



MONASH
BUSINESS
SCHOOL

Using Remote Sensing Data to Understand Fire Ignition During the 2019-2020 Australia Bushfire Season

Weihaio Li

EBS Honours student

Research Plan

for Honours, supervised by Prof. Di Cook

Department of
Econometrics &
Business Statistics

20 May 2020

1 Introduction

In Australia, bushfires have historically caused massive loss of property and life. Since the 1967 Hobart bushfires, the insurance claims for building losses was greater than \$A10 million (McAneney, Chen & Pitman, 2009). The Australian bushfire season in 2019-2020, compared with other major bushfires in history, had a more devastating impact on the environment and properties. According to the Parliament of Australia report (2020) in this past bushfire season, 3094 houses have been destroyed and over 17M hectares of land burned. These two figures are the highest in history. Fortunately, only 33 people including firefighters died compared to 173 in the Black Saturday fires and 47 in Ash Wednesday fires (Parliament of Australia, 2020). Understanding the cause of the 2019-2020 Australian bushfires is a crucial step to provide information for future policy and fire risk management.

The research is also motivated by the lack of knowledge about the cause of bushfire in Australia. Beale and Jones (2009) state in their research that the cause is only known for 58.9% of fires. Historically, about 50% are due to deliberate and suspicious ignitions, 35% by accidents and 6% natural, such as lightning. Providing a probable cause for bushfires may help investigators in practice.

2 Research aims and questions

This research will explore the potential methods of fire ignition during 2019-2020 Australia bushfires season based on the new availability of satellite data, and provide a model to predict the fire risk for locations around the country. Hotspots data from the JAXA's Himawari-8 satellite, will be combined with weather data from the Australian Bureau of Meteorology, and other data on roads, recreation sites, vegetation, to study locations of fire ignition and predict bushfire risk. Satellite data provides the opportunity to more objectively study ignition locations, particularly in less accessible regions.

The overall research question is whether we can provide a probabilistic estimate of aim of the cause of fire ignition for the 2019-2020 bushfires. Several smaller research questions that will help towards the overall aim are:

1. Using satellite hotspots data can we detect ignition time and location?
2. Can data from other sources including vegetation, weather, proximity to road and recreation site help to inform ignition type?
3. How do the characteristics of 2019-2020 bushfires compare to historical bushfires?

4. Can we make a useful model for the fire risk across Australia? What predictors including fire indexes, proximity to road and recreation site, weather and vegetation are useful for modelling fire risk?

3 Review of literature

Existing research on bushfire modelling can be divided into two main categories: simulation and analytical modelling.

In simulation modelling, Keane et al. (2004) develop landscape fire succession models (LFSMs) to simulate fire behaviour spatially, including fire ignition and fire spread, and accounting for fire and vegetation dynamics. Bradstock et al. (2012) used a model called FIRESCAPE which involved simulating fire behaviours with fuel and weather conditions. Simulation methods are cost-effective and time-effective ways to model bushfires, but ignition cause is not considered (Clarke et al., 2019). These methods seldomly address the ignition types that we are interested in.

Analytical modelling is another common way to build bushfires models. The general framework for analysing bushfires ignitions is a generalised additive model (GAM). Bates, McCaw and Dowdy (2018) used it for predicting the number of lightning ignitions. Some studies include a logit link (FIXME: why is this different to previous sentence) to extend the model for bushfires ignition cause prediction (Read, Duff and Taylor, 2018; Zhang, Lim and Sharples, 2017). Mixed-effect models have also been considered to incorporate spatial dependence and weather factors (Duff, Cawson and Harris, 2018). Simpler models have also been used, including multiple linear regression, negative binomial regression and generalised logistic regression (Cheney, Gould, McCaw and Anderson, 2012; Plucinski, McCaw, Gould and Wotton, 2014; Collins, Price and Penman, 2015). Particularly, instead of using a model, some researches performed statistical testing and exploratory data analysis to test hypotheses about bushfires (Miller et al., 2017; Dowdy, Fromm and McCarthy, 2017).

Common covariates for ignition analysis are weather conditions, vegetation types, topographic information and anthropogenic variables. In addition, various fire danger indexes have been used in modelling. Some studies choose to use index variables developed by McArthur such as Forest Fire Danger Index (Clarke et al., 2019; Read et al., 2018), while others choose to use indexes developed by Canadian Forestry Service such as Canadian Fire Weather Index and Drought Code (Plucinski et al., 2014). We doubt that these indexes are useful for fire ignition prediction because they are extracted from weather and vegetation information. However, it may help improve our model performance because it can be viewed as generated from feature engineering.

Although numerous studies for ignitions analysis have applied semiparametric and parametric methods, little analytics attention has been paid to more complex models like the tree-based model, support vector machine and artificial neural network. These tools are well developed in machine learning, which will also be considered for this research.

Most of the existing works of bushfire ignition analysis focus on a certain areas such as south-eastern Australia (Clarke et al., 2019), or a certain state such as Victoria (Read, Duff & Taylor, 2018). Little of existing works have applied prediction across Australia.

4 Project design

4.1 Data collection

A focus of this work is to utilise open source data, and collate these data sets to provide a data fusion with which to tackle the research questions. The motivating data source is satellite hotspots, which is different from what has been analysed previously in the literature. The next paragraphs describe how this and other data is accessed, and pre-processed for later analysis. Table 3 summarises the data sources.

- **Hotspots:** Hotspot data is downloaded from the JAXA's Himawari-8 satellite. This satellite is positioned in geostationary orbit at 140 degrees east longitude, and the revisit period is 10-minute. Its management system - JAXA's P-Tree system, provides WildFire observation product with 2km spatial resolution. (P-Tree System, 2020). Details on how to download this data are provided by [ozjim post].
- **Weather:** Weather data were collected from the Australian Bureau of Meteorology, by using an R package - Bomrang (Adam, Mark, Hugh & Keith, 2020). Due to the limitation of APIs provided by the package, we crawled data from BOM's website for extra information.
- **Fuel layer:** To characterise the fuel we used forest of Australia (2018) from the Australian Bureau of Agriculture and Resource Economics and Sciences. It is a fuel layer contains the vegetation information across Australia.
- **Fire origins:** Fire origins are the existing records of historical bushfires ignition which is downloaded from the Department of Environment, Land, Water Planning. This dataset is used to understand the characteristic of each type of ignition. It helps us to identify the causes of 2019-2020 bushfires.

- Roads: FIXME
- Recreation sites: FIXME

4.2 Data preprocessing

4.2.1 Data types

The hotspot and weather data is in CSV format, which can be processed generally with the `tidyverse` tools (Wickham et al., 2020) in R. Fuel layer and other map data are presented as geospatial objects, which can be processed using the tools in the `sf` (Pebesma, 2020).

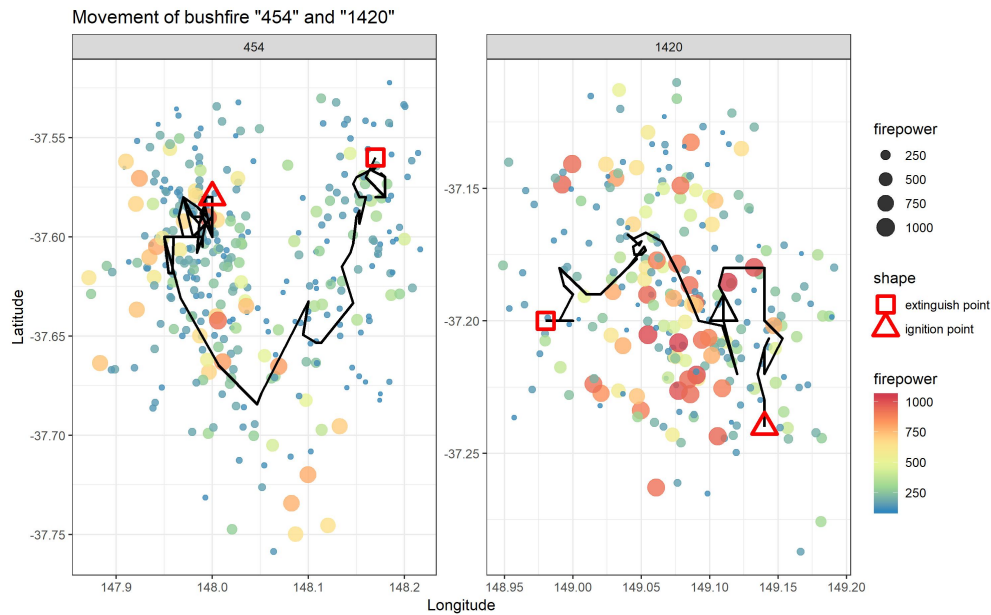
4.2.2 Identifying unique fires using spatiotemporal clustering

Hotspots are point data obtained by light detection from the satellite. They are snapshots of bushfires every 10 minutes. In order to study ignition, the data needs to be processed to identify the start, and movement of a fire.

The initial steps in the processing were to subset to only Australian locations, keep only hotspots above a threshold firepower (irradiance over 100 watt per square metre), and set the starting time to be when annual fire restrictions were lifted.

A clustering algorithm was developed to identify fire clusters, and details can be found in Table 4. Each hotspot a cluster membership which we called *fire_id*. Other characteristics of each fire id including its centroid, starting time, ending time and movement are recorded. The motivation for algorithm is the behaviour of real bushfire which can be summarised as two parameters, the potential distance a fire can spread in an hour, and the time a fire can stay smouldering but undetectable before flaring up again, represented by r_0 and t_0 , respectively. These two parameters have been used to determine if a new hotspot belongs to an existing or new cluster.

Figure ?? shows two fires and their life time movement as computed by the algorithm.



Compiled

data

The end result of the data pre-processing is a set of data tables, that are uniquely identified by individual primary key and related by a set of foreign keys. Figure 1 shows these tables and the keys which allow information to be compared across tables. Technically, this is called a conceptual entity relationship diagram (ERD). It will be useful for developing the data model to be used in both modelling and a web interface for communicating fire risk.

4.3 Exploratory data analysis

Preliminary examination of the compiled data to examine spatiotemporal trends and relationships between variables will be conducted using graphical and numerical summaries.

4.4 Modelling

Two types of models will be developed and tested. One is the ignition method prediction model, which is used for analysing ignition type of 2019-2020 bushfires and predicting the causes of future bushfires. Another is the fire risk model. It will learn the features of hotspots and predict the probability of hotspots occurrence.

- Predicting ignition method: Historical causes of bushfires will be used as training data. Multiclass response variable models, using vegetation, topographic and climate information as independent variables, will be developed and assessed. Statistical models, generalized linear model, and generalized additive model will be compared with these computational models, random forest, support vector machine and artificial neural networks.

- Modelling fire risk: The framework of fire risk modelling is similar to ignition method prediction. It will be a binary classification problem. The aim of this model is to predict the probability of bushfire ignition in a cell of the grid. The grid will be designed to fit the shape of Victoria with spatial resolution equal to 50km (Figure 2). The training data will be the hotspots data instead of bushfires history. Models considered will be the same as ignition method prediction.

5 Timeline

Table 1 summarises the work done to date. Table 2 maps out the plan for completing the thesis research.

Table 1: *Research plan till week 9*

Timeline	Tasks
Week 2	Geographic data background reading
Week 3	Collect Remote sensing data (JAXA himawari-8 satellite) and explore BOM weather data APIs (Bomrang)
Week 4	Collect Road Map (OpenStreetMap) and read articles in SpatioTemporal data visualization and modelling
Week 5	Develop clustering algorithm for remote sensing data
Week 6	Test diferent hyperparameters for clusetring
Week 7	Exploratory data analysis on fire clusters and data integration
Week 8	Feature planning for the shiny app
week 9	Write research proposal and prepare the first presentation

Table 2: *Research plan since June*

Timeline	Tasks
June - July	Modelling fire ignition and fire risk
August	Consolidate findings and create mockups of the shiny app
September	Develop the shiny app and perform different levels of testing
October	Write thesis and prepare the second presentation

6 Expected outcomes

It is expected that we will have probabilistic predictions for ignition cause for the 2019-2020 for each of the fires identified from the hotspot data. A useable model of fire risk will be produced and made accessible in an interactive web application.

Appendix A

Table 3: *Data information*

Data set name	Spatial Resolution	Temporal resolution	Time
Hotspots data - JAXA's Himawari-8 satellite	$0.02^{\circ} \approx 2km$	Per 10 minutes	2015-2020
Weather data - Australian Bureau of Meteorology		Daily	2019-2020
Map - OpenStreetMap	2m		2020
Fuel layer - Australian Bureau of Agriculture and Resource Economics and Sciences	100m		2018
Victorian CFA fire stations - Department of Environment, Land, Water & Planning	20m		2020
Victorian Recreation sites - Department of Environment, Land, Water & Planning	10m		2020
Fire Origins - Department of Environment, Land, Water & Planning	100m		1972-2019

Table 4: A clustering algorithm for hotspots

Algorithm 1 Hotspots cluterling

input: Hotspots dataset $H : (\text{Hour_id}^{(n)}, \text{Coordinates}^{(n)})$, $n = 1, 2, \dots, N$
 An empty dataset $F : (\text{Fire_id}^{(m)}, \text{Coordinates}^{(m)}, \text{Active}^{(m)})$, $m = 1, 2, \dots$
 An empty vector $K \in \mathbb{N}_1^n$
 A distance hyperparameter $r_0 \in \mathbb{R}^+$
 A time hyperparameter $t_0 \in \mathbb{N}^+$

output: A vector $K \in \mathbb{N}_1^n$ contains memberships of hotspots
 A dataset F contains fire clusters information including memberships, latest centroids and time from last updated

- 1: select subset $H_c \in H$ where $\text{Hour_id} == 1$
- 2: calculate distance matrix D for Coordinates in H_c
- 3: assign 1 to a zero adjacency matrix A for where $D \leq r_0$ // hotspots with relative distance less or equal to r_0 will be considered belong to the same cluster
- 4: create undirected unweighted graph G from A
- 5: record memberships of G to K
- 6: record clusters classes to Fire_id and record clusters centroids to Coordinates of F
- 7: set Active in F to t_0 // Active clusters are fire being observed in the last t_0 hour
- 8: **for** $\text{hour} = 2, \dots, T$ **do**
- 9: let $\text{Active} - 1$ and select subset $F_c \in F$ where $\text{Active} \geq 0$
- 10: select subset $H_c \in H$ where $\text{Hour_id} == \text{hour}$
- 11: append Coordinates from F_c to H_c
- 12: repeat step 2 - 4
- 12: **for** $h_i = \text{each hotspot in } H_c$ **do**
- 13: **if** h_i share the same membership as one of active clusers in F_c **then**
- 14: copy the corresponding Fire_id of the nearest active cluser to K
- 15: **else** copy the membership from G to K
- 16: **end if**
- 17: **end for**
- 18: update F for clusters involed in current timestamp and reset corresponding Active to t_0
- 19: **end for**

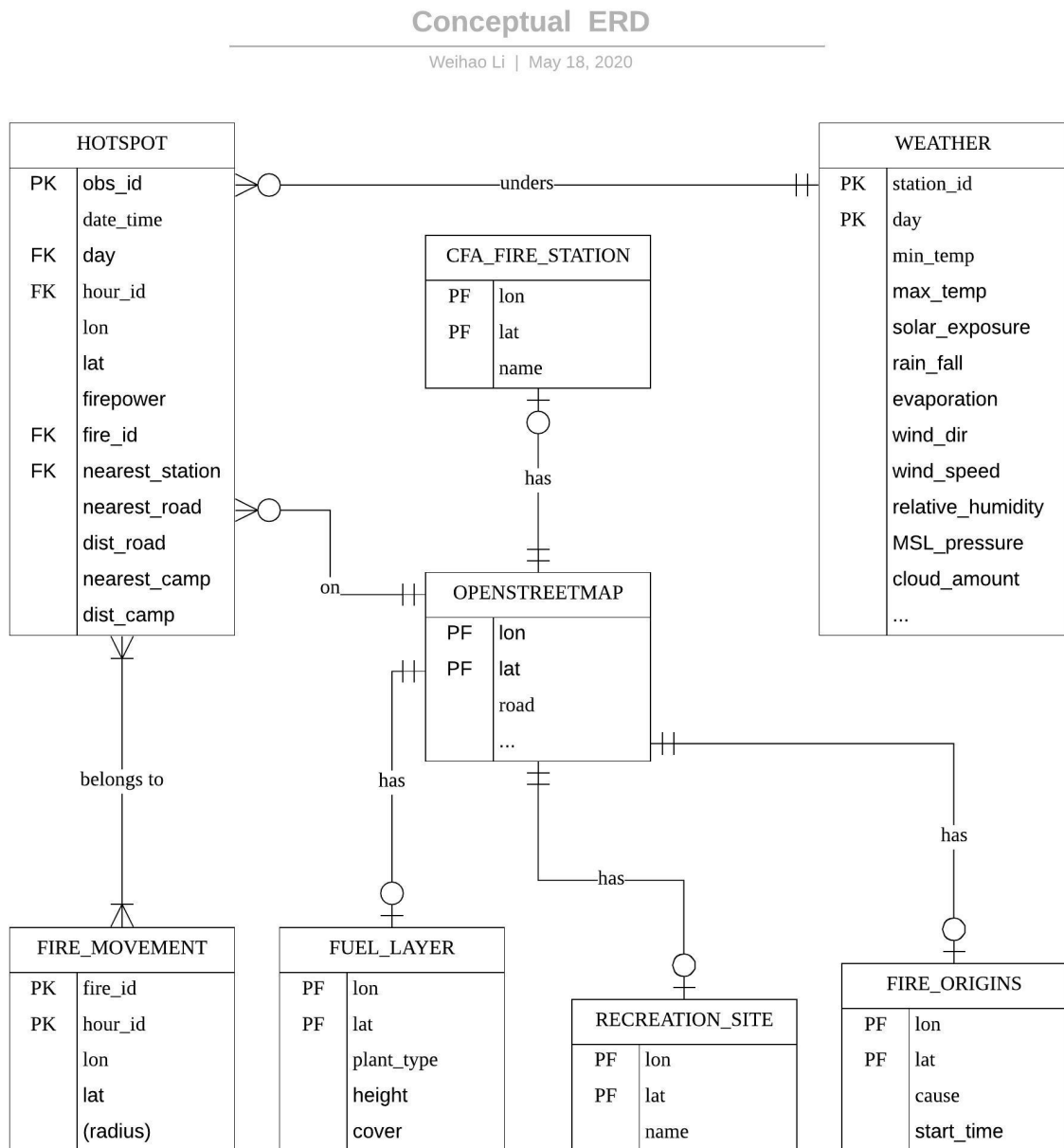


Figure 1: Entity relationship diagram illustrating the relational tables of the compiled data. Tables correspond to processed hotspot data, weather, local facilities like CFA sites. This data structure is useful for the data modeling and web app development.

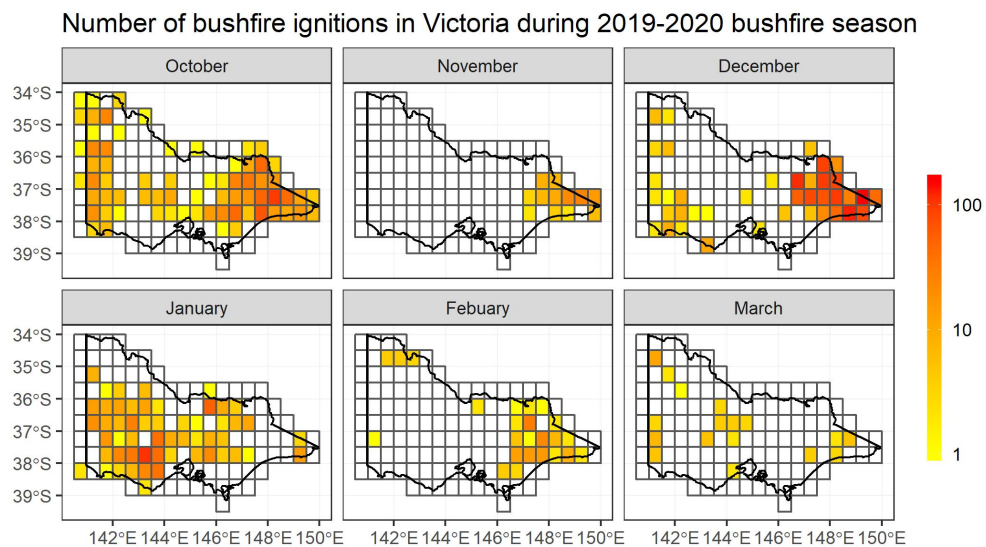


Figure 2: The grid map illustrating the general situation of Victoria during 2019-2020 Australia bushfire season. The severest time for Victoria was December when the massive amount of bushfires ignited in Eastern Victoria. Places like East Gippsland suffered from this devastating crisis.