# A clustering algorithm to organize satellite hotspots data for the purpose of tracking bushfires remotely

*by Weihao Li, Emily Dodwell, and Dianne Cook*

**Abstract** An abstract of less than 150 words.

## Introduction

Bushfires are a major problem for Australia, and many other parts of the globe. There is concern that as the climate becomes hotter, and drier, that the impact of fires becomes much more severe and extensive. In Australia, the 2019-2020 fires were the worst on record causing extensive ecological damage, as well as damage to agricultural resources, properties and infrastructure. The Wollemi pine, rare prehistoric trees, required special forces intervention to prevent the last stands in the world, in remote wilderness areas, from being turned into ash.

Contributing to the problem is that many fires started in very remote areas, locations deep into the temperate forests ignited by lightning, that are virtually impossible to access or to monitor. Satellite data provides a possible solution to this, particularly remotely sensed hot spot data, which may be useful in detecting new ignitions and movements of fires. Understanding fires in remote areas using satellite data may provide some help in developing effective strategies for mitigating bushfire impact.

This work addresses this topic. Using hot spot data, can we cluster in space and time, in order to determine (1) points of ignition and (2) track the movement of bush fires.

This paper is organised as follows. The next section provides an introduction to the literature on spatiotemporal clustering and bush fire modeling and dynamics. Section Algorithm describes the clustering algorithm, and section Application illustrates how the resulting data can be used to study bush fire ignition.

## Background

### Spatiotemporal clustering

### Bushfire modeling

## Algorithm

### Data pre-processing

### Steps

This algorithm runs in a temporal manner. Starting from the first hour of the first day or the bushfire season, hotspots are grouped, and then agglomerated spatially. This proceeds to the next hour.

**1. Divide hotspots by hour**

The algorithm starts by dividing hotspots into subgroups given their hours since the first observed hotspot as shown in Figure 2. Notice the unit of time is an arbitrary choice. Theoretically, it can be replaced with any other units not larger than the total length of time and not less than the temporal resolution of the data. Normally, the algorithm will be sensitive to noises and unobserved hotspots if a small unit of time is used while losing details of bushfires movement if a large unit of time is used.

One hour is chosen to be the default unit of time because the temporal resolution of our data is 10 minutes. Besides, since bushfires usually last longer than 12 hours, treating hotspots within an hour as a whole is reasonable.

**2. Start from the first hour**

The algorithm first selecting the hotspots in the first hour as shown in Figure 3.

**3. Connect reachable hotspots**

We try to formalize the meaning of "cluster" to better illustrate the connection between this algorithm and the underlying bushfire behaviour. When bushfires spread over time, the satellite will

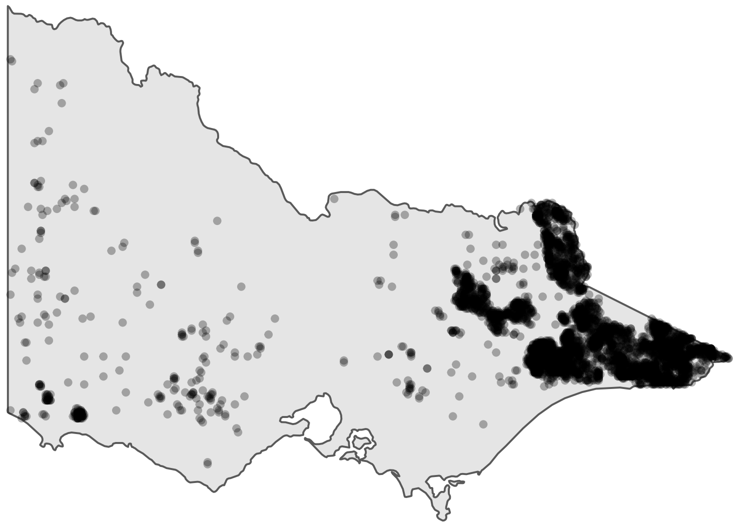


**Figure 1:** Hotspot locations in Victoria during 2019-2020 season.

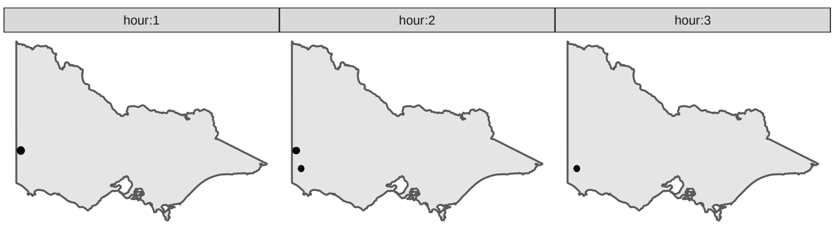



**Figure 2:** Step 1. Hotspots in the first 3 hours of the bushfire season.

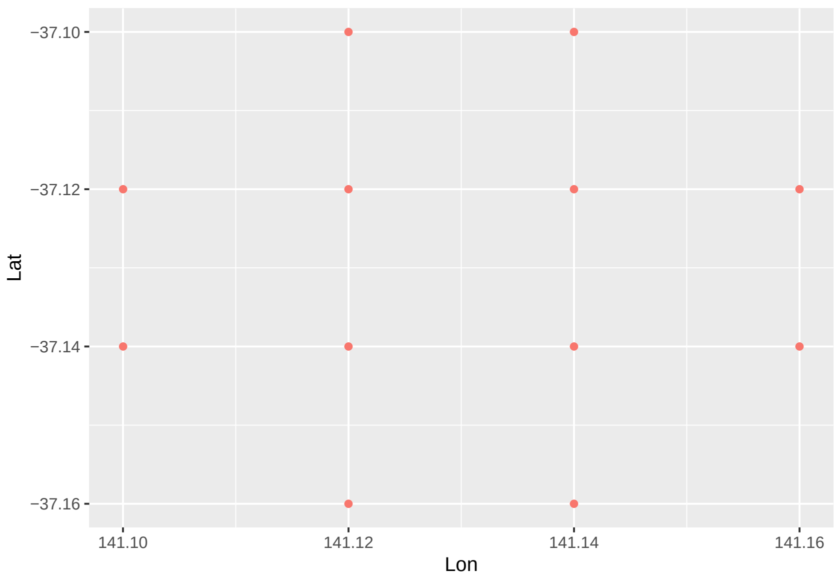



**Figure 3:** Step 2. Hotspots in the first hour.

record a series of hotspots along its trajectory. Thus, if two hotspots are close to each other within a certain distance in a short period of time, they could be considered as a single bushfire.

**Definition 1 (directly reachable)** *A point p is directly reachable from a point q with respect to AdjDist, if the distance between point p and q is less or equal to AdjDist.*

**Definition 2 (reachable)** *A point p is reachable from a point q with respect to AdjDist, if there is a chain of points* $p_1$, $p_2$, ..., $p_n$, $p_1 = p$, $P_n = q$ *such that* $P_n$ *is directly reachable from* $p_{n-1}$.

In this step, the algorithm connects all pairs of directly reachable hotspots with respect to $AdjDist = 3000m$ in the first hour as shown in Figure 4. The result of this step is an undirected and unweighted graph. $AdjDist$ is the first parameter used in this algorithm which controls the connectivity between hotspots. It is introduced for the purpose of parameterizing the speed of bushfires.

**4. Assign memberships to hotspots**

The algorithm then clusters hotspots into groups based on the connectivity between hotspots as shown in Figure 5. Different memberships will be assigned to vertices in different components. Thus, if two hotspots are reachable from each other, they will be assigned the same membership. In the first hour, because all hotspots are reachable from other hotspots, they are assigned the same membership.

**5. Move to the second hour or beyond**

We need to introduce the second parameter used in this algorithm, which is ActiveTime. This parameter controls the time frame of hotspots that will be included by the algorithm in each iteration. More specifically, this time frame is defined as $[max(t - ActiveTime, 1), t]$, where $t$ is the current timestamp of the algorithm. For instance, if $ActiveTime = 24$ and the algorithm is clustering the hotspots in the 36th hour, then the time frame will be $[12, 36]$.

We define this time frame for the purpose of associating the hotspots in the current timestamp with hotspots observed in previous hours. By imposing this time frame, we provide an invariant task at each iteration, which is clustering hotspots in the current hour given the clustered hotspots in previous hours.

Move to the hour $t$, the algorithm will select hotspots in hour between $max(t - ActiveTime, 1)$ and $t$, including $max(t - ActiveTime, 1)$ and $t$. Given $t = 2$ and $ActiveTime = 24$, the hotspots in the first two hours will be selected as shown in Figure 6.

**6. Connect reachable hotspots**

This step is equivalent to step 3 and is only used for illustration purpose. Given the hotspots in the first two hours, the algorithm connects all pairs of reachable points as shown in Figure 7.

**7. Find the nearest hotspot in the previous hours within the same component for each hotspots in the current timestamp**

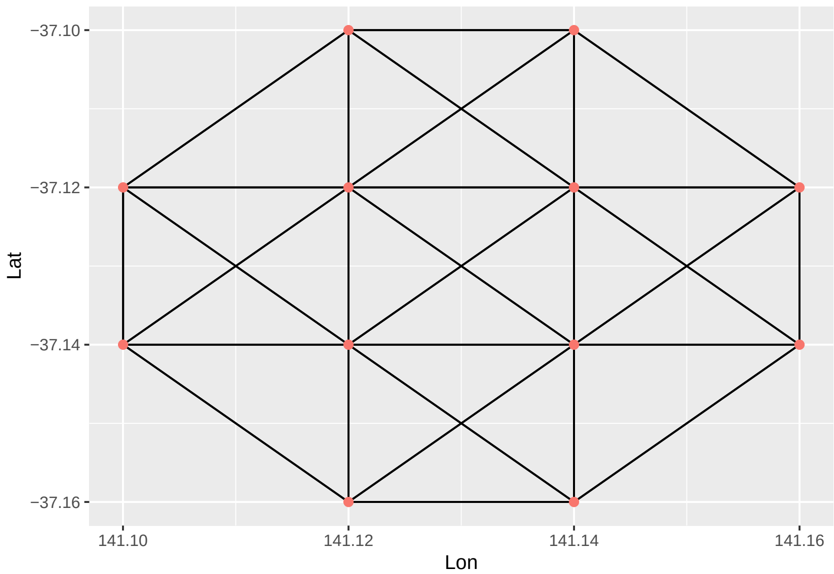



**Figure 4:** Step 3. All pairs of directly reachable hotspots in the first hour are connected given the AdjDist is 3000 metres. There is only one component in the graph

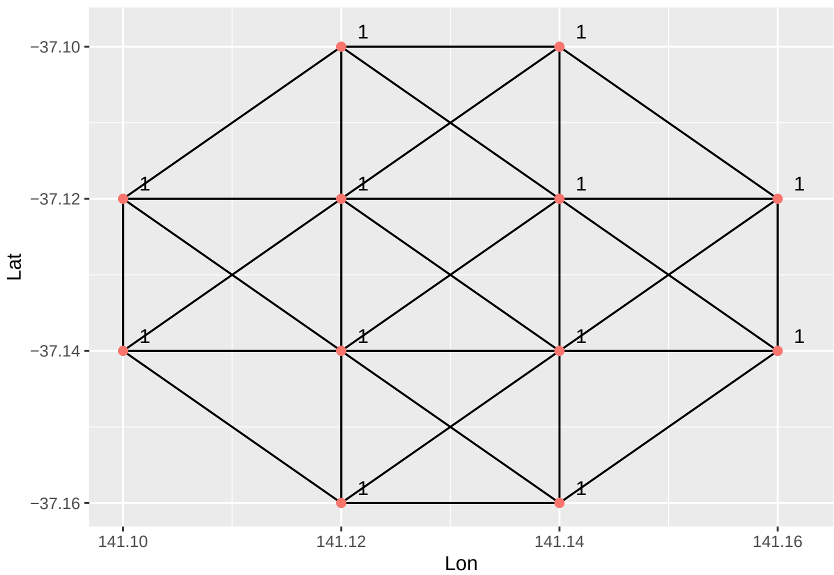



**Figure 5:** Step 4. All hotspots in the first hour are reachable from each other. They are assigned the same membership.

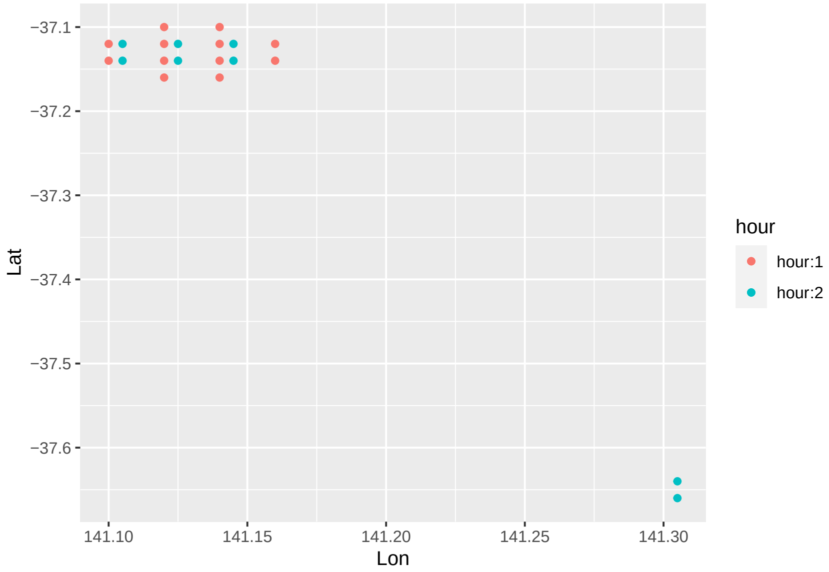



**Figure 6:** Step 5. Hotspots in the first two hours. Due to overlapping, hotspots in the second hour are adjusted by 0.005 degree in longitude.

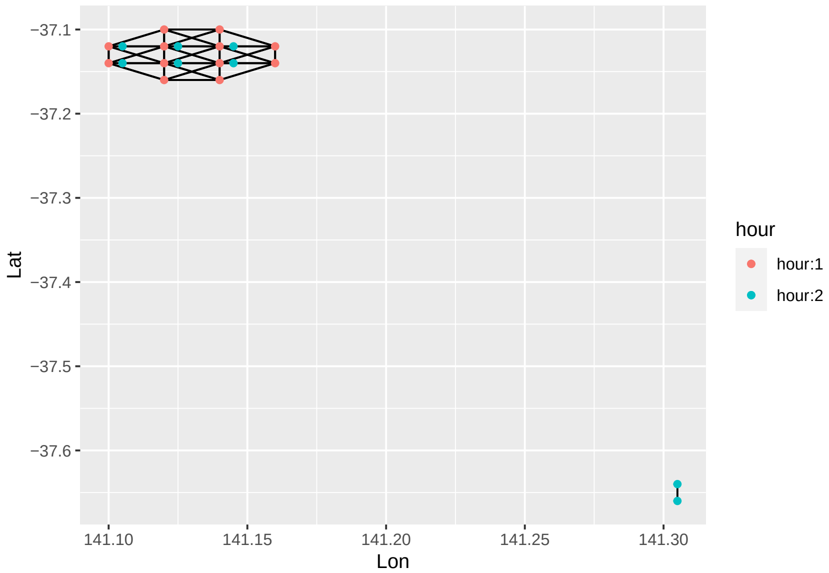



**Figure 7:** Step 6. All pairs of directly reachable hotspots in the first two hours are connected given the AdjDist is 3000 metres. There are two components in the graph.

According to the result from step 6, the algorithm will then find the nearest hotspot in the previous hours within the same component for each hotspots in the current timestamp which is the second hour as shown in Figure 8. This step along with step 8 is to solve the problem when a hotspot in the current timestamp share the same component with multiple hotspots in the previous hours. And, we need to decide which membership in the previous hours it should inherit.

**8: Assign membership to hotspots in the current timestamp**

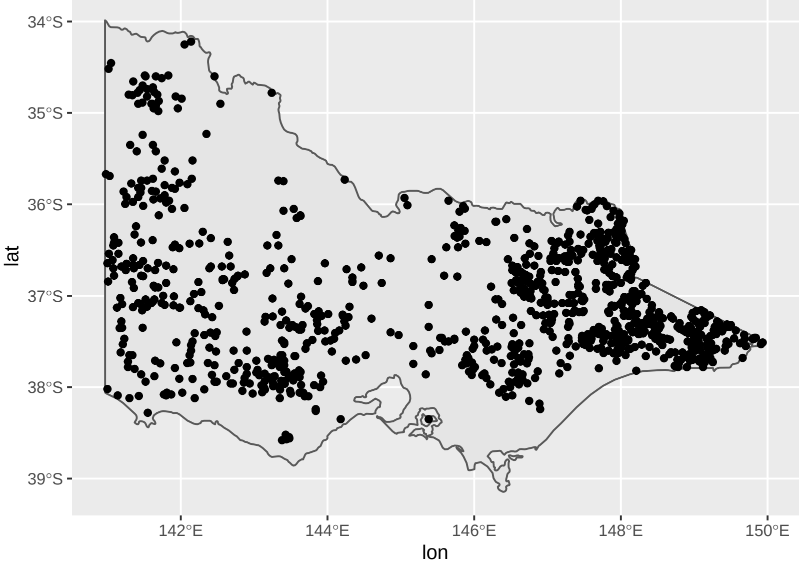
### Effects of parameter choices

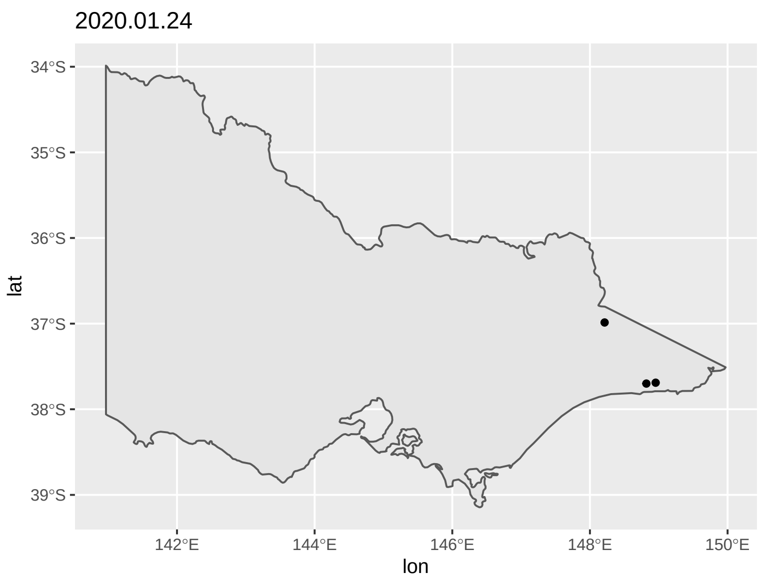
There are two parameters that can be tuned in this algorithm. They are `adj_dist`, which is the density distance and `active_time`, which is the .

## Application

### Determining the ignition point and time for individual fires

Show ignition points for a particularly heavy day and another for a particularly light day



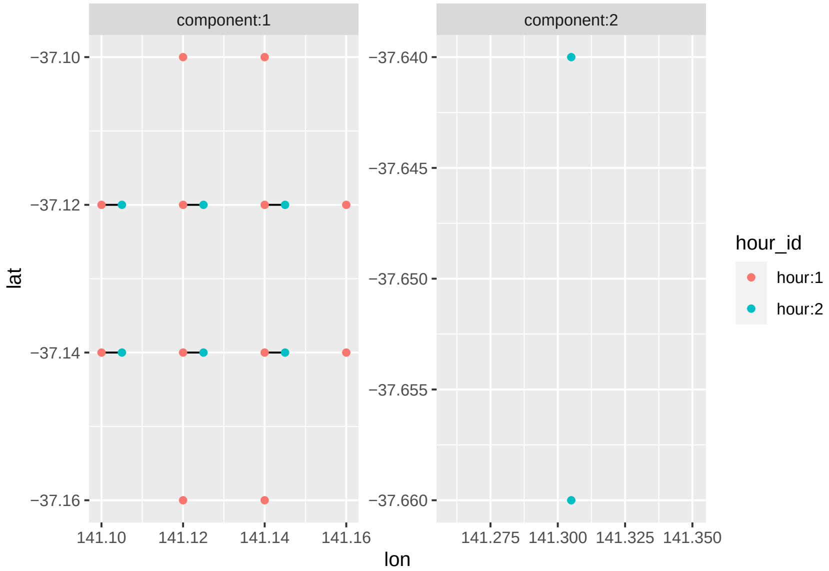
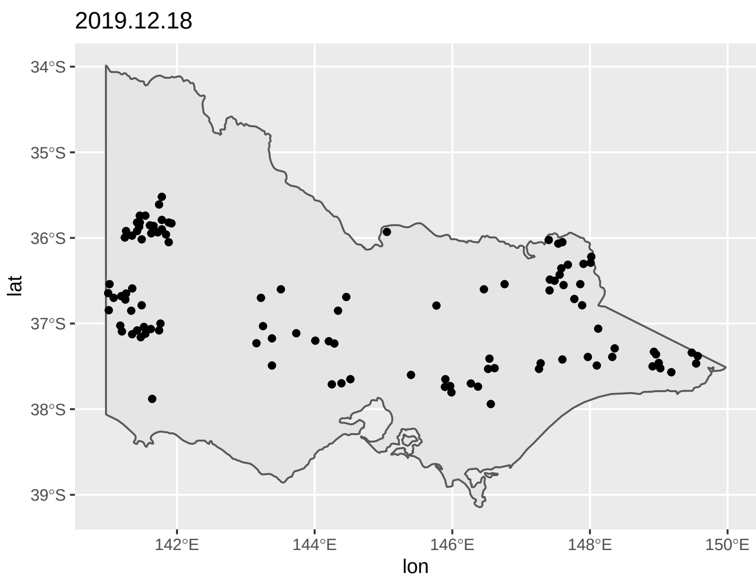



**Figure 8:** Step 7. Hotspots in the second hour is connected with its nearest hotspot in the first hour within the same component. Due to overlapping, hotspots in the second hour are adjusted by 0.005 degree in longitude.




### Tracking fire movement

Display showing how a fire moves over time, maybe two or more fires

### Allocating resources for future fire prevention

Merging data with camp sites, CFA, roads, . . .

### Summary

### Acknowledgements

- The code and files to reproduce this work are at XXX
- Data on hotspots can be downloaded from XXX

*Weihao Li*
*Monash University*
*line 1*
*line 2*

wlii0039@student.monash.edu

*Emily Dodwell*
*AT&T*
*line 1*
*line 2*

emily@research.att.com

*Dianne Cook*
*Monash University*
*line 1*
*line 2*

dicook@monash.edu