolling average for the years 2015-2019 inclusively, as recorded by the MODIS instrument. The red line shows the cumulative fire area for 2020, the grey lines show the other years in the range, and the green line shows an average of the grey lines — an average of the cumulative fire areas for 2015-2019 inclusively.

Figure 5.1 shows the 15 day rolling average of total fire area for 2020, and the average of the years 2015-2019 inclusively. It shows that fire area is generally low in the beginning of the year and begins to pick up around June. The five year average peaks for about six weeks centred around August and then levels off for the rest of the year. 2020 however experiences huge peaks in August and September. Figure 5.1 clearly shows that the months of August and September were dramatically worse than would be expected based on the five year average. The rolling average area of land consumed by fire on 7 September is in fact more than three times greater than the sum of the previous five years on that date. Figure 5.1 clearly shows that the wildfires in California over August and September were indeed significantly worse than anything the state had experienced in the previous five years.

It is also worth noting the effect that applying the 15 day rolling average has on the data. This rolling average is calculated for the current date, seven days previously, and seven days after. This smooths out short-lived blips in the data.

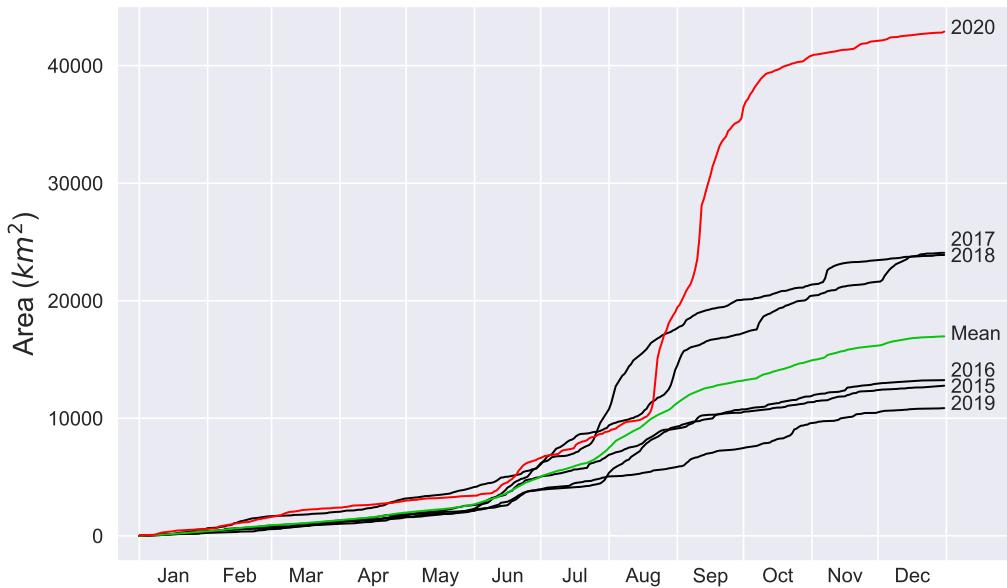


Figure 5.2: Cumulative fire area calculated for each year from 2015-2020 inclusively using data from the MODIS instrument. The line in red represents the cumulative fire area through 2020, and the black lines show the cumulative fire area for the years 2015-2019. The green line is the mean cumulative fire area of the years 2015-2019; the average of the black lines.

Figure 5.2 shows the cumulative fire area for 2015-2020 inclusively. 2020 has the largest cumulative fire area of any of the five previous years, almost twice as much as the next worst year and significantly above the mean — almost double.

The 2020 line experiences a significant spike halfway through August. This marks the steepest part of any of the lines in Figure 5.2, suggesting that August 2020 saw the fastest rate of fire growth. This is sustained more or less until the end of September, which corresponds to the peaks seen in this time period in Figure 5.1.

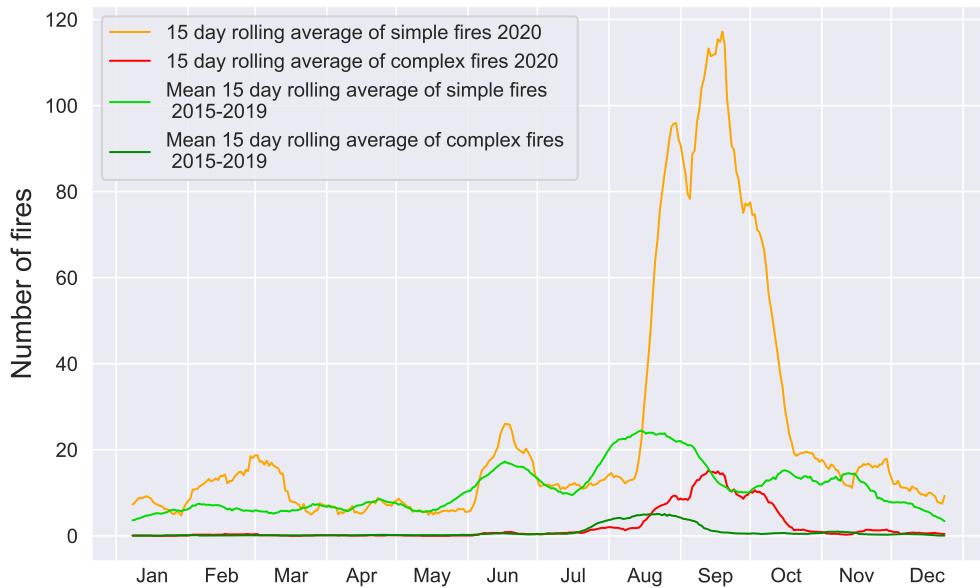


Figure 5.3: 15 day rolling average number of fires for 2020 and mean value for 2015-2019 inclusively. The rolling average number of simple fires throughout 2020 is shown in orange, with the rolling average of complex fires shown in red. The mean rolling average of simple fires for the years 2015-2019 inclusively is shown in light green, and the mean rolling average of complex fires in shown in dark green.

Figure 5.3 shows the results of the number of fires determined by the flood fill and grouping algorithms discussed in Chapter 4.2.1. The simple fire lines show the same basic trends seen in Figure 5.1, 2020 is slightly above average in February and June, and large peaks across August and September.

The complex fire lines also offer interesting insight into the events in 2020. They show that complex fires do not typically develop until the summer months. This could be for one of two reasons. Firstly, it could be because these months see a higher number of individual fires, increasing the chance that a complex fire will develop. Secondly, and more likely, it could be that complex fires only develop in these months because this is when the total fire area is high. Larger fires are potentially more likely to split or merge, and therefore are more likely to become complex fires. This could suggest that there is a certain minimum area threshold for complex fires to form. This requires an analysis of data on the average area per fire as, if this hypothesis is true, the uptick in complex fires should be accompanied with an increase in average area per fire.



Figure 5.4: 15 day rolling average of the mean fire area recorded by the MODIS instrument for 2020, shown in red. The mean value for the years 2015-2019 inclusively is shown in green.

Figure 5.4 shows the average area of each fire group per day for 2020 and the average for the five previous years. This figure shows that the average area of a single fire group in 2020 in the months of June, July, August, and September was significantly greater than the average fire for the previous five years.

These months also appear to have a periodic cycle with the mean fire area peaking for roughly two weeks at a time before returning to a lower level. This could be due to the way fire groups grow and shrink. It is possible that the peaks seen in the 2020 line in Figure 5.4 are indicative of a fire group growing larger, up to a point where it reaches its maximum size, and then decreasing. This kind of lifespan could be influenced by a fire using up all of its available resources of flammable material, or human firefighting intervention.

This figure also lends weight to the theory that complex fires only develop in the summer as this is when average fire area is greatest. Figure 5.4 shows that in 2020, and for the average of the five previous years, the largest mean area of a fire is seen in the summer months.

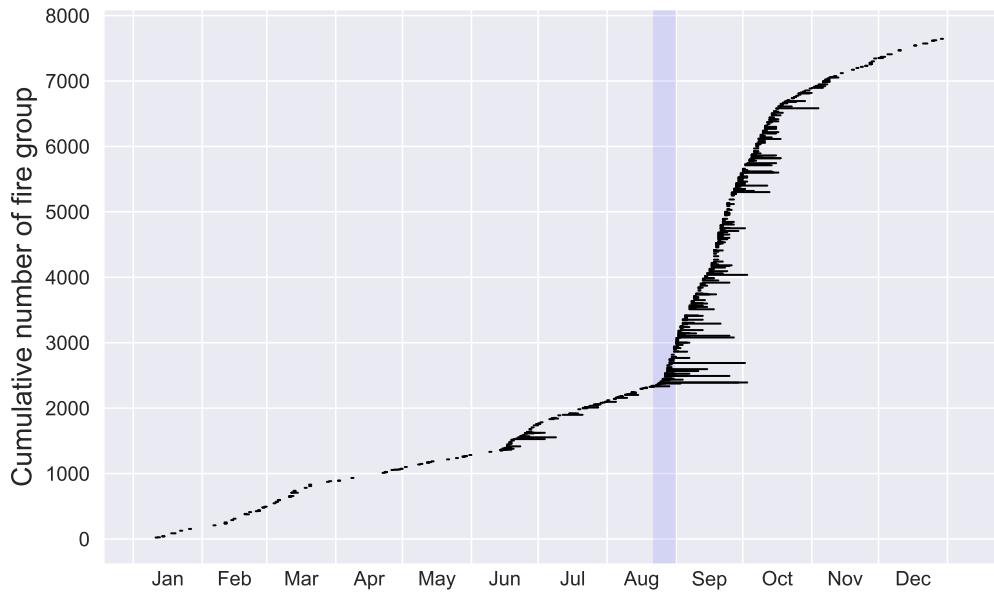


Figure 5.5: Graph of the lifetime of individual fire groups for 2020. The lines are drawn from the start date to end date for each fire group, thus showing all the days on which the fire group was burning. The area highlighted is the last ten days of August, which marks the period of time in which the longest lasting fires were initiated.

Figure 5.5 shows the lifetimes of each fire group in 2020. Each line represents a distinct fire group and is drawn from the first date that the group is detected by the MODIS instrument, to the last. It shows that up until June there were few fires burning each day, if any, and that these were all short lived. Mid-June experiences more frequent fires which tended to burn for longer. July and the beginning of August are relatively quiet compared to the rest of August and the first half of October. This period is characterised by a sharp increase in the number of fires, with many of the fires burning for several weeks. By early November the number of fires has reduced with fires only burning for a few days.

What is noteworthy from this graph is that the longest lasting fires of the year were started in the last third of August, highlighted in blue. This suggests that the peak in fire area seen in September may in fact be due to fires that were started in August. This period also marks a significant increase in the daily number of fires, also shown in Figure 5.2.

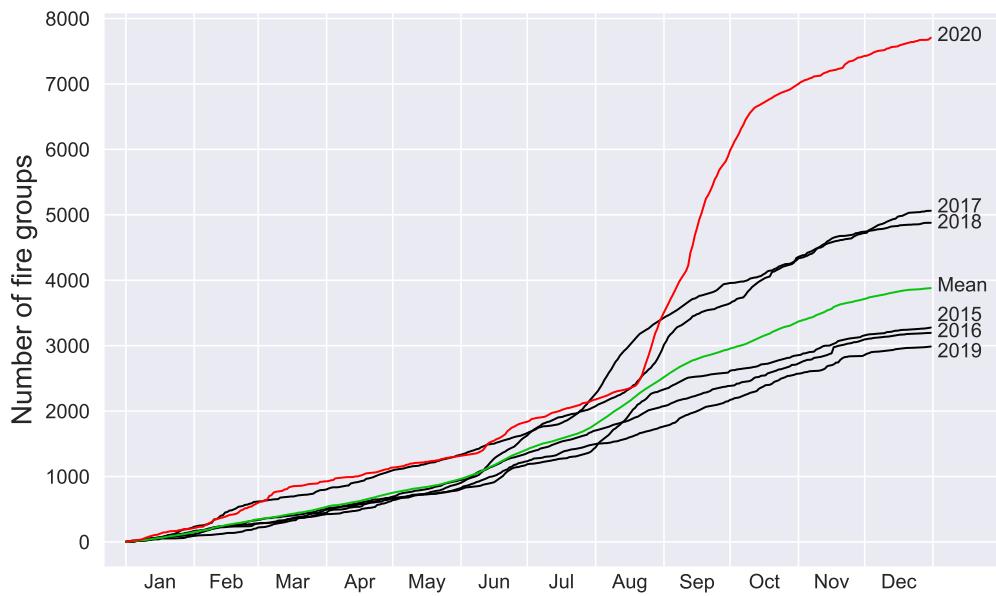


Figure 5.6: Cumulative number of distinct fire groups for the years 2015-2020 inclusively. 2020 is shown in red, the years 2015-2019 are shown in black, and the mean of the years 2015-2019 is shown in green. All lines are labelled accordingly.

Figure 5.6 shows the cumulative fire group count for 2020 in red, as well as the five previous years in black. Also shown, in green, is the average of the five previous years. As to be expected, this graph looks similar to Figure 5.2 showing the cumulative area of fire, with the last five years ranking in the same order as for total fire area. The line for 2020 appears to show a steady increase in the cumulative number of distinct fire groups up until the beginning of June. Within the first week of June there is an uptick in the rate of growth of the number of fires, but this is dwarfed by the rate seen from mid-August through to early October. For October through to the end of the year the rate of growth returns to a rate similar to the beginning of the year. Once again, 2020 is significantly above the mean — almost twice as many fire groups.

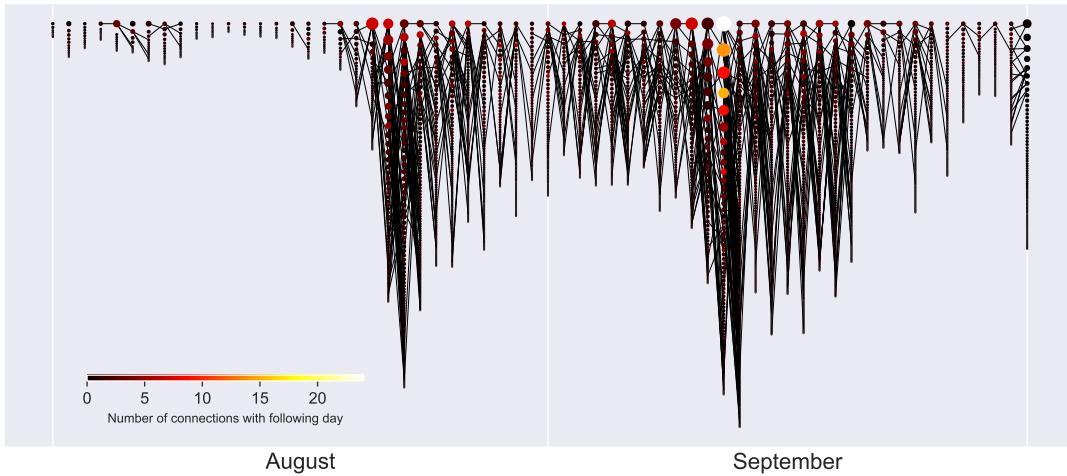


Figure 5.7: A network map of the interactions of the distinct fire groups present on each day from 01/08/20 - 01/10/20. Each node represents a distinct fire group. The size of each node is proportional to the size of the fire group it represents. The colour of each node represents the number of connections it has with fire groups in the next day according to the colour bar shown. Only connections between the largest fires are plotted to aid in visualisation.

Figure 5.7 shows a network map for the fire groups across August and September. Each individual fire group is plotted as a node on this map. The connections between fires — fire that overlap one day to the next — are also plotted. The colour of the node corresponds to the number of connections it has with fires in the following day. Each pair of fires can only connect to each other once, regardless of how many pixels they overlap by. The size of each node corresponds to the area of the fire group it represents. The fire groups have been sorted to order the node from largest to smallest area. It shows that the beginning of August was relatively quiet, with few, short-lived fires. The fires start to pick up around the middle of August. These days appear to show the fastest rate of increase in number of fire groups per day. The number of fire groups and their average size then decreases towards the beginning of September. These measures peak again around mid-September, then steadily decrease towards the end of the month. At the very end of the month there is the beginning of a possible new peak.

There are a few features to highlight from this map. After the first peak in August, the total area and number of fire groups remains elevated at least until the beginning of October, identifying the days around 20 August as a watershed period. There also appears to be a recurring trend regarding single-day peaks in number of fire groups. The single-day highs in August and September are both preceded by a day with large fire groups which split into many smaller ones.

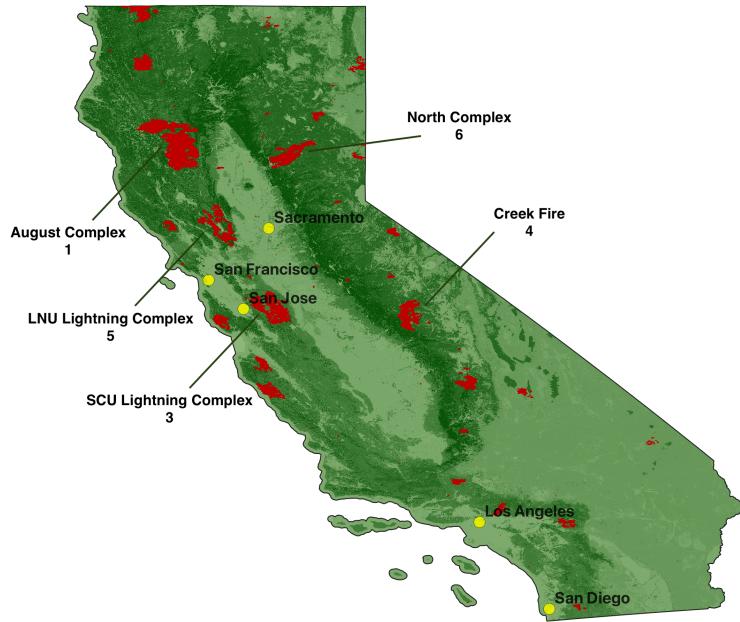


Figure 5.8: A GIS map of California showing wildfires over August and September. Land cover data for the state of California is shown in green on this map, with areas of darker green representing areas of land more densely covered with vegetation. Wildfires are shown in red, with large fires labelled according to the names given them by CalFire, and ranked by area of all recorded wildfires, as seen in Table 2.1.

Figure 5.8 shows the MOD14A2 data overlaid on a map of California, with all large fires annotated. The number by each annotation is the ranking of the fire in terms of area on record. As stated in Chapter 1, five of the six largest fires ever recorded were burning in August or September 2020. This figure shows the single worst week ever in terms of total fire area. The fires appear to be mostly along the western edge of the Central Valley, as well as to the north. This type of data will be used in conjunction with other geospatial data to determine whether there is any correlation between the meteorological and wildfire data sets.

Figures 5.1 through 5.8 show that in 2020 wildfires in California were larger individually, were more numerous, and covered a larger total area than any year since 2015. They support the literature discussed in Chapter 2 which claimed that 2020 was the worst year for wildfires in California in recent history.

5.2 Analysis of NOAA data

The data from NOAA is analysed in much the same way that the MODIS data was. Data for 2020 is compared to data for 2015-2019 inclusively to determine how 2020 compares to previous years.

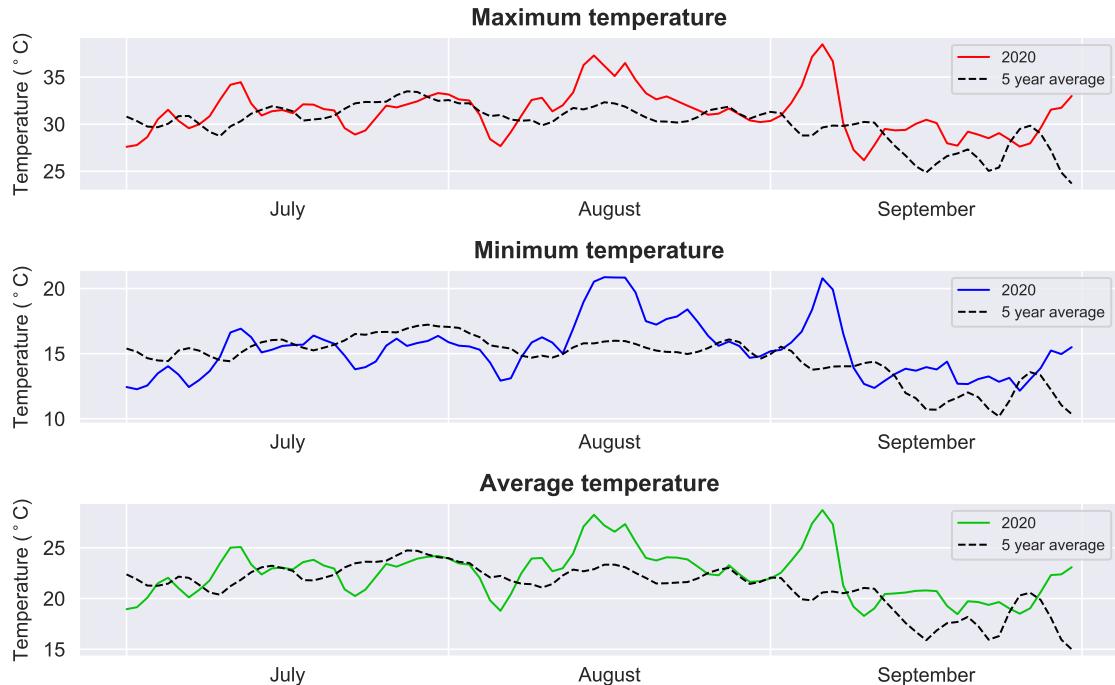


Figure 5.9: Daily NOAA maximum, minimum, and average temperature data for the state of California between 01/07/20 - 01/10/20. Also shown are the five year averages for the years 2015-2019 across the same months; July, August, and September. The values are calculated by taking the mean value of all recorded values at the 1203 statewide recording stations.

Figure 5.9 shows temperature data for 2020 and an average of the years 2015-2019 inclusively. This data is collated from measurements taken at 1203 stations across California. The three different temperature measures, namely daily maximum temperature, daily minimum temperature, and daily average temperature all follow the same trend. The month of July 2020 is relatively close to the five year average for July, with a small peak around 10 July, and a small dip around 20 July. August however sees a sizeable peak around the middle of the month which is sustained for about two weeks. What is noteworthy about this period is that the greatest difference from the five year average is seen in the minimum temperature data. This suggests that the area was subjected to temperatures which remained above average for a period of 1-2 weeks with no respite. A comparatively normal end of August is met with a large peak in the first week of September. The rest of September

remains slightly higher than average for the most part, and picks up again towards October.

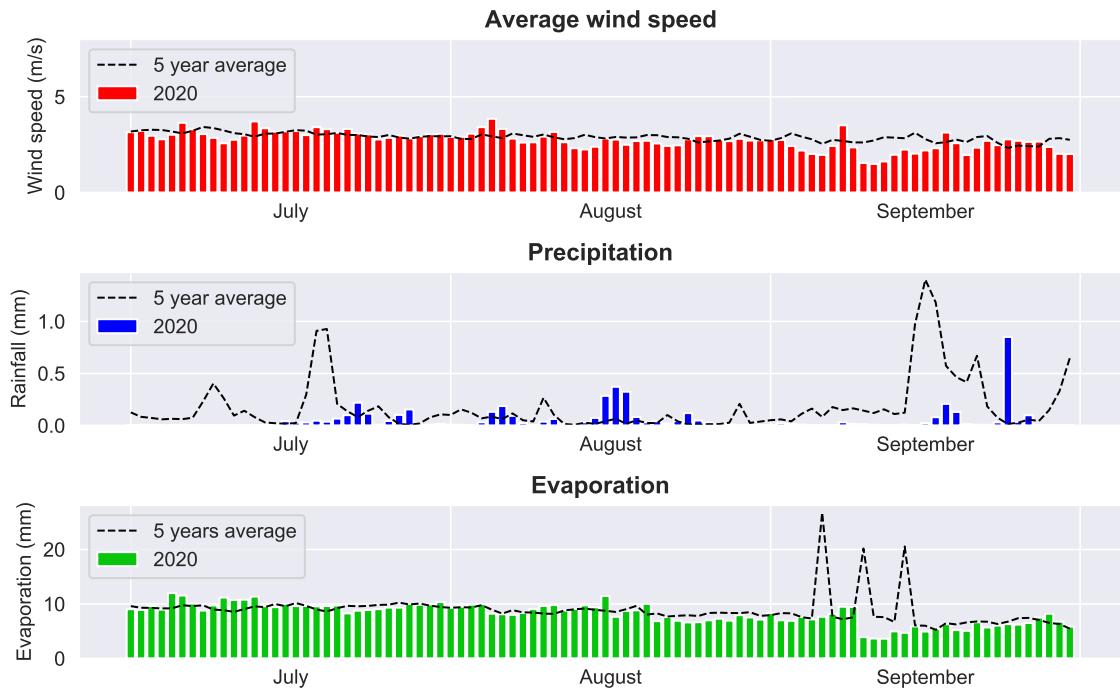


Figure 5.10: Daily NOAA average wind speed, precipitation, and evaporation data for 01/07/20 - 01/10/20. Also shown are the five year averages for the years 2015-2019 across the same months; July, August, and September. The values shown are the average of the data across the 1203 statewide recording stations.

Figure 5.10 shows data for average wind speed, average precipitation, and average evaporation, all averaged across the whole of California. The trends in these data are much less pronounced than the previous data set. The average wind speeds across California in 2020 seem to be largely in line with the average for the previous five years. The only point of note is the second week of September which shows lower than average wind speeds.

The average precipitation across California for 2020 is also largely in line with the five year average. This plot shows some of the shortcomings of this kind of analysis. As these are the summer months precipitation is usually very low. This means that one storm potentially has the ability to significantly affect the value of the mean precipitation. It appears that this might be the case for the two peaks in the five year average seen in the middle of July and in the middle of September. These are likely significantly influenced by one event in one year, rather than a regularly occurring peak in precipitation in these specific weeks every year.

This effect is also seen in the average evaporation plot. It is worth noting that of the 1203 stations from which this data is drawn, not all stations collect data for all the parameters

discussed in this report. It could be that some locations experience extremes of weather that are not represented here as there is simply not a station in close proximity to collect the appropriate data.

The precipitation data set is an average of all the available readings across the state of California. It could be the case that performing an amalgamation of the results in this manner is not appropriate over such a large area, where latitude, climate, altitude, and other influencing factors vary so substantially. To address this, further analysis is performed for the same measurements relating to just one single county in California — Sonoma County. Sonoma was chosen because it experienced wildfires during August 2020.

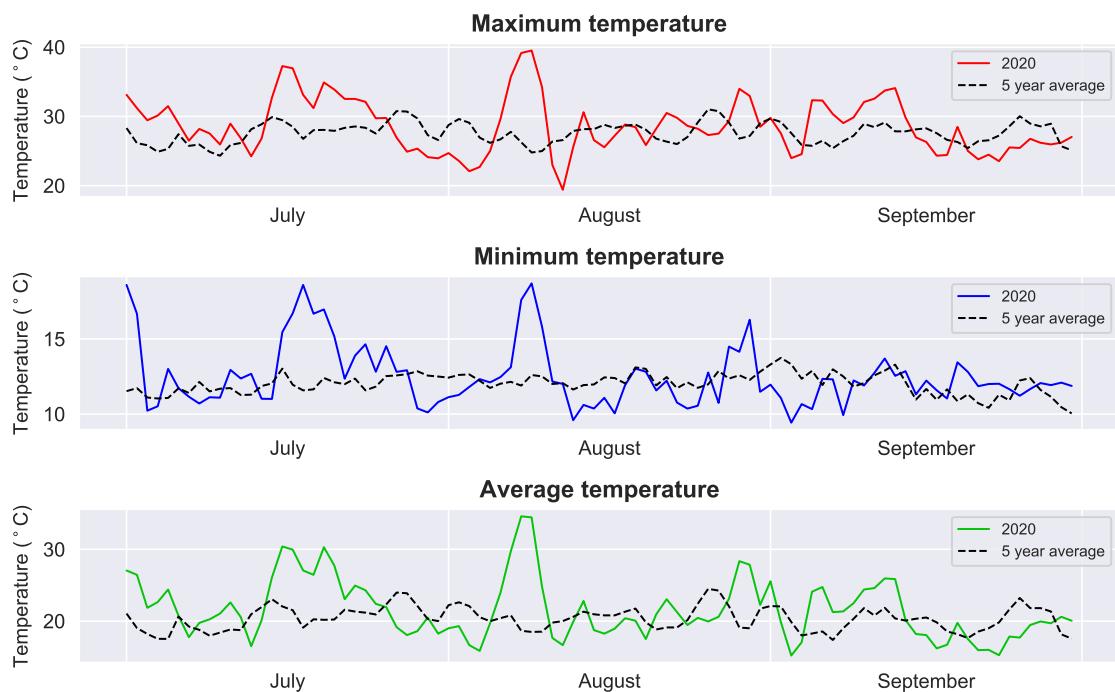


Figure 5.11: Daily NOAA maximum, minimum, and average temperature data for Sonoma county in the state of California between 01/07/20 - 01/10/20. Also shown are the five year averages for the years 2015-2019 across the same months; July, August, and September. The values are calculated by taking the mean value of all recorded values at the 1203 statewide recording stations.

The temperature plots shown in Figure 5.11 show roughly similar patterns to those in Figure 5.9. Towards the middle of each month in 2020 there is a spike which is above average for the same time period in the previous five years. The largest spike, which is in August, actually occurs slightly earlier than the August peak for California as a whole. It is worth drawing attention to the amount of variation there is in the measurements for average daily temperature. At the beginning of August the average daily temperature

changes from around 16°C to around 35°C in the space of just a few days. This could be an accurate account of average daily temperature in the region, but this large variation does raise questions of how realistic this data is and whether there are any other phenomena influencing these values.

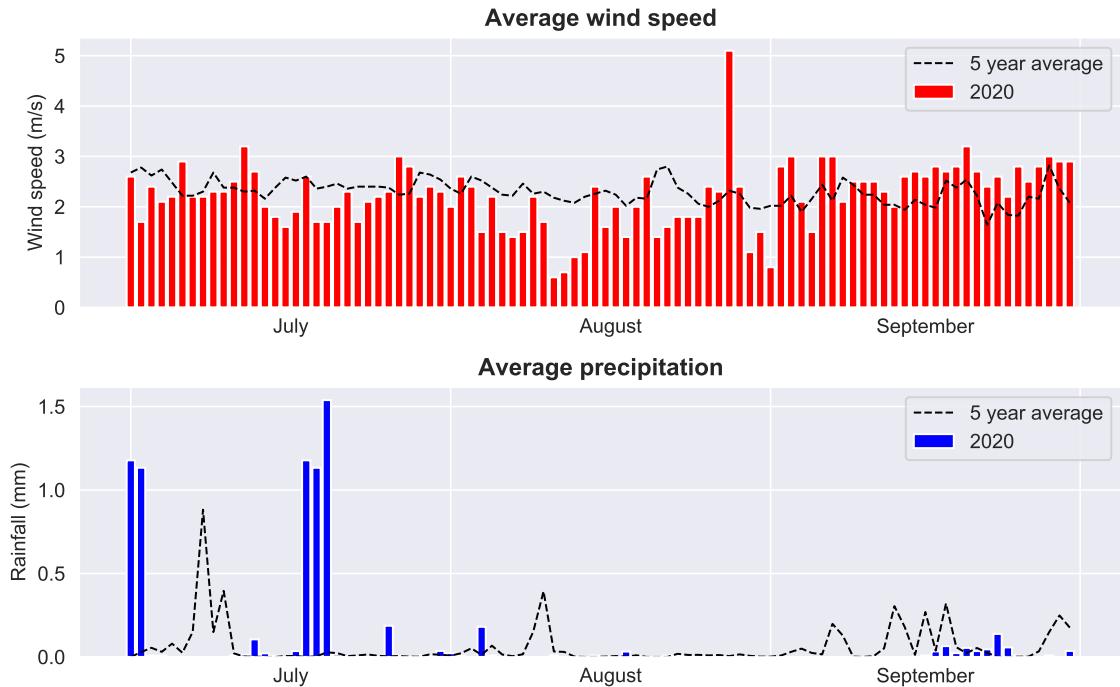


Figure 5.12: Daily NOAA average wind speed, and precipitation data for 01/07/20 - 01/10/20. Also shown are the five year averages for the years 2015-2019 across the same months; July, August, and September. The values shown are the average of the data across the 1203 statewide recording stations.

Figure 5.12 shows average wind speed and average precipitation for Sonoma County (evaporation data is not recorded in Sonoma County). It appears from the average wind speed plot that these months in 2020 were fairly typical, with perhaps August seeing slightly less strong winds than previous years. The precipitation graph also does not provide great insight because, as mentioned before, there are few data points available for these months.

5.3 Analysis of lightning strike data

The lightning data, unlike all other data analysed in this report, is collected for the years 2016-2020, as this was provided by Vaisala. The majority of lightning strikes recorded are cloud-to-cloud strikes. These are not relevant to this investigation into wildfires, as only cloud-to-ground strikes can ignite a fire. In the following analysis the term lightning strike

refers only to ground strikes. The sample area for the lighting data excludes the south-east corner of California, around the Mojave desert. Much of the land in this region has little vegetation and is not subjected to many wildfires.

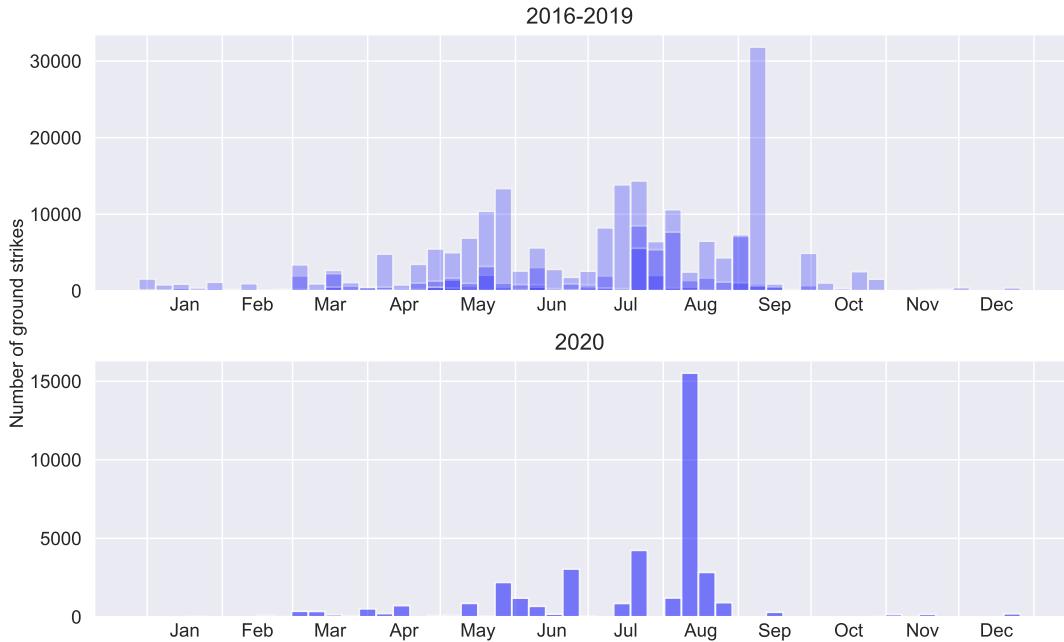


Figure 5.13: Bar chart plots of the lightning strike data, specifically cloud-to-ground lightning strikes. The data has been plotted weekly for ease of interpretation. The chart for 2016-2019 has the weekly total number of lightning strikes for each year overlaid onto each other. The 2020 graph plots the weekly number of lightning strikes in 2020. Note the different scales on the y-axis.

Figure 5.13 shows bar charts of the lightning strike data from Vaisala. The lightning strike data confirms that August did indeed experience a surge in lightning strikes, marking the worst days of the whole year by a large margin. 2020 does not however particularly stand out against the other years. It does mark the second highest weekly peak, but it was exceeded by a week in mid-September 2017. In fact, of the years sampled, 2020 marks the lowest total volume of lightning strikes. Whilst the wildfire area for 2017 is the next highest after 2020 between 2015 and 2020, it is still significantly lower. This suggests that lightning strikes alone do not fully explain the wildfire trends seen in California.

5.4 Correlations

The MODIS data are considered in the context of the NOAA and lightning strike data. This discussion aims to highlight correlations in the data, not to identify the causal

relationship between variables. For example, wildfires and air temperature often experience peaks at the same time. It could be the case that higher than average air temperatures provided the conditions needed for large wildfires to form. It could also be the case that a large area of burning could cause a temperature rise.

Other trends in the data become more apparent when geographical locations are combined with temporal location. The MOD14A2 data and the MCD64A1 data can be loaded in GIS and combined with the other data sets to determine if there are any correlations.

In the following GIS maps, the base layers consist of wildfires plotted on a map of California. The MOD14A2 and MCD64A1 data are combined in the red regions to show the full extent of the fires. The base map also includes a land cover layer, which is useful in explaining the boundaries of some of the wildfires. All data shown are from the month of August 2020.

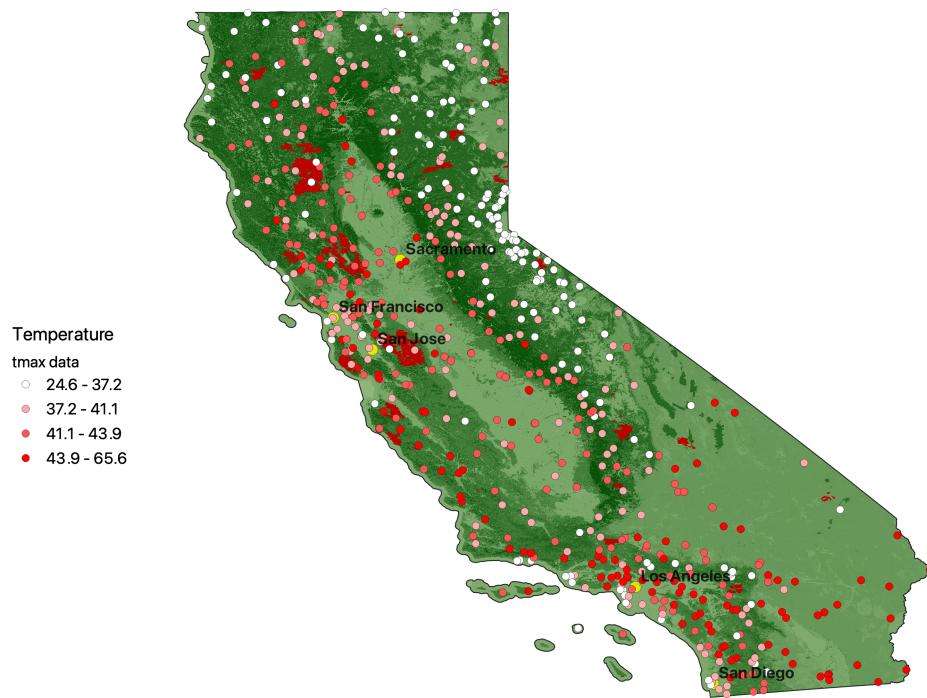


Figure 5.14: A GIS map of maximum temperature distribution across the state of California. USGS land cover data is shown in green, with areas of darker green representing areas of land more densely covered with vegetation. Areas of fire in the month of August 2020 are shown in dark red. The coloured points on the map are placed at the location of the recording station, and their colour corresponds to the maximum recorded air temperature at that station in the month of August 2020. The legend in the figure shows the colour banding used for the points.

Figure 5.14 shows maximum recorded air temperature in August at each station across California. The points represent the location of each station in the NOAA database, and the values of the data shown in the legend are in Celsius. It appears that the west side of the Central Valley may have experienced hotter temperatures than the east side. As stated earlier, the causal relationship between air temperature and wildfires cannot be determined from these data alone, but it is worth mentioning nonetheless. The gradations in the colour bar used to populate this map have been chosen to highlight the differences in maximum temperature between the east and west sides of the valley. While the west did experience higher maximum temperatures, they were only slightly higher than the east, typically by a few Celsius.

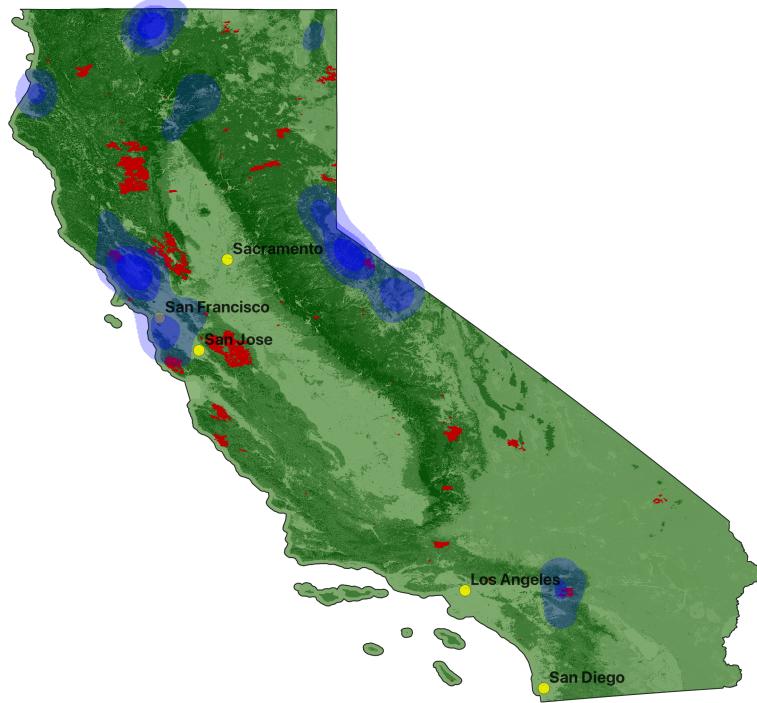


Figure 5.15: A GIS map of the state of California showing a precipitation heatmap for August 2020. The base layers are the same as those in Figure 5.14. The areas in blue represent areas which saw precipitation during the month of August 2020. The darker blue areas experienced a greater amount of precipitation than the lighter areas.

Figure 5.15 shows a heatmap of precipitation across California in August 2020, shown in blue, with areas of darker blue indicating areas which saw a greater total amount of precipitation. The precipitation clusters around the coast by San Francisco and also the mountains in the east. While this map shows the precipitation covered a relatively large area, it should be recalled from Figure 5.10 that the total volume of precipitation is still extremely low.

There is little useful information about why wildfires formed where they did to extract from this map. There are a few small fires that overlap with this precipitation. It could be that the precipitation was so low that it did not influence the growth of the fires significantly, or that this precipitation brought with it lightning strikes which initiated these fires. This data serves mainly to exclude areas of land where a wildfire is unlikely to start, as the area is damp.

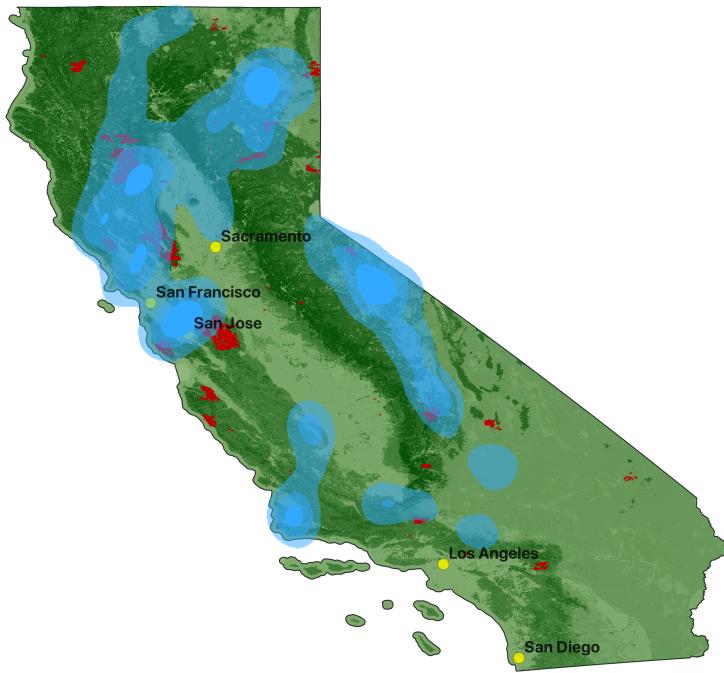


Figure 5.16: A GIS map showing lightning strikes across the state of California in August 2020. The base layers of the map are the same as in Figure 5.14. The areas of light blue represent areas which experienced lightning strikes in this time period, with darker areas receiving more lightning strikes than lighter areas.

Figure 5.16 shows a heatmap of the locations of ground lightning strikes in August. The lightning strikes appear to be concentrated around the northern parts of the state, avoiding the Central Valley. The strikes on the west of the valley seem to overlap with some of the largest wildfires in the state. The other concentration of strikes, to the east of the valley, occurs over what is the Sierra Nevada mountain range. This area of land is at a much higher elevation and has a lot less tree cover, potentially explaining why so few wildfires burned there.

As was explained by the CalFire report in Chapter 2, these lightning strikes all occurred within a few days of each other. The total number of lightning strikes shown on this map constitute nearly 50% of the yearly average number of lightning strikes for the state of California.

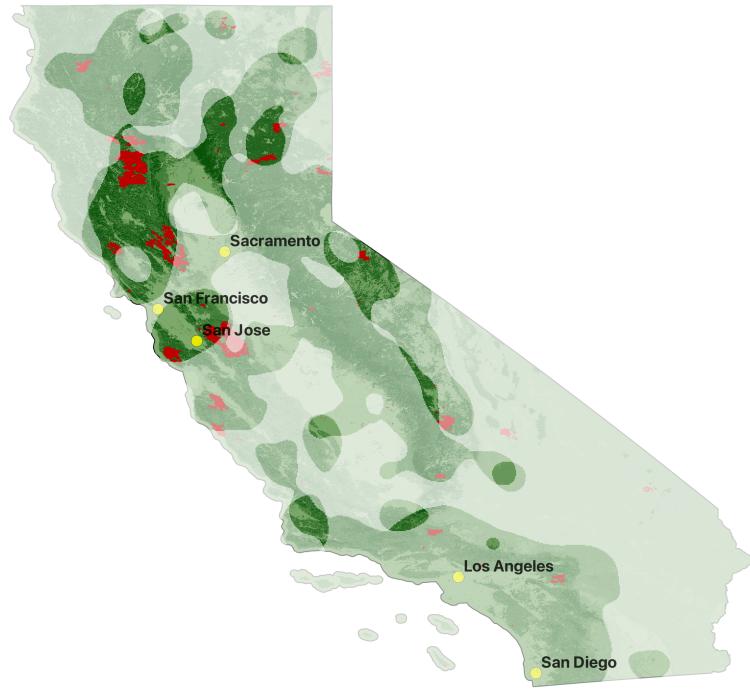


Figure 5.17: A GIS map showing a combination of the variables displayed in Figures 5.14 - 5.16. Areas greyed out correspond to areas with a low maximum air temperature, high precipitation, and few lightning strikes. This leaves the areas in colour as areas that were hot, dry, and saw many lightning strikes in the month of August 2020. The base layers of this map are the same as those in Figure 5.14.

Figure 5.17 shows a combination of all the variables discussed in Figures 5.14 - 5.16. Areas are greyed out successively if they have either a low maximum temperature, high precipitation, or few to no lightning strikes. The areas left in colour therefore are hot, dry areas that experienced a high number of lightning strikes — prime wildfire conditions.

A trend does appear to emerge from this data. The area to the north west of the Central Valley appears to fit all these criteria, and is indeed the location which sees the largest fires. Other smaller fires to the east of the valley also often lie in the hot spots created by this analysis. This suggests that there may be a link between area of overlap of all of these factors and an increased probability of wildfires. As mentioned previously this analysis cannot determine the causal relationship between these factors, just that they do indeed appear to correlate.

5.5 PCA of wildfire growth and meteorological factors

The data discussed in Chapter 5.4 can be analysed further to quantify the correlation between them. The data are broken down and analysed on a weekly basis for three different counties in California; Marin, Napa, and Sonoma. The data are gathered for the three counties over the five weeks that span August, giving a set of 15 individual data points. The data considered in this analysis are maximum, minimum, and average temperatures, and lightning strikes. As mentioned previously, other meteorological measures are not consistently available across counties and are therefore not considered.

For the PCA the data are formulated such that each data point has a target variable associated with it, in this case, a simple marker which indicates whether there have been new fires in each county that week or not. The component variables are the temperature and lightning strike data. PCA is carried out on this data set to determine whether there is any clustering between weeks that have fire growth and weeks that do not.

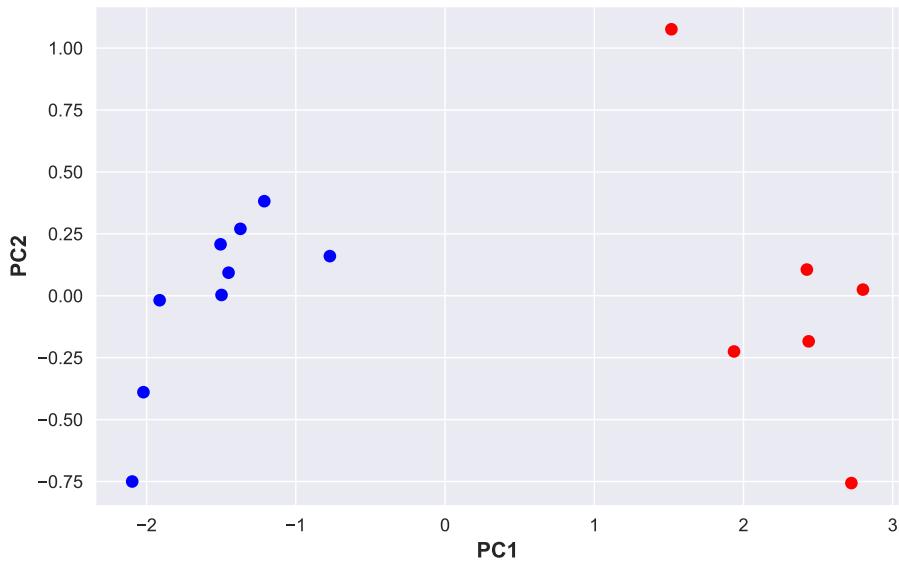


Figure 5.18: Plot of the data as transformed by the PCA as described in Chapter 4. Weeks in which a county experienced wildfire growth are shown as points in red, and weeks where a county did not experience wildfire growth are shown as points in blue. The plot shows clear grouping around the PC1 axis.

Component vector	Lightning strikes	Average temperature	Maximum temperature	Minimum temperature	Total component variance
PC1	0.48376609	0.51343503	0.49951064	0.50283592	93.0%
PC2	-0.78051897	0.31456039	0.53125752	-0.09801705	4.72%

Table 5.1: Composition of first two principal component vectors from average air temperature, maximum air temperature, minimum air temperature, and number of lightning strikes. Also shown are the contributions of each principal component to the total variance of the data.

Figure 5.18 and Table 5.1 show the results of the PCA described above. There does indeed appear to be clustering along PC1 of points with no fire growth (blue) and points with fire growth (red). The contributions of each variable to the PC1 vector are roughly equal, meaning they all exert similar influence on whether or not a fire starts in a certain week. This analysis shows that if a county experiences a week of high temperatures and lots of lightning strikes then it is likely to also experience wildfire growth. This classification could be used to identify future times where a county is at risk of wildfire growth.

6. Conclusions and future works

This chapter summarises the results presented in Chapter 5 and gives the main conclusions of these results.

The main aims of this report are to identify quantitatively how extreme 2020 was for wildfires in California, and to identify possible causes for those wildfires. This report shows that 2020 was indeed an extraordinary year for California. The wildfires in 2020 were greater in number, larger in area, and burned longer than wildfires in the previous five years.

The data sourced for this report, and the different analytical techniques employed proved effective tools in achieving the aims and objectives set out in Chapter 3. They provide relevant insight into the events in California in 2020, and offer a background of information with which to compare it to.

August 2020 is identified as a critical month; the beginning of a sharp increase in wildfire metrics across the board. The total area of wildfires increases in late August and peaks in mid-September. It has been shown that some of the largest fires burning in September were in fact started in August. A link has been highlighted between environmental factors in August and the rise in the number of wildfires. 2020 had the hottest August in over 125 years, and much of the state was experiencing various levels of drought. The state also experienced a large number of lightning strikes in the space of just a few days.

The wildfires visualised in GIS show just how much of the land area of California they covered. Geospatial data concerning temperature, precipitation, and lightning strikes, as well as land cover overlaid onto these images show how these different parameters overlap to create areas of land particularly vulnerable to wildfire ignition.

Principal Component Analysis conducted with county data including wildfire growth and meteorological data shows clear clustering based on whether or not a county experienced wildfire growth in a week. This analysis shows that lightning strikes and air temperature are both good indicators of how vulnerable an area is to wildfires.

2020 marks a year of extremes in California. Whilst it has not been discussed in detail in this report, it is hard to miss how as time goes on more and more extremes in weather seem to become more and more common. This report sheds light on the situation in California, and should act as a reminder that climate change poses an imminent threat to all life on this planet.

This report considers various data sources for its analysis. A number of different analytical techniques are employed to answer the questions posed in Chapter 1. Any future work on this subject would fall into one of two categories. Either more data would be gathered

and analysed in new ways to gain further insight into the questions raised in this report, or entirely new questions spawned by the work in this report could be pursued.

An obvious place to start for future work on wildfires in California would be to consider more data. This could take the form of a similar methodology as this report, but with data for many more years in the past, to better identify long term trends in wildfires in California. It could also be an improvement in the PCA performed in this report. This could be achieved by simply increasing the number of data points. The process could be repeated for all counties in California, and for every week for the last five years. This would give an improved result as a larger data set would help to identify all factors affecting the growth of wildfires. The MODIS instrument produces at least two other products that could be useful in any further analysis.

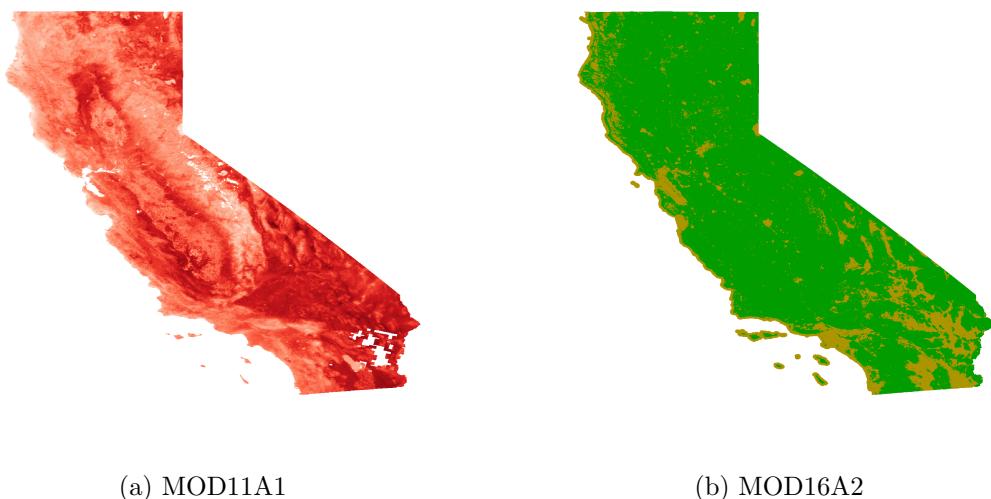


Figure 6.1: Example GIS maps of possible additional MODIS data to be used in future work.

Figure 6.1 shows samples of the MODIS Land Surface Temperature and Emissivity (MOD11) and Evapotranspiration (MOD16) data products. Both these data sources could be used to supplement the NOAA data used in this report. The drawback of the NOAA data is that it is only available at the discrete locations of the stations, and is also dependent on whether or not that station happens to have the necessary equipment to take measurements of the desired parameters. These data sets however offer information in 1km^2 grid squares, just as with the MOD14 data.

The ideal methodology would be similar to that used in this report, but with a much higher temporal and spatial resolution. Instead of collecting the various data on a county by county basis, it could be collected for each 1km^2 grid square. It could also be collected daily, further increasing the number of data points available.

This report does however also raise the possibility of new questions and new research

opportunities. The lightning strikes discussed in this report constitute proximate causes of the worst wildfire season in California's history. It would be interesting to investigate further the ultimate causes of those wildfires. An investigation could be undertaken to try to understand why 2020 in particular was so bad for lightning strikes, particularly dry ones. As touched upon in Chapter 2, the cause of the lightning strike surge in August was due to tropical storm Fausto bringing dry lightning — lightning strikes without precipitation — to the region. Further research could be carried out to determine how the frequency and intensity of tropical storms over California has changed in recent years. Similarly, it would be interesting to obtain data about how often these storm bring dry lightning strikes, as this is obviously a prime wildfire ignition source.

Bibliography

- [1] CalFire. (November 3, 2020). “Top 20 largest california wildfires,” [Online]. Available: https://www.fire.ca.gov/media/4jandlh/top20_acres.pdf (visited on November 26, 2020).
- [2] J. E. for Agence France-Presse — Getty Images. (December 3, 2019). “The creek fire in madera county, calif.,” [Online]. Available: <https://www.nytimes.com/2020/09/09/climate/nyt-climate-newsletter-california-wildfires.html> (visited on November 26, 2020).
- [3] WHO. (April 6, 2021). “Wildfires,” [Online]. Available: https://www.who.int/health-topics/wildfires#tab=tab_1 (visited on April 6, 2021).
- [4] National Interagency Fire Center. (December 21, 2020). “National large incident year-to-date report,” [Online]. Available: <https://gacc.nifc.gov/sacc/predictive/intelligence/NationalLargeIncidentYTDReport.pdf> (visited on March 24, 2021).
- [5] D. L. for ABC News. (October 10, 2020). “Damage from california’s wildfires estimated at \$10 billion, experts say,” [Online]. Available: <https://abc7news.com/california-wildfires-cost-of-cal-fire-stanford-wildfire-research/6897462/> (visited on March 24, 2021).
- [6] CalFire. (2020). “2020 fire siege,” [Online]. Available: <https://www.fire.ca.gov/media/hsviuuv3/cal-fire-2020-fire-siege.pdf> (visited on March 24, 2021).
- [7] A. C. Mulkern. (August 24, 2020). “Fast-moving california wildfires boosted by climate change,” [Online]. Available: <https://www.scientificamerican.com/article/fast-moving-california-wildfires-boosted-by-climate-change/> (visited on March 24, 2021).
- [8] The Guardian. (October 6, 2020). “California wildfires spawn first ‘gigafire’ in modern history,” [Online]. Available: <https://www.theguardian.com/us-news/2020/oct/06/california-wildfires-gigafire-first> (visited on November 26, 2020).
- [9] Z. Guzman. (Jul. 16, 2015). “The california drought is even worse than you think,” [Online]. Available: <https://www.cnbc.com/2015/07/16/the-california-drought-is-even-worse-than-you-think.html> (visited on April 27, 2021).

- [10] K. M. F. M. M. N. V. G. IMAGES. (August 16, 2020). “Lightning fills the sky above the bay bridge as dawn breaks in san francisco, calif., on sunday, aug.16, 2020.” [Online]. Available: <https://www.forbes.com/sites/rachelsandler/2020/08/19/california-sees-10849-lightning-strikes-in-72-hours-as-wildfires-rage/?sh=57d84911c860> (visited on April 27, 2021).
- [11] The LA Times. (August 16, 2020). “Moisture from tropical storm fausto fuels northern california storms,” [Online]. Available: <https://www.latimes.com/california/story/2020-08-16/moisture-from-tropical-storm-fausto-fuels-northern-california-storms> (visited on November 26, 2020).
- [12] NASA. (August 17, 2020). “NASA finds short-lived fausto faded fast,” [Online]. Available: <https://blogs.nasa.gov/hurricanes/2020/08/17/fausto-eastern-pacific-ocean/> (visited on March 27, 2021).
- [13] NOAA. (Sep. 4, 2020). “National temperature and precipitation maps,” [Online]. Available: <https://www.ncdc.noaa.gov/temp-and-precip/us-maps/> (visited on March 27, 2021).
- [14] United States Drought Monitor. (August 20, 2020). “Map archive,” [Online]. Available: <https://droughtmonitor.unl.edu/Maps/MapArchive.aspx> (visited on March 27, 2021).
- [15] NASA. (November 26, 2020). “Spacecraft icons,” [Online]. Available: <https://science.nasa.gov/toolkits/spacecraft-icons> (visited on November 26, 2020).
- [16] NASA. (November 26, 2020). “Terra: the EOS Flagship,” [Online]. Available: <https://terra.nasa.gov/> (visited on November 26, 2020).
- [17] ESA. (March 2, 2020). “Polar and sun-synchronous orbit,” [Online]. Available: https://www.esa.int/ESA_Multimedia/Images/2020/03/Polar_and_Sun-synchronous_orbit (visited on March 25, 2021).
- [18] NASA. (March 25, 2021). “MODIS Components,” [Online]. Available: <https://modis.gsfc.nasa.gov/about/components.php> (visited on March 25, 2021).
- [19] NASA. (March 25, 2021). “MODIS,” [Online]. Available: <https://modis.gsfc.nasa.gov> (visited on March 25, 2021).
- [20] NASA. (February 22, 2021). “FIRMS FAQ,” [Online]. Available: <https://earthdata.nasa.gov/faq/firms-faq> (visited on March 30, 2021).
- [21] C. O. Justice, L. Giglio, D. Roy, L. Boschetti, I. Csiszar, D. Davies, S. Korontzi, W. Schroeder, K. O’Neal, and J. Morisette. New York, NY: Springer New York, 2011, pp. 661–679, ISBN: 978-1-4419-6749-7. DOI: 10.1007/978-1-4419-6749-7_29. [Online]. Available: https://doi.org/10.1007/978-1-4419-6749-7_29.