MONASH BUSINESS SCHOOL

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

Weihao Li
EBS Honours student

Research Plan
for Honours, supervised by Dianne Cook and Emily Dodwell

Department of Econometrics & Business Statistics

20 July 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 were greater than $A10 million (McAneney, Chen, & Pitman, 2009). The 2019-2020 Australian bushfires, compared with other major bushfires in history, had a more devastating impact on the environment and properties. According to a report by Lisa Richards, Nigel Brew and Smith (2020), published by the Parliament of Australia, 3094 houses were destroyed and over 17M hectares of land burned. These two figures are the highest in history. Fortunately, fewer lives were lost: 33 (including firefighters) compared to 173 in the Black Saturday fires and 47 in Ash Wednesday fires (Lisa Richards, Nigel Brew & Smith, 2020).

The research is also motivated by the lack of knowledge about the cause of bushfire in Australia. Beale and Jones (2011) 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% by natural, such as lightning. Understanding the cause of the 2019-2020 Australian bushfires is a crucial step to provide information for future policy and fire risk management.

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. Satellite data provides the opportunity to more objectively study ignition locations, particularly in less accessible regions.

The overall research aim is to provide probabilistic estimates of the cause of fire ignition for the 2019-2020 bushfires. Several smaller research questions that will be addressed 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 sites 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) developed landscape fire succession models (LFSMs) to simulate spatial fire behaviour, 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- and time-effective ways to model bushfires, but ignition cause is not considered (Clarke et al., 2019). These methods seldomly addressed the ignition types that we are interested in.

Analytical modelling is another common way to build bushfires models. The general framework for analysing bushfires ignition is a generalised additive model (GAM). Bates, McCaw, and Dowdy (2018) used it with a Gaussian distribution for predicting the number of lightning ignitions. Some studies implemented GAM with a multinomial distribution to predict ignition probability (Read, Duff, & Taylor, 2018; Zhang, Lim, & Sharples, 2017). Mixed-effect models have also been considered to incorporate spatial dependence and weather factors (Duff, Cawson, & Harris, 2018). Simpler models have also been used, including multiple linear regression, negative binomial regression and generalised logistic regression (Cheney et al., 2012; Plucinski et al., 2014; Collins, Price, & Penman, 2015). Some research used statistical tests and exploratory methods to assess hypotheses about bushfires (Miller et al., 2017; Dowdy, Fromm, & McCarthy, 2017).

Common covariates for ignition analysis are weather conditions, vegetation types, topographic information and environmental impact of human activities. 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, Duff, & Taylor, 2018), while others choose to use indexes developed by the 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, this type of feature engineering could improve the models.

Although numerous studies for ignitions analysis have applied semi-parametric and parametric methods, little use of machine learning models has been made. These approaches will also be considered for this research.

Most of the existing work on bushfire ignition cause focuses on south-eastern Australia (Clarke et al., 2019), and Victoria (Read, Duff, & Taylor, 2018). Models have not been applied 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. This data is collated with weather records, fuel layer, location of roads and recreation sites, and historical data on ignition cause. The spatial and temporal details of each dataset are provided in Table 3. Descriptions and links to the data sources can be found in Table 4. These two tables are provided in the Appendix.

4.2 Data pre-processing

4.2.1 Data types

The hotspot and weather data is in CSV format, which can be processed generally with the `tidyverse` (Wickham, 2017) tools in R (R Core Team, 2019). Fuel layer and other map data are presented as geospatial objects, which can be processed using the tools in the `sf` (Pebesma, 2018).

4.2.2 Identifying unique fires using spatiotemporal clustering

Hotpots are light detections from the satellite, provided at 10-minute intervals. To study ignition, it needs to be processed to identify the start, indicate the likelihood of being the same fire, and track the movement of a fire. Not all hotspots are fires, and the recommended threshold of fire power over 100 (irradiance over 100 watts per square metre) is used to filter the data. Only data recorded after the published fire restrictions dates are used.

A clustering algorithm was developed to group hotspots into unique fires. The details are in Table 5 in the Appendix. Each hotspot will be assigned a cluster membership which is called the *fire_id*. Other characteristics of each fire, including its centroid, starting time, ending time and movement are recorded. The motivation for this algorithm is bushfire dynamics which can be summarised using two parameters, the potential distance a fire can spread in an hour, and the time a fire can stay smouldering but undetectable by satellite 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 1 shows two fires and their lifetime movement as computed by the algorithm.

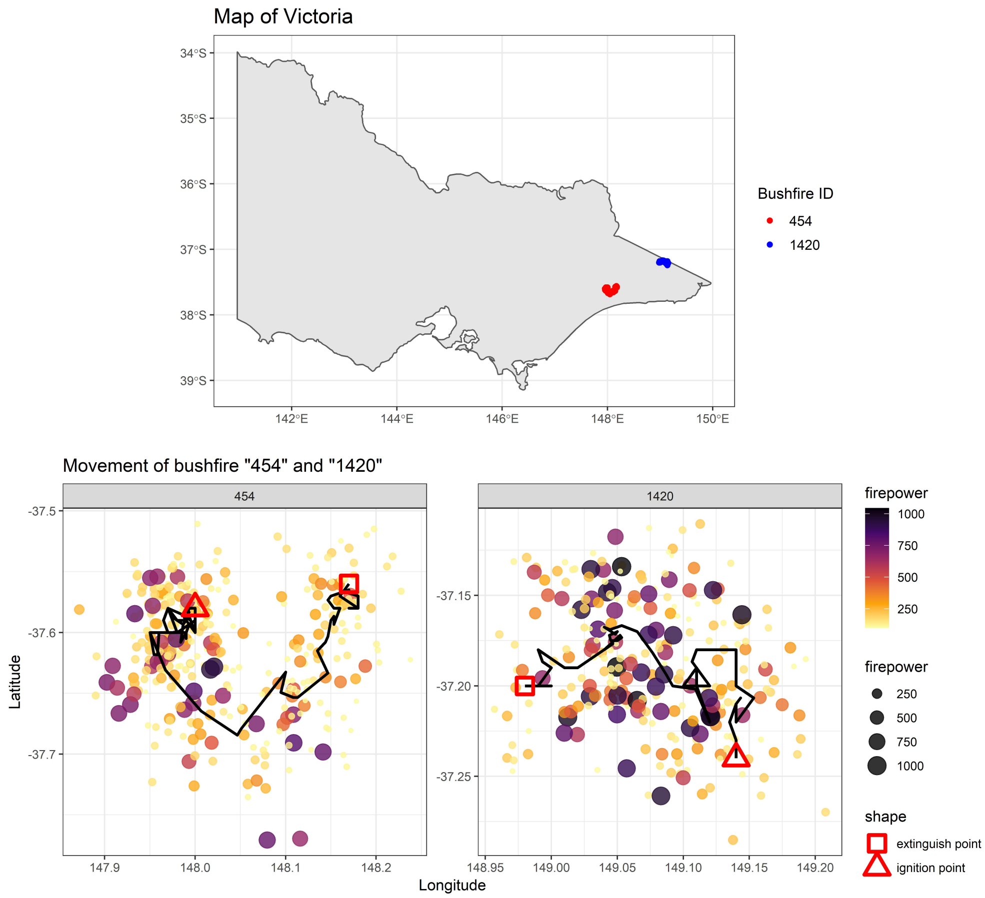

Figure 1: *The movement path and hotspots map shows the bushfire behaviour of bushfire "454" and bushfire "1420", which is two clusters from the results of the clustering algorithm. The triangle is the ignition point of bushfire and the rectangle is the extinguish point. The black line is the movement of bushfire centroid. Points with red color and larger size are hotspots with higher fire power which is measured by watt per square metre. We can know that bushfires ignited with low fire power, and will get hotter when time pass, then finally extinguish with not enough fuel.*

4.2.3 Compiled data

The end result of the data pre-processing is a set of data tables, that are uniquely identified by an individual primary key and related by a set of foreign keys. Figure 2 summarizes 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 a future hotspot occurring.

- Predicting ignition method: Historical causes of bushfires will be used as training data. Multi-class 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 the ignition method prediction, although it will be a binary classification problem. This model aims to predict the probability of bushfire ignition in the cell of a spatial grid. The grid will be designed to fit the shape of Victoria with spatial resolution equal to 50km (Figure 3). 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: *Completed work*

Timeline	Tasks
Week 2	Geographic data analysis 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 different hyperparameters for clustering
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*

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

6 Expected outcomes

It is expected that we will have probabilistic predictions for ignition cause for each of the fires identified from the hotspot data during 2019-2020 Australian bushfire season. A usable 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	$0.02° \approx 2km$	Per 10 minutes	2015-2020
Weather data		Daily	2019-2020
Road Map	2m		2020
Fuel layer	100m		2018
Victorian CFA fire stations	20m		2020
Victorian Recreation sites	10m		2020
Fire Origins	100m		1972-2019

Table 4: *Description of datasets*

Dataset	Description
Hotspots data	Hotspot data is downloaded from P-Tree System (2020). Details on how to download this data are provided by Williamson (2020)
Weather data	Weather data were collected from the Australian Bureau of Meteorology, by using an R package - Bomrang (Sparks et al., 2020)
Fuel layer	To characterise the fuel we used forest of Australia from Australian Bureau of Agricultural and Resource Economics and Sciences (2018)
Fire Origins	Fire origins are the existing records of historical bushfires ignition which is downloaded from Department of Environment, Land, Water & Planning (2019)
Road Map	Roads map are features contained in OpenStreetMap provided by OpenStreetMap contributors (2020)
Victorian Recreation sites	Recreation sites are location information of picnic and camping sites within Victorian forest, which is downloaded from Department of Environment, Land, Water & Planning (2020-b)
Victorian CFA fire stations	CFA stations are location information of CFA fire station within Victoria, which is downloaded from Department of Environment, Land, Water & Planning (2020-a)

Table 5: *A clustering algorithm for hotspots*

Algorithm 1 Hotspots clutering

---

input: Hotspots dataset H : (Hour_id$^{(n)}$, Coordinates$^{(n)}$), n = 1, 2, ... N
An empty dataset F : (Fire_id$^{(m)}$, Coordinates$^{(m)}$, Active$^{(m)}$), m = 1, 2, ...
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 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 a relative distance less or equal to r_0 will be considered belonging 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** hour = 2, ... T **do**
9 : let Active -1 and select subset $F_c \in$ F where Active ≥ 0
10 : select subset $H_c \in$ H where Hour_id == hour
11 : append Coordinates from F_c to H_c
12 : repeat step 2 - 4
12 : **for** h_i = each hotspot in H_c **do**
13 : **if** h_i share the same membership as one of active clusters in F_c **then**
14 : copy the corresponding Fire_id of the nearest active cluster to K
15 : **else** copy the membership from G to K
16 : **end if**
17 : **end for**
18 : update F for clusters involved in current time-stamp and reset corresponding Active to t_0
19 : **end for**

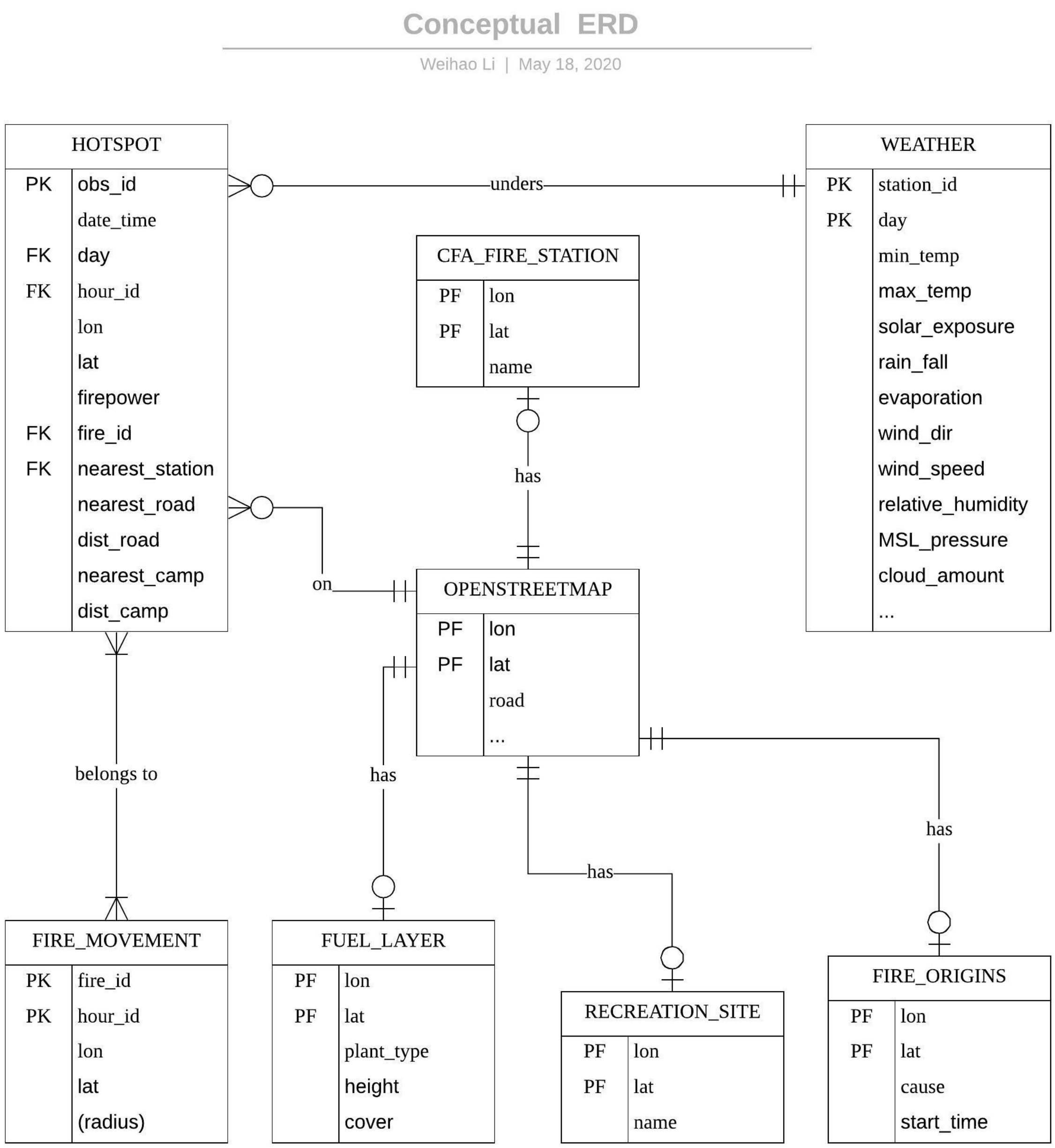

Figure 2: *Entity relationship diagram illustrating the relational tables of the compiled data. Tables correspond to processed hotspot data, weather, local facilities like CFA sites. This data structure is useful for data modelling and web app development.*

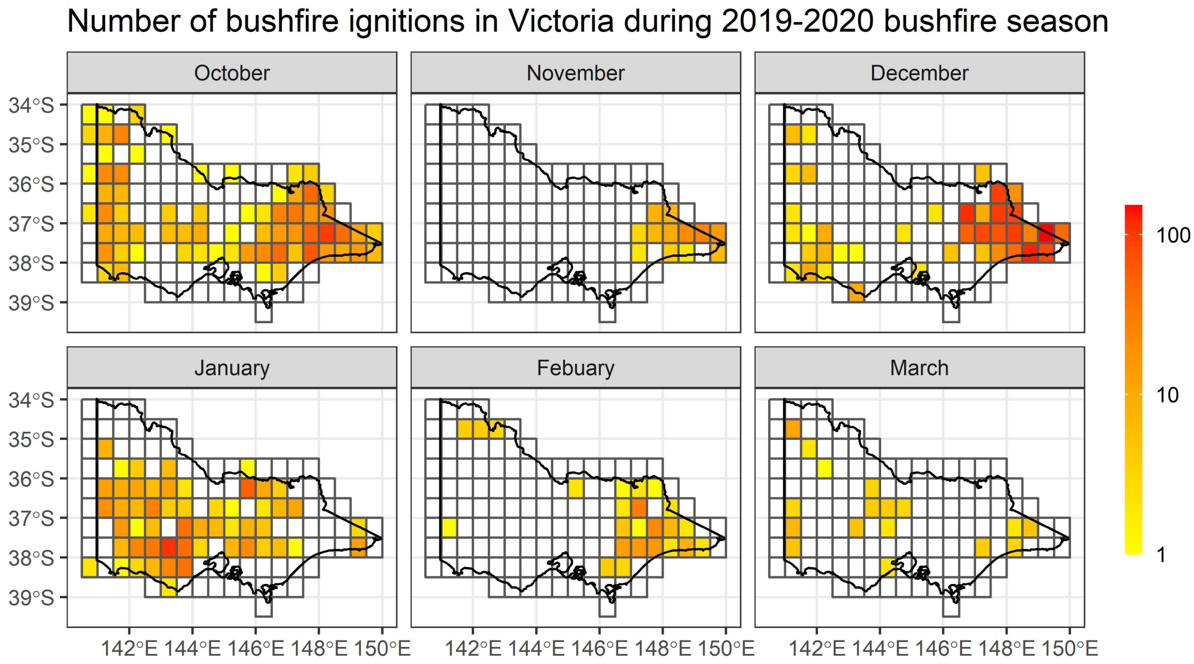

Figure 3: *The grid map illustrating the general situation of Victoria during 2019-2020 Australia bushfire season. The severest time for Victoria was December when the massive amount of bushfires ignited in Eastern Victoria. Places like East Gippsland suffered from this devastating crisis.*

References

Australian Bureau of Agricultural and Resource Economics and Sciences. (2018). Forests of Australia. Retrieved May 21, 2020, from https://www.agriculture.gov.au/abares/forestsaustralia/forest-data-maps-and-tools/spatial-data/forest-cover

Bates, BC, McCaw, L, & Dowdy, AJ. (2018). Exploratory analysis of lightning-ignited wildfires in the warren region, western australia. *Journal of environmental management*, **225**, 336–345.

Beale, J, & Jones, W. (2011). Preventing and reducing bushfire arson in australia: A review of what is known. *Fire technology*, **47**(2), 507–518.

Bradstock, RA, Cary, GJ, Davies, I, Lindenmayer, DB, Price, OF, & Williams, RJ. (2012). Wildfires, fuel treatment and risk mitigation in australian eucalypt forests: Insights from landscape-scale simulation. *Journal of Environmental Management*, **105**, 66–75.

Cheney, NP, Gould, JS, McCaw, WL, & Anderson, WR. (2012). Predicting fire behaviour in dry eucalypt forest in southern australia. *Forest Ecology and Management*, **280**, 120–131.

Clarke, H, Gibson, R, Cirulis, B, Bradstock, RA, & Penman, TD. (2019). Developing and testing models of the drivers of anthropogenic and lightning-caused wildfire ignitions in south-eastern australia. *Journal of environmental management*, **235**, 34–41.

Collins, KM, Price, OF, & Penman, TD. (2015). Spatial patterns of wildfire ignitions in south-eastern australia. *International Journal of Wildland Fire*, **24**(8), 1098–1108.

Department of Environment, Land, Water & Planning. (2019). Fire Origins - Current and Historical. Retrieved May 21, 2020, from https://discover.data.vic.gov.au/dataset/fire-origins-current-and-historical

Department of Environment, Land, Water & Planning. (2020-a). CFA - Fire Station. Retrieved May 21, 2020, from https://discover.data.vic.gov.au/dataset/cfa-fire-station-vmfeat-geomark_point

Department of Environment, Land, Water & Planning. (2020-b). Recreation Sites. Retrieved May 21, 2020, from https://discover.data.vic.gov.au/dataset/recreation-sites

Dowdy, AJ, Fromm, MD, & McCarthy, N. (2017). Pyrocumulonimbus lightning and fire ignition on black saturday in southeast australia. *Journal of Geophysical Research: Atmospheres*, **122**(14), 7342–7354.

Duff, TJ, Cawson, JG, & Harris, S. (2018). Dryness thresholds for fire occurrence vary by forest type along an aridity gradient: Evidence from southern australia. *Landscape Ecology*, **33**(8), 1369–1383.

Keane, RE, Cary, GJ, Davies, ID, Flannigan, MD, Gardner, RH, Lavorel, S, Lenihan, JM, Li, C, & Rupp, TS. (2004). A classification of landscape fire succession models: Spatial simulations of fire and vegetation dynamics. *Ecological modelling*, **179**(1), 3–27.

Lisa Richards, Nigel Brew, & Smith, L. (2020). *2019-20 Australian bushfires-frequently asked questions: a quick guide*. Parliament of Australia. https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp1920/Quick_Guides/AustralianBushfires

McAneney, J, Chen, K, & Pitman, A. (2009). 100-years of australian bushfire property losses: Is the risk significant and is it increasing? *Journal of environmental management*, **90**(8), 2819–2822.

Miller, C, Plucinski, M, Sullivan, A, Stephenson, A, Huston, C, Charman, K, Prakash, M, & Dunstall, S. (2017). Electrically caused wildfires in victoria, australia are over-represented when fire danger is elevated. *Landscape and Urban Planning*, **167**, 267–274.

OpenStreetMap contributors. (2020). Planet dump retrieved from https://planet.osm.org.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*.

Plucinski, M, McCaw, W, Gould, J, & Wotton, B. (2014). Predicting the number of daily human-caused bushfires to assist suppression planning in south-west western australia. *International Journal of Wildland Fire*, **23**(4), 520–531.

P-Tree System. (2020). JAXA Himawari Monitor - User's Guide. Retrieved May 21, 2020, from https://www.eorc.jaxa.jp/ptree/userguide.html

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Read, N, Duff, TJ, & Taylor, PG. (2018). A lightning-caused wildfire ignition forecasting model for operational use. *Agricultural and Forest Meteorology*, **253**, 233–246.

Sparks, AH, Carroll, J, Goldie, J, Marchiori, D, Melloy, P, Padgham, M, Parsonage, H, & Pembleton, K. (2020). *bomrang: Australian government bureau of meteorology (bom) data client* [R package version 0.7.0]. https://CRAN.R-project.org/package=bomrang

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'* [R package version 1.2.1]. https://CRAN.R-project.org/package=tidyverse

Williamson, G. (2020). Example code to generate animation frames of Himawari-8 hotspots. Retrieved May 21, 2020, from https://gist.github.com/ozjimbob/80254988922140fec4c06e3a43d069a6

Zhang, Y, Lim, S, & Sharples, JJ. (2017). Wildfire occurrence patterns in ecoregions of new south wales and australian capital territory, australia. *Natural Hazards*, **87**(1), 415–435.