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

Weihao Li

EBS Honours student, Supervised by Dianne Cook and Emily Dodwell

Department of Econometrics and Business Statistics

wlii0039@student.monash.edu

16 Oct 2020

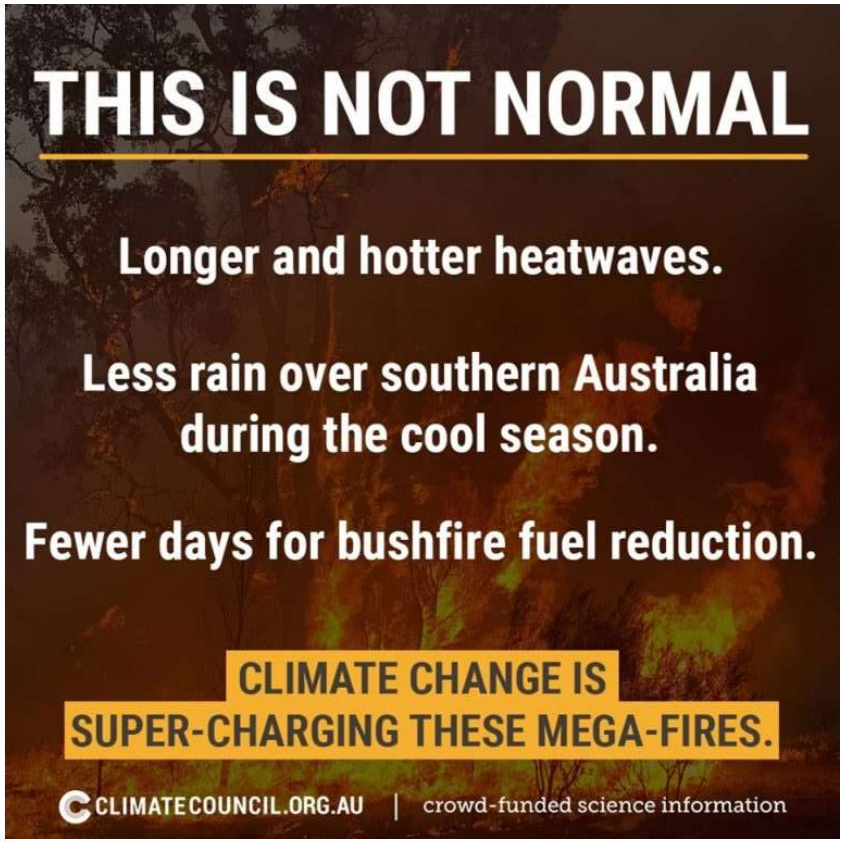
Introduction

2019-2020 Australia bushfire season

- A catastrophic and unprecedented disaster.
- 3094 houses were destroyed, 17M hectares of land burned and 33 lives were lost.

Introduction

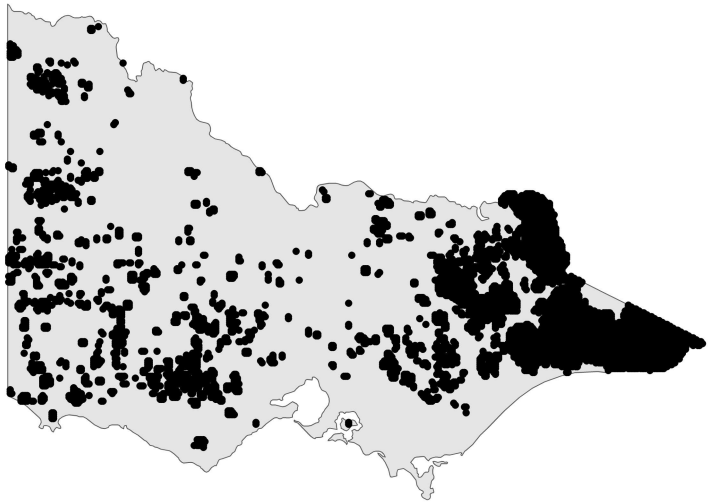
Lightning or Arson?

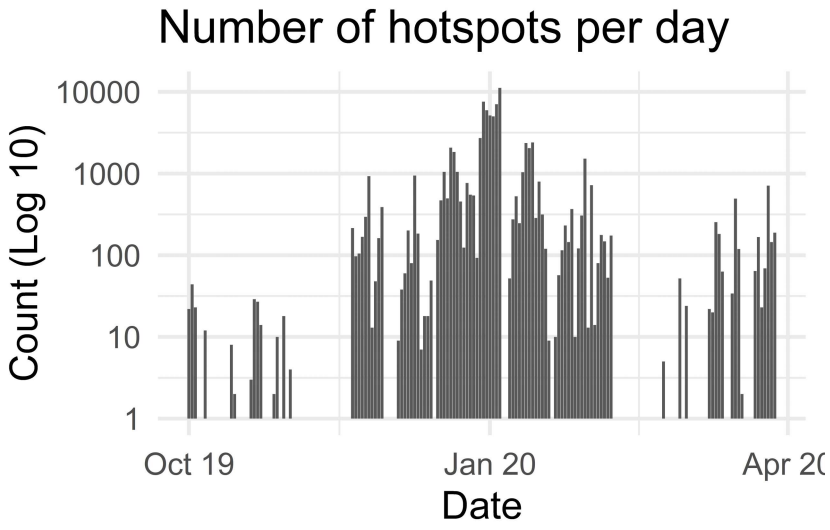


Remote sensing data

Remote sensing data are the reflected energy from the earth collected by remote sensors carried by a satellite or an aircraft.

Japan Aerospace Exploration Agency provides a hotspot product taken from the **Himawari-8** satellite.





🎯 Research questions

(1) Can we detect **bushfire ignitions** from hotspot data?

(2) Can we accurately predict the **causes** of bushfires during the 2019-2020 Australia bushfire season?

Literature review

Loboda and Csiszar (2007)

- Fire Spread Reconstruction (**FSR**)
- Clustered hotspots in a temporal manner to detect the bushfire ignitions

Read, Duff, and Taylor (2018):

- Fitted a logistic regression generalised additive model to lightning-caused ignitions in Victoria
- Used weather conditions, vegetation types, topographic information and environmental impact of human activities as covariates

Others: Ester et al. (1996), Zhang, Lim, and Sharples (2017), Zumbrunnen et al. (2012)

Data Sources

Historical fire origins: Department of Environment, Land, Water and Planning

Remote sensing data: Japan Aerospace Exploration Agency

Supplementary

Wind speed data: Commonwealth Scientific and Industrial Research Organisation and Automated Surface Observing System

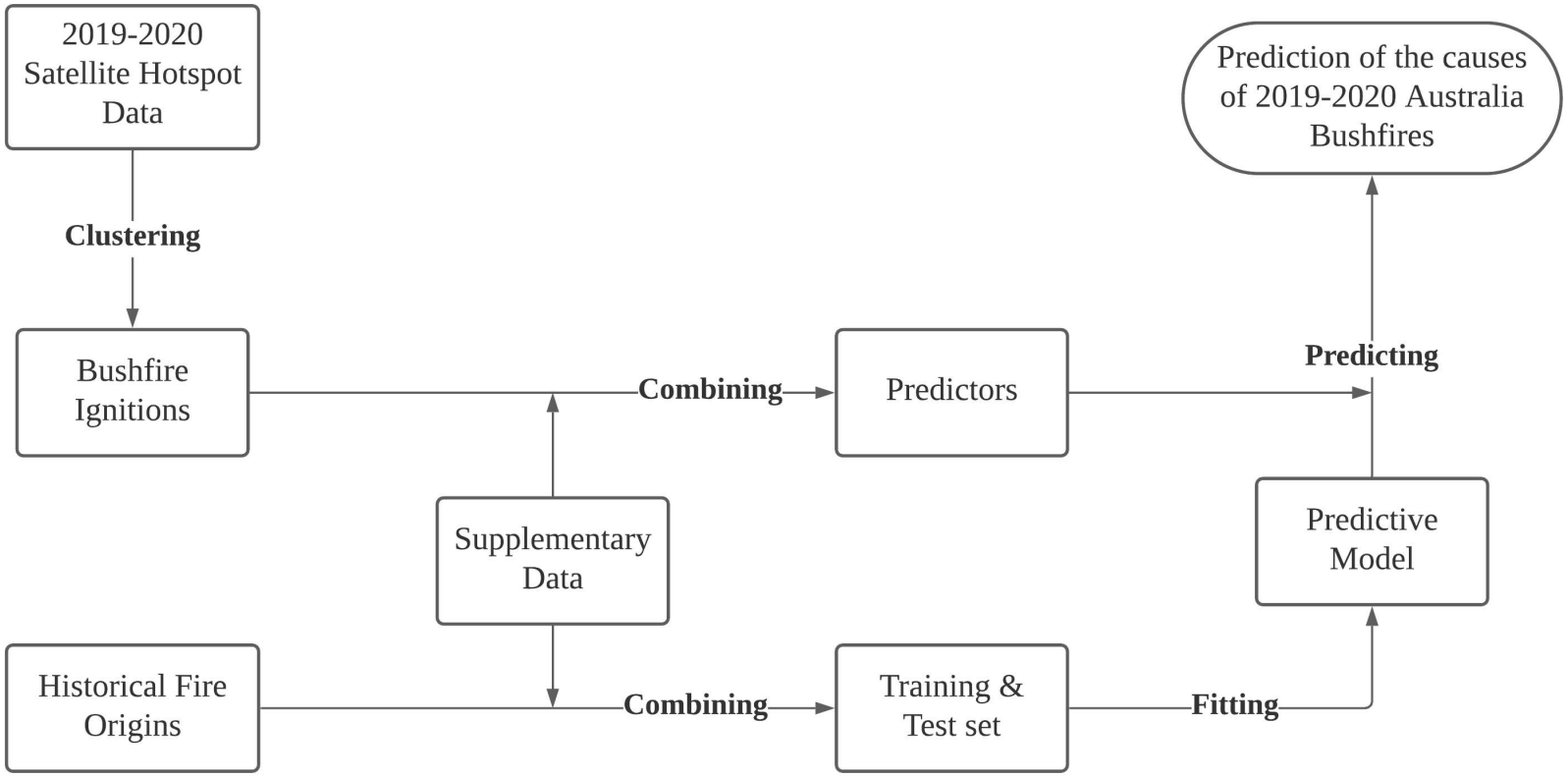
Temperature, Rainfall and Solar exposure: Bureau of Meteorology

Road map: OpenStreetMap

Fuel layer: Australian Bureau of Agricultural and Resource Economics

Fire stations and Recreation sites: Department of Environment, Land, Water and Planning

Diagram


2019-2020 Satellite Hotspot Data
Clustering
Bushfire Ignitions
Combining
Predictors
Predicting
Prediction of the causes of 2019-2020 Australia Bushfires
Supplementary Data
Predictive Model
Historical Fire Origins
Combining
Training & Test set
Fitting

Detect bushfire ignitions from hotspot data

We developd a spatio-temporal clustering algorithm to track bushfires from noisy satellite hotspot data.

1 Divide hotspots into different blocks

$$[max(t-24,1),t], t = 1, 2, \ldots, T\ (hours)$$

2 Cluster hotspots in each block

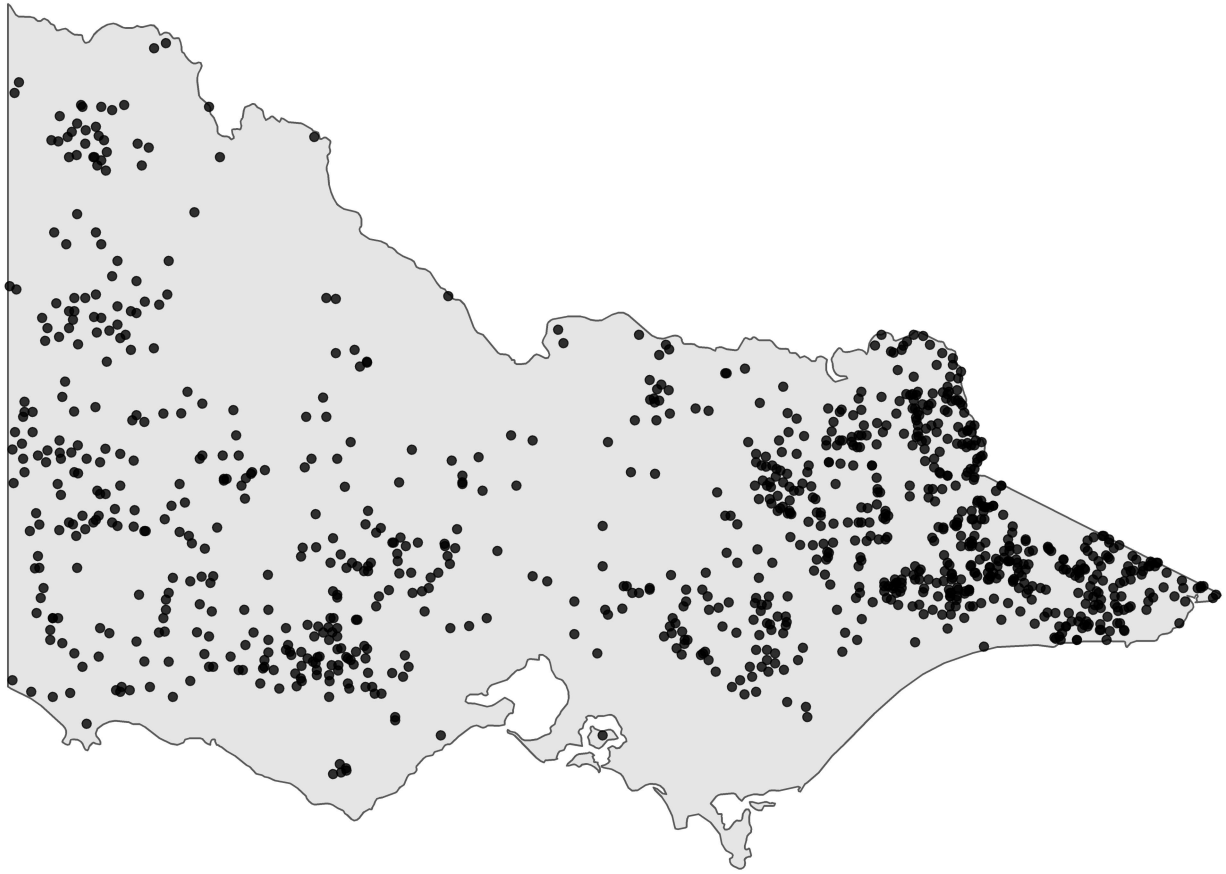
- Connect two nodes if geodesic between two nodes <= 3km.
- A connected component is a cluster.

3 Update the clustering results recursively

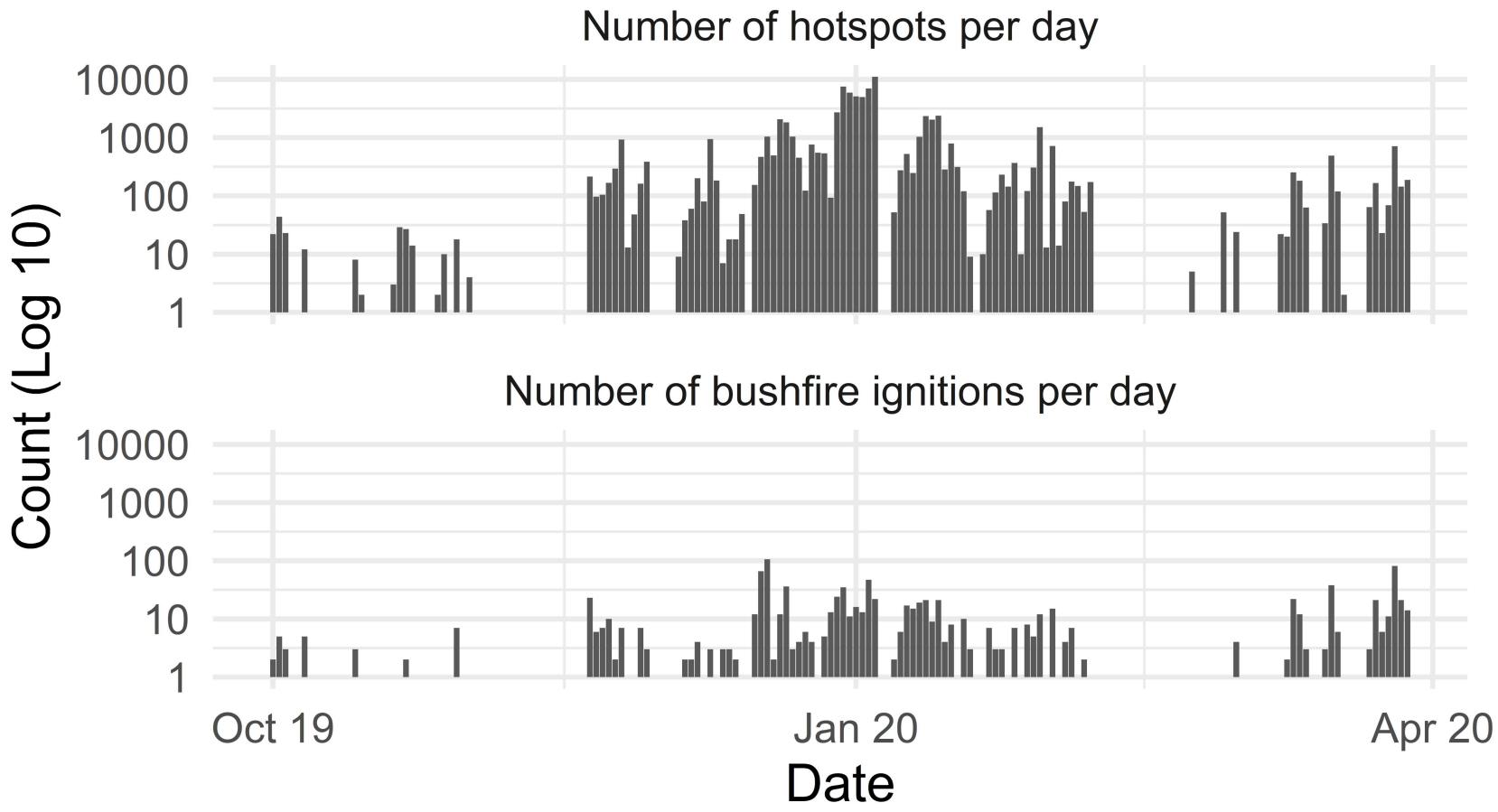
- Start from the second hour, till the end.
- If a hotspot is also in the last hour block, update its membership. And then hotspots who share the component with it will also be updated.

4 Compute ignition locations

Clustering result



Clustering result



Data integration

🚀 Used **`raster`** to extract values in **Grid** format data by projecting the bushfire ignitions on the grid:

- Forest type, forest height class, forest crown cover and average daily wind speed

🚀 Used **`sf`** to calculate the geodesic in **Simple features GIS** data for every bushfire ignition:

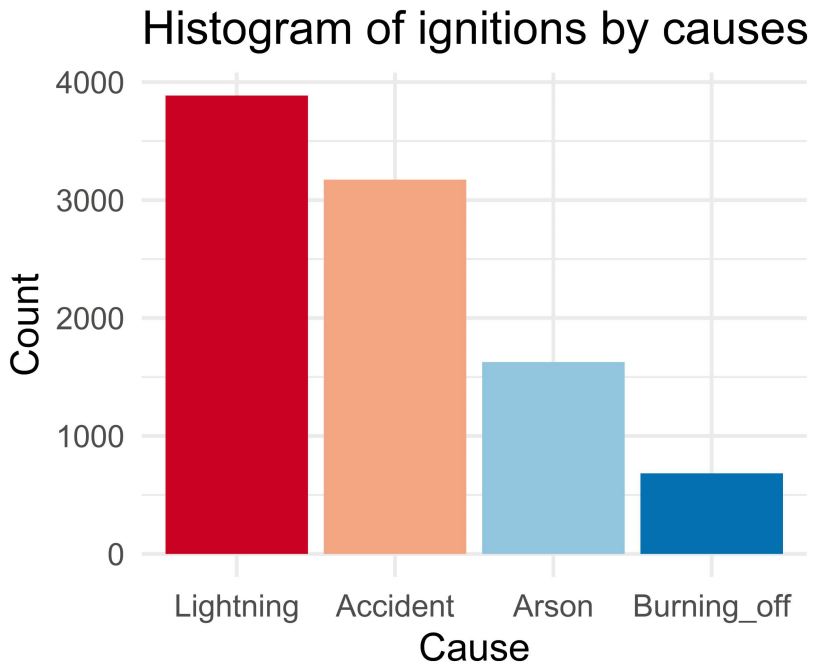
- Proximity to the nearest CFA station, the nearest camping site and the nearest road

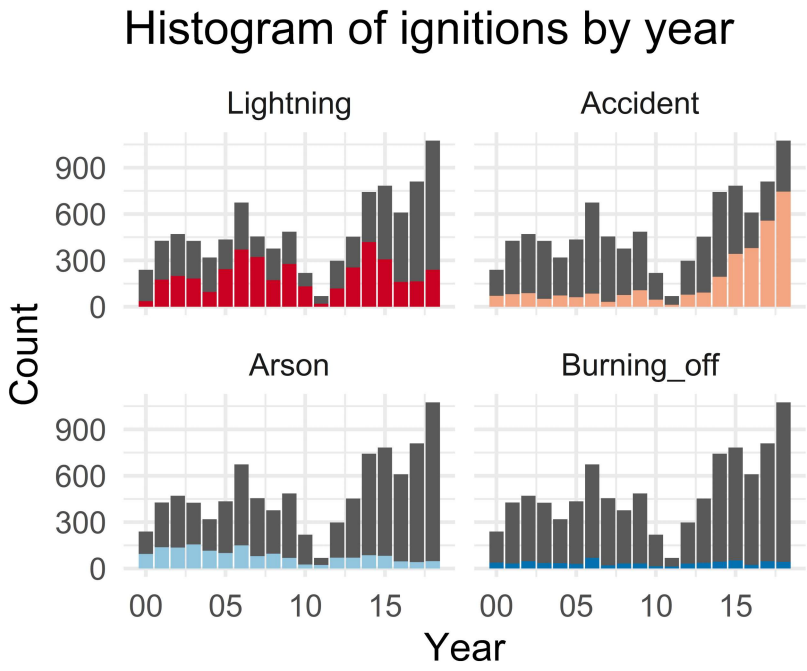
🚀 Used **`tidyverse`** to summarize the weather statistics:

- Temperature, rainfall, solar exposure and wind speed
  - 1-day average, 7-day average, ..., 2-year average

Overview of historical bushfire ignitions

Lightning and accident were the two main sources of historical bushfire ignitions, which took up 41% and 34% respectively. There were 17% bushfires caused by arson.

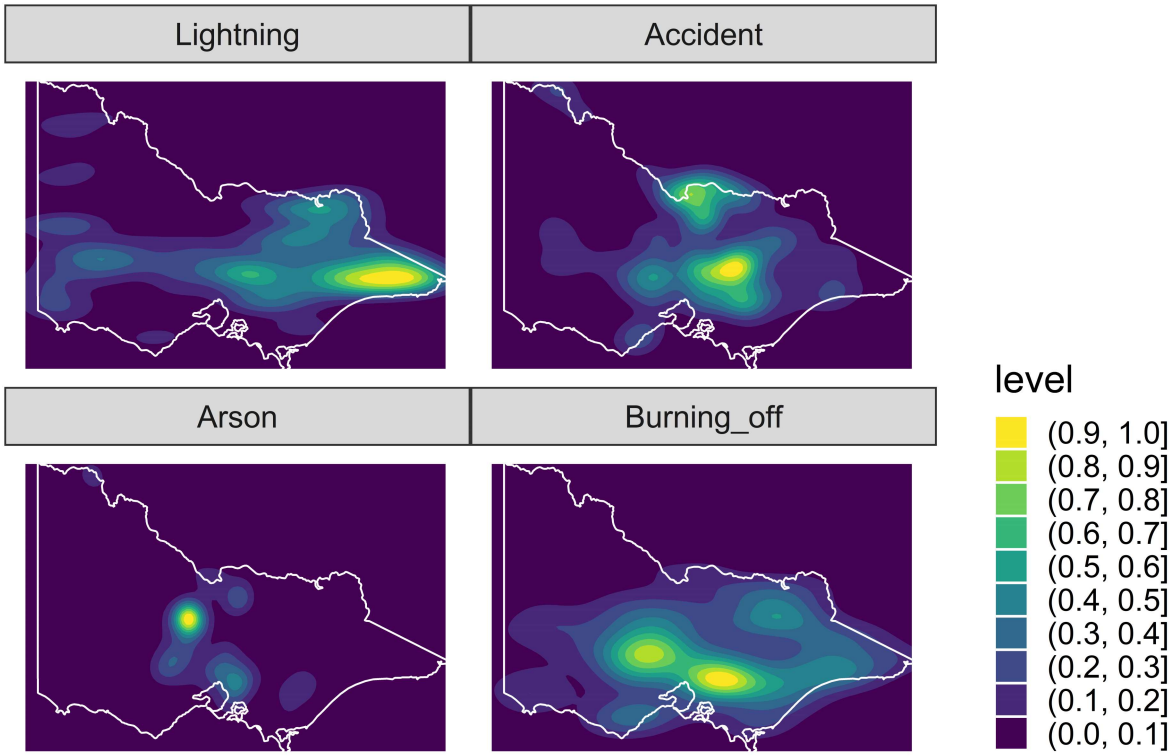




Spatial distribution of historical ignitions

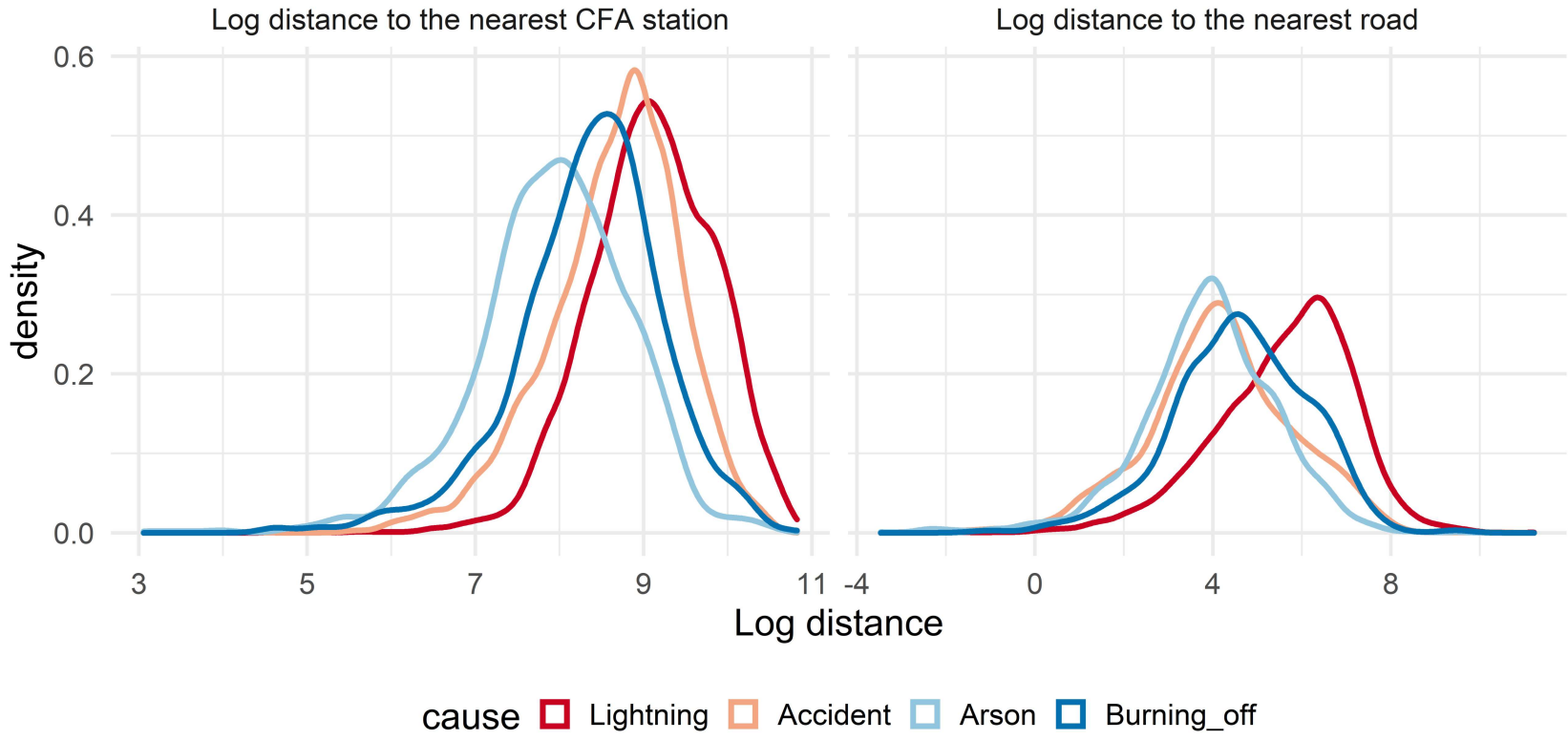
Lightning-caused bushfires were mostly in the east of Victoria. Bushfires caused by arson were near Bendigo.

2D conditional density plot of historical bushfire ignitions



Proximity

Lightning-caused bushfires were far away from the CFA station and road which made it less reachable. In contrast, bushfires caused by arson were close to CFA station and road.



Modelling

We used a **random forest** model to classify different causes of bushfire ignition.

Random forest (Breiman, 2001) is an ensemble learning method for building tree-based prediction models. It generates a certain number of decision trees by using the bootstrap aggregating technique and take the majority vote in the case of decision trees as the prediction.

We kept 80% of the data as training set, which was 7497 observations.

And the remaining 1872 observations was used as test set.

Other candidate models were also tested, but their performance was worse than random forest.

Model	Accuracy	Muti-class AUC
Multinomial logistic regression	0.5272	0.7424
GAM multinomial logistic regression	0.6779	0.8233
Random forest	0.7495	0.8795
XGBoost	0.7388	0.8752

Feature selection

We needed to select the most important features out of 55, but the global variable importance could be bias and misleading.

We used **`lime`**

Assumes machine learning models are linear at the local scope

1 Randomly sample data points around an given observation

2 Pass sample data points to the model to obtain predictions

3 Fit a lasso regression on the predictions with the sample data points

4 Select the most important variables from the regularization path

5 repeat 1-4 for 100 observations. Variables being selected most frequently are most important . We kept the top 10 most important variables.

Model performance

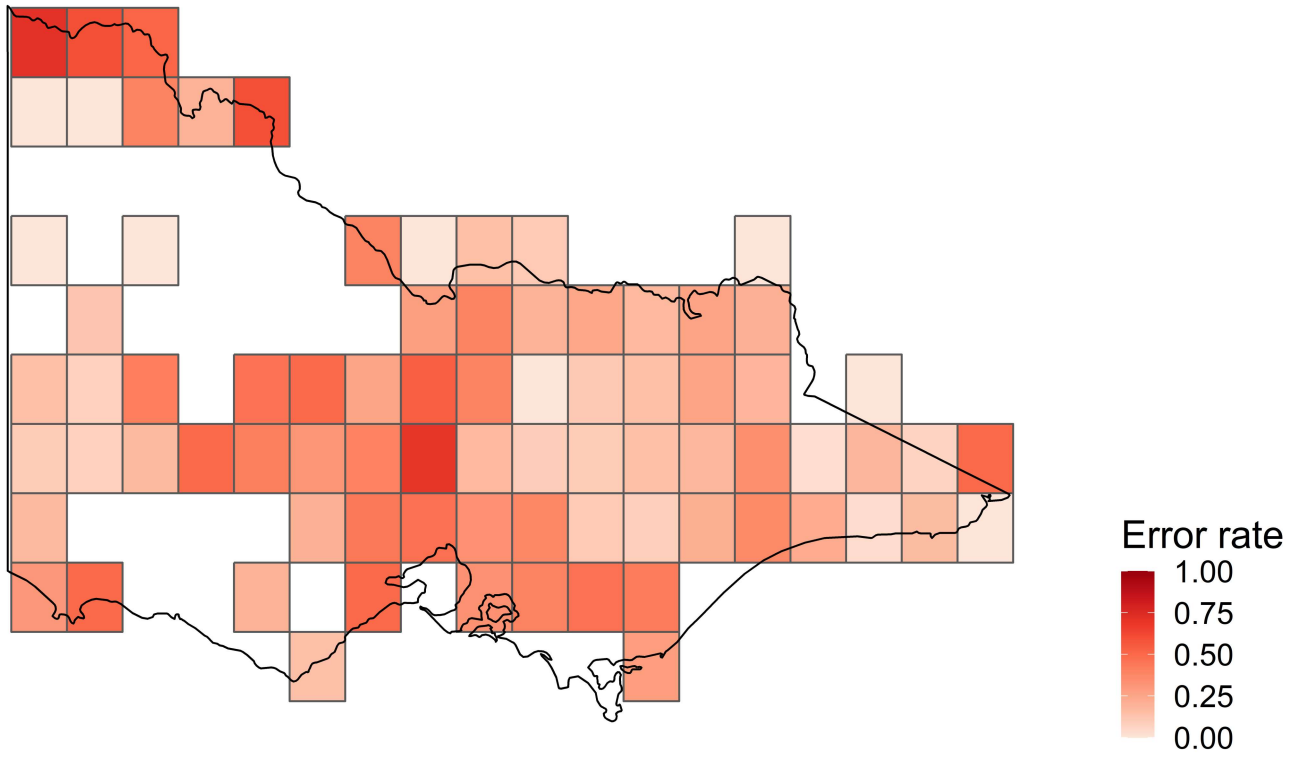
The overall accuracy of our model was 74.95%.

- Lightning-caused and accident-caused ignitions could be easily classified from other causes.
- The model was not very confident with arson and burning off.

	Lightning	Accident	Arson	Burning_off	Total
Prediction:Lightning	703 (90.5%)	77 (12.1%)	50 (15.4%)	44 (32.4%)	874
Prediction:Accident	51 (6.6%)	494 (77.9%)	89 (27.4%)	38 (27.9%)	672
Prediction:Arson	18 (2.3%)	55 (8.7%)	175 (53.8%)	22 (16.2%)	270
Prediction:Buring_off	5 (0.6%)	8 (1.3%)	11 (3.4%)	32 (23.5%)	56
Total	777	634	325	136	1872

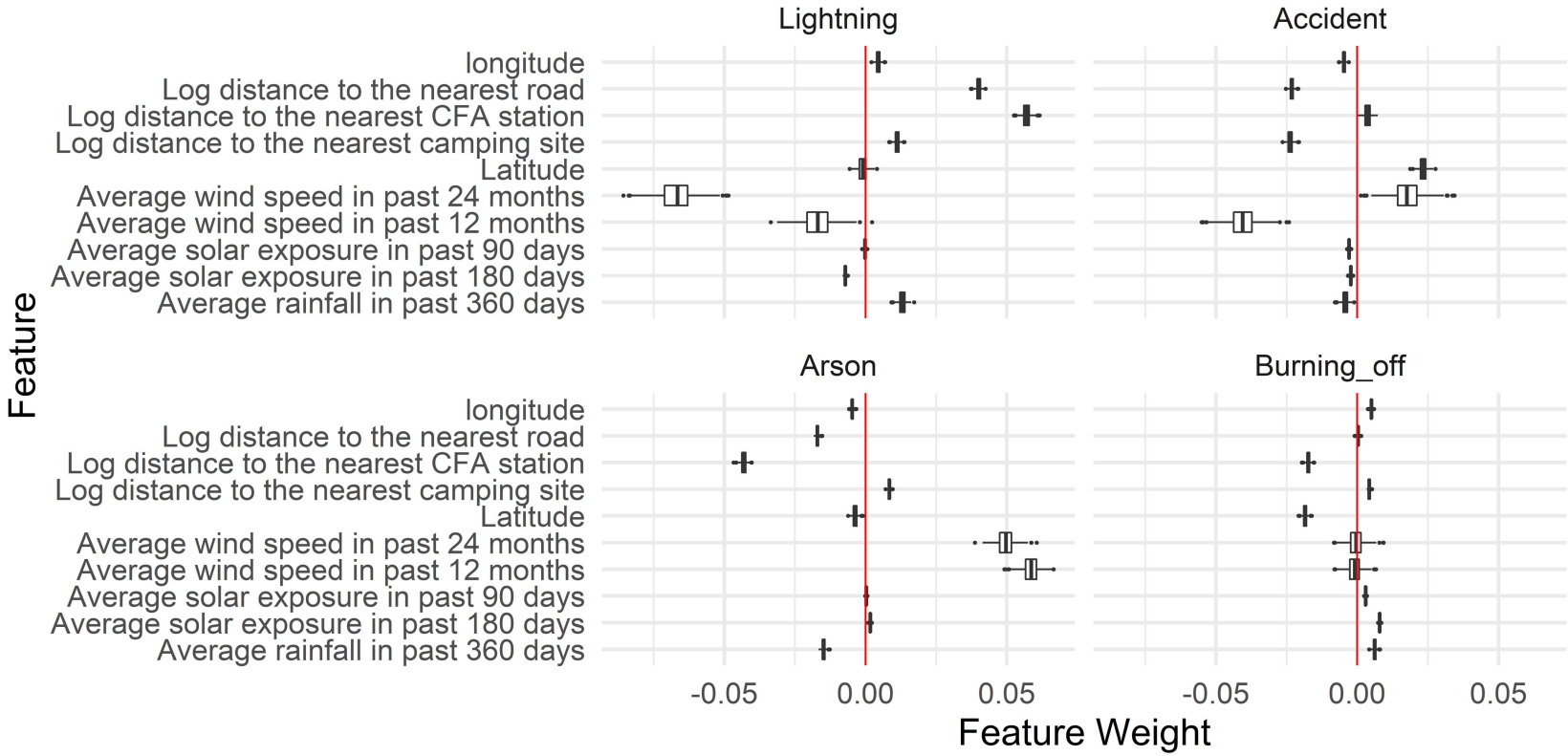
Model diagnostic

Our model correctly predicted most of the cases in the east of Victoria. However, it performed worse near Melbourne. Besides, our model didn't fit well in the north-west of Victoria.



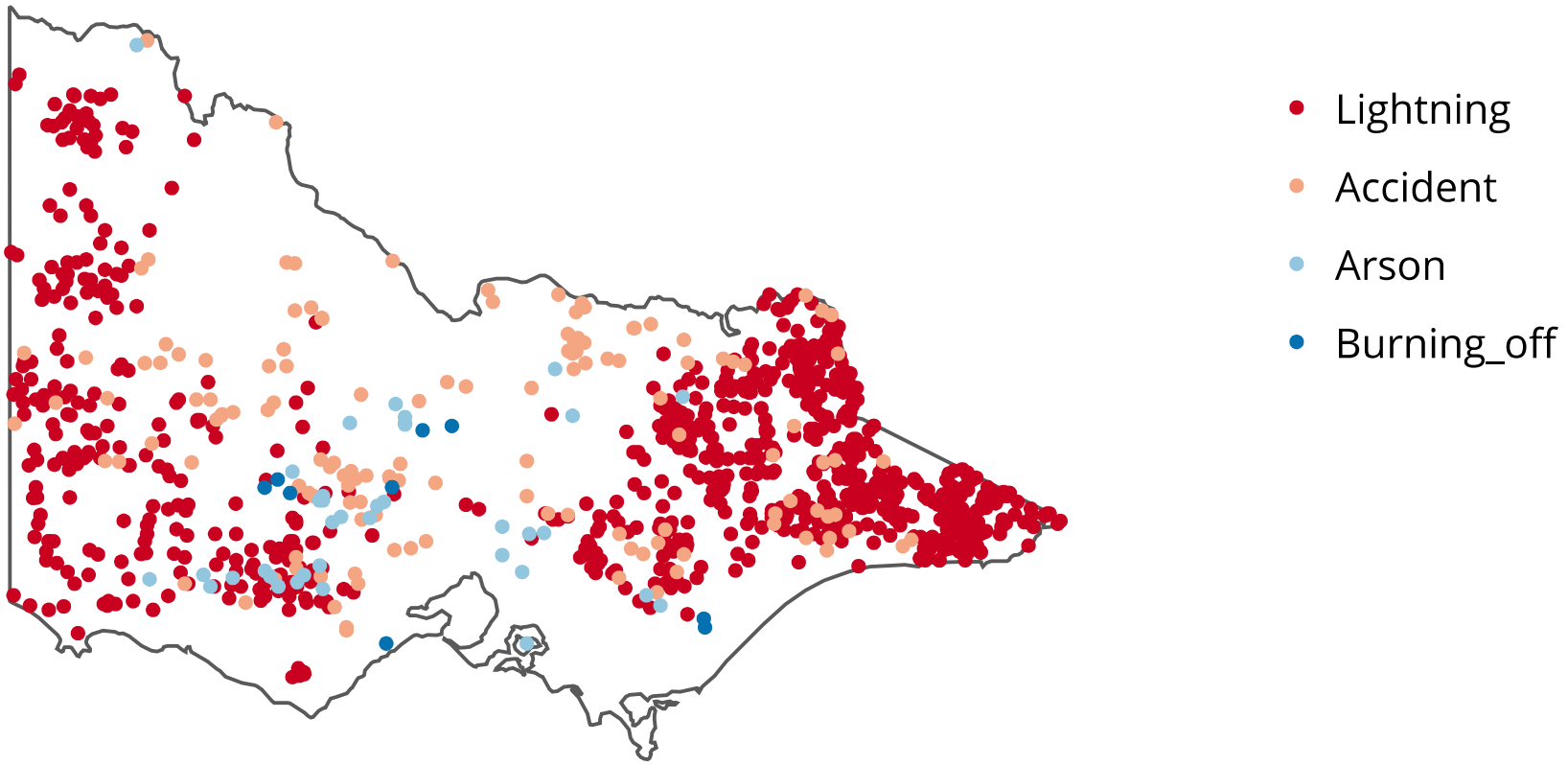
Model interpretation

The feature weights can be seen as the marginal effects. They were the scaled coefficients of the lasso regression produced by **`lime`**. Proximity to the nearest road, proximity to the nearest road and average wind speed had a significant impact on the prediction.



Results

Prediction of the causes of 2019-2020 Australia bushfires



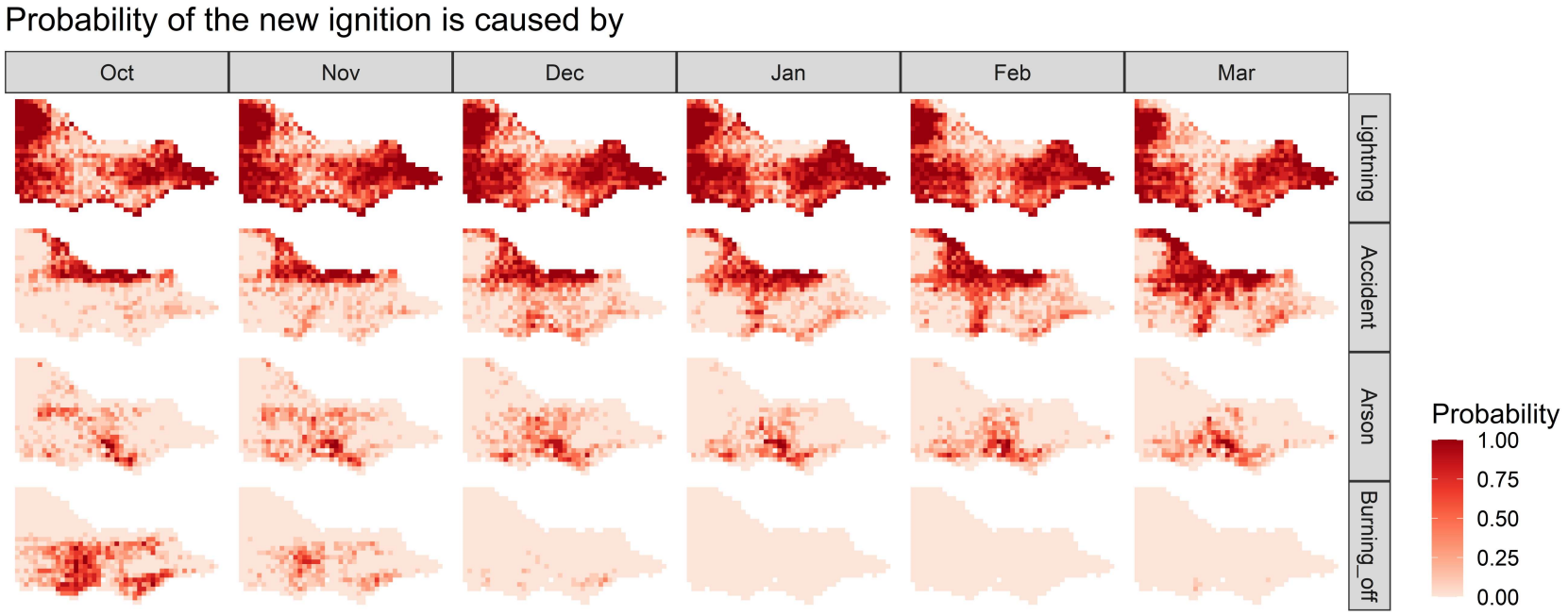
Summary of the predictions

- Most majority of the bushfires in 2019-2020 season were caused by **lightning**.
- 138 bushfires caused by accidents which took up 14% of the total fires. Most of them were ignited in March.
- 37 bushfires were caused by arsonists, and over half of them were in March.
- Very few planned burns were predicted after October 2019 which suggests the correctness of our model.

Cause	Oct	Nov	Dec	Jan	Feb	Mar	Total
Lightning	19	57	315	266	32	149	838 (82%)
Accident	3	8	34	13	0	80	138 (13.5%)
Arson	2	2	10	2	0	21	37 (3.62%)
Burning_off	7	0	2	0	0	0	9 (0.88%)

Spatial characteristics of different causes

This map is based on the assumption that the long-run weather condition of the new ignition is similar to 2019-2020 Australia bushfire season.



Policy implication

- CFA and FFMV may need to consider delivering planned burns on a larger scale in remote areas of Victoria to reduce the risk of lightning-ignited bushfire.
- CFA may need to reveal the motivations behind the rose of the number of accident-caused bushfire in March 2020 to reduce the controllable impact on bushfire ignition.
- Our research shows arson was unlikely to be the main cause. Public needs to be cautious with controversial claims that have little evidence support. Don't be a part of the disinformation campaign.

Summary

1. Algorithm to detect bushfire ignition from hotspot data
2. Model to predict the cause of bushfire ignition
3. Prediction of the causes of the 2019-2020 Australia bushfires
4. A complete and adaptable workflow for monitoring and understanding new ignition from hotspot data

Risk of bushfire ignited by different methods

Using the lightning-ignited bushfire as an example, our model only produced

$$P(L|S, \mathcal{F})$$

where L = the bushfire ignited by lightning, S = the bushfire ignited and $\mathcal{F}$ is other information of the ignition, such as location, time, weather conditions, etc.

A future extension would be predicting $P(S|\mathcal{F})$ and combining this model to produce

$$P(L, S|\mathcal{F}) = P(L|S, \mathcal{F})P(S|\mathcal{F})$$

Thanks for listening!

I would like to thank my supervisors,

Dianne Cook and **Emily Dodwell**,

for your help and guidance.

I would also thank **Xueyan Zhao** and **Heather Anderson** for your coordination in this year.

Acknowledgements

Slides produced using Rmarkdown with xaringan styling. Raw files for slides, research report and data analysis are available on MY GITHUB

Questions?

Author: Weihao Li

Department of Econometrics and Business Statistics

wlii0039@student.monash.edu