

# Development of spatial models for bushfire occurrence in South-Eastern Australia

**Y. Zhang**<sup>a</sup>, **S. Lim**<sup>a</sup> and **J.J. Sharples**<sup>b</sup>

<sup>a</sup> *School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia*

<sup>b</sup> *School of Physical, Environmental and Mathematical Sciences, The University of New South Wales, Canberra, ACT 2610, Australia*

*Email: [yang.zhang3@student.unsw.edu.au](mailto:yang.zhang3@student.unsw.edu.au)*

**Abstract:** Australia is one of the most flammable continents in the world. The southeastern region of the continent, where the population is densely settled, is susceptible to low frequency and high intensity bushfires that can threaten human lives and cause extensive loss of properties. This paper describes the development and validation of spatial models for bushfire occurrence in South-Eastern Australia, especially in New South Wales, Victoria and the Australian Capital Territory.

The active fire database from the Moderate Resolution Imaging Spectroradiometer (MODIS) is used as original data source of fire activity over the 11-year period 2003-2013. Those active fire detections are grouped into burning events using the Fire Spread Reconstruction approach (FSR) algorithm based on the spatial and temporal proximity between fire detections. Fire occurrence point is defined as the ignition point of each identified event. Univariate and multiple logistic regression models are investigated for the comprehensive understanding of determinants contributing to the spatial distribution of bushfires. The probability of bushfire occurrence in South-Eastern Australia is also studied for the prediction of future fire occurrence.

Bushfires in the study area are significantly influenced by both environmental and anthropogenic variables. The mean annual precipitation positively influences the fire incidence, because the semi-arid regions lack the fuels necessary for a fire to start, while the coastal regions with abundant rain provide ample fuels for fire ignition. This finding is inconsistent with that at a small landscape scale. Fire probabilities are different regarding various land cover types. Forests are most likely to burn because they are covered by heavy fuel loads. Savannas are equivalently fire-prone because they are fundamentally easy to ignite. Permanent wetlands are also susceptible to fire possibly due to the influence of climate change and urban expansion. Shrublands are less fire-prone because of the low-level shrub canopy cover. Fires are also found to distribute in areas near the zero meso-scale elevation residual contour, which is consistent with the previous finding. Anthropogenic variables also show predictive power because of the influence of human activities on fire occurrence.

The final model for the probability of bushfire occurrence include mean annual precipitation, MODIS land cover, distance to zero meso-scale elevation residual contour, distance to secondary road and distance to railway. The bushfire probability map was generated accordingly. From the information provided by the quantitative statistics and the bushfire probability map, bushfires in the study area mostly likely to occur in coastal and mountainous areas close to various types of infrastructure and zero meso-scale elevation residual contours, as well as on forests, savannas and permanent wetlands, while they rarely occurred inland. It is concluded that the proposed model provides practical guidance for fire management actions in South-Eastern Australia.

**Keywords:** *Bushfire occurrence, spatial pattern, MODIS*

## 1. INTRODUCTION

Australia is known as one of the most fire-prone lands in the world. Quick responses to fire activities are required in the southeastern region of the continent because bushfires in this densely settled area can cause massive loss of lives and properties (Russell-Smith *et al.*, 2007). Accurate predictions of when and where bushfires will occur are crucial in this regard. Statistics on fire occurrence can be used to model the patterns of fire activities and identify fire-prone areas. The results of these models can support planning and decision-making in fire management activities.

Environmental and anthropogenic factors regulate the spatial patterns of bushfires across multiple scales. At landscape or sub-regional scales, the traditional fire environment triangle illustrates the influence of environmental variables (weather, fuels, and topography) on fire behaviour. At broader scales, ignitions, climate and vegetation are three factors that are considered to be significant in regulating the fire regime. Fire climate is found to affect bushfire occurrence (e.g. Bradstock *et al.*, 2009; Preisler *et al.*, 2004). Land cover types affect bushfire incidence due to their connection with fuel load and type (Gumming, 2001). Fuel moisture content and air temperature (Sharples, 2009) are influenced by topography, which indirectly affects the chance of fire ignition. Anthropogenic variables are also powerful indicators of fire incidence in many regions around the world (e.g. Vilar *et al.*, 2010).

Fire occurrence studies have been conducted in North America (e.g. Preisler *et al.*, 2004), Europe (e.g. Vilar *et al.*, 2010) and Australia (e.g. Bradstock *et al.*, 2009), however, none of them were undertaken at a sub-continental scale in south-eastern Australia. High accuracy satellite-based monitoring systems such as the Moderate Resolution Imaging Spectroradiometer (MODIS) make it possible to understand fire occurrence at a broad scale. Specifically, the active fire products (Giglio *et al.*, 2003) derived from MODIS datasets have been used to study spatial and temporal patterns of bushfire occurrence. In some countries, the Fire Spread Reconstruction (FSR) algorithm is used to recreate fire development and identify ignition points in fire events (Loboda & Csiszar, 2007). However, the relationship between MODIS-based fire locations and their determinants in Australia is not well understood.

In this study, fire occurrence points are identified from MODIS active fire locations during the time period from 2003-2013 using the FSR algorithm. Binary logistic regression models are used to explore the relationships between these fire occurrence points and other explanatory variables, and to generate a fire occurrence probability map accordingly. This study aims to extend the knowledge about the distribution of bushfires and provide practical guidance for fire management actions at a broad landscape scale in South-Eastern Australia.

## 2. METHODOLOGY

### 2.1. Study Area

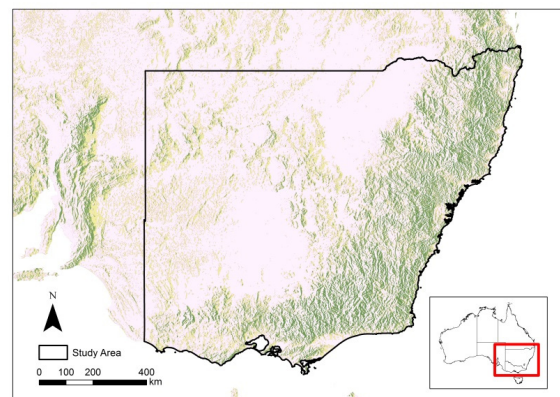
The study area covers the territory of New South Wales, Victoria and Australian Capital Territory, with an area of 1,050,000 km<sup>2</sup> (Figure 1). Infrequent high-intensity summer–autumn fires occur in this area due to the winter–spring rainfall activities (Murphy *et al.*, 2013).

### 2.2. Data Description

Fire occurrence, climate, vegetation, topography and anthropogenic data are collected and prepared for statistical analysis as listed in Table 1.

#### *Climate*

Climate data used in this study includes the mean annual precipitation, the mean January minimum temperature and the mean July maximum temperature. The January minimum temperature and the July maximum temperature are chosen because they maximise the spatial variability of temperature gradients (Syphard *et al.*, 2008). The precipitation data for 2004-2013 and the temperature data for 2006-2013 are obtained from the Bureau of Meteorology. The mean January minimum temperature and the mean July maximum temperature are calculated by averaging yearly extreme values picked from the daily mean maximum and the daily mean minimum temperature raster layers. The mean annual precipitation is the mean



**Figure 1.** Location of the study area

of the monthly total rainfall for the year. All the generated raster layers are resampled at a 1 km resolution using nearest neighbour technique.

**Table 1.** Variables analysed in the regression models explaining fire occurrence in South-Eastern Australia

| Variable   | Source                    | Description  |
|--|---------------------------|--|
| Climate  |                           |  |
| Mean annual precipitation                              | Bureau of Metrology       |  |
| Mean January minimum temperature                       | Bureau of Metrology       |  |
| Mean July maximum temperature                          | Bureau of Metrology       |  |
| Vegetation   |                           |  |
| CLUM land use  | Department of Agriculture | 6 primary classes  |
| MODIS land cover                                       | NASA                      | 6 primary classes  |
| NDVI   | NASA                      |  |
| EVI  | NASA                      |  |
| Topography   |                           |  |
| Elevation  | NASA                      |  |
| Slope  | NASA                      | Derived from elevation grid  |
| Northwestness  | NASA                      | Derived from elevation grid<br>$NW = \cos ([\text{aspect}] - 1, 0, \cos (([\text{aspect}] + 45) * \pi / 180))$ |
| Distance to zero meso-scale elevation residual contour | NASA                      | Derived from elevation grid  |
| Anthropogenic Data                                     |                           |  |
| Distance to primary road                               | OSM                       | Mean Euclidean distance  |
| Distance to secondary road                             | OSM                       | Mean Euclidean distance  |
| Distance to railway                                    | OSM                       | Mean Euclidean distance  |
| Population density                                     | ABS                       |  |
| Fire occurrence  | NASA                      | Identified from MODIS active fire detections   |

### Vegetation

The catchment scale land use of Australia map (CLUM) is used in this study. It was updated in March 2014 and is published by the Department of Agriculture. According to the Australian Land Use and Management (ALUM) Classification version 7, the CLUM data is categorised into 6 primary classes: (1) conservation, natural environments; (2) production from relatively natural environments; (3) production from dryland agriculture and plantations; (4) production from irrigated agriculture and plantations; (5) intensive uses; (6) water. The 50 m resolution dataset is resampled at a 1 km resolution using the majority algorithm.

The MODIS 500 m Land Cover Type product (MCD12Q1) is also used in this study. The original 17 classes are grouped into 6 primary classes to take the influence of primary vegetation types on fire occurrence into account. The dataset of 2003 is chosen and resampled at a 1 km resolution using the majority algorithm. The Collection 5 MODIS global monthly Vegetation Index product (MYD13A3) with a spatial resolution of 1 km is utilised. Both Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) in January 2003 are generated.

### Topography

Elevation, slope, aspect and *distance to zero meso-scale elevation residual contour* (McRae, 1992) are used in this study. The Global Digital Elevation Model Version 2 (GDEM V2) 30 m data obtained from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is used as the elevation data. It is resampled at a 1 km resolution by calculating the mean of the cell values within each 1 km<sup>2</sup> rectangular block. Slope and aspect variables are derived from the elevation data. To incorporate the variable aspect into linear statistics, a linear variable ‘northwestness’ is derived by the cosine transformation of the circular aspect variable. The variable *distance to zero meso-scale elevation residual contour* is generated by removing micro- and macro- scale variation of elevation, leaving only the meso-scale residual and generating contours accordingly (McRae, 1992).

### Anthropogenic Data

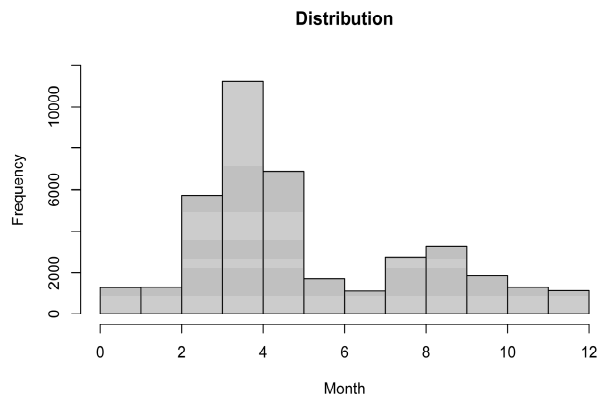
The influence of the human activity on bushfire occurrence is quantified based on the network of roads and railways that make up the traffic system. The primary road, secondary road and railway information is derived from the database of OpenStreetMap (OSM). The Euclidean distance to the nearest road or railway is calculated to generate 1 km resolution distance maps. The population density information in 2003 is also

incorporated. The 1 km resolution map is generated according to the estimated residential population of each Local Government Area (LGA) obtained from Australian Bureau of Statistics (ABS).

*Fire Occurrence*

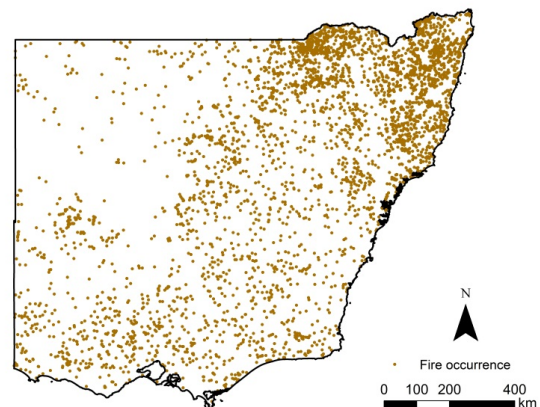
The Collection 5 MODIS global monthly fire location product (MCD14ML) (Giglio *et al.*, 2003) is used in this study. The MCD14ML dataset is a combination of level 2 MOD14 / MYD14 active fire products collected from the MODIS instrument on Terra and Aqua. The spatial resolution of this product is 1 km.

Fire occurrence points are identified from the MODIS active fire detections using the Fire Spread Reconstruction approach (FSR) (Loboda & Csiszar, 2007). This method groups active fire points into burning events based on the spatial and temporal proximity between fire detections. The upper limits of spatial and temporal distance between two points are set at 4 km and 10 days, respectively. The number of active fire points within an individual group ranges from 1 to 13337 in this study. The earliest points within each group are selected to be the candidate ignition points. A given group may contain multiple candidate ignition points, but only the point with the smallest object ID is assumed to be the ignition point to minimize the influence of spatial and temporal autocorrelation. The identified 19668 ignition locations are used as fire occurrence points in the modelling process.



**Figure 2.** Monthly distribution of fire ignitions in South-Eastern Australia from 2003 to 2013

CLUM Land use and MODIS land cover data are used to filter the fire occurrence points. Fire incidences that are located on intensive used areas (e.g. industrial and residential areas), waterbodies and barrens are removed to reduce the influence of commission error. The histogram of the monthly distribution of the remaining fire occurrence points (Figure 2) exhibits upper fluctuations in spring and autumn, which are associated with prescribed burning and agricultural fires (e.g. stubble burning), therefore only fire occurrence points within the typical fire danger season (November to February) are used in this study.



**Figure 3.** Fires occurrence points during 11 fire seasons of 2003-2013 in South-Eastern Australia

In total there were 4110 fire occurrences during the fire seasons of 2003-2013 in the study area, which are utilised in the modelling process. The locations of these events are shown in Figure 3.

**2.3. Modelling approach**

To estimate the probability of fire occurrence, a logistic regression model was developed. Let  $P_i$  be the probability of a fire occurrence in a cell  $i$ , and  $x_{ij}$  be the value of the  $j$ th covariate in the cell  $i$ , then the logistic regression model is defined as:

$$P_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni}) / (1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni}))$$

where  $\beta_0$  is an intercept and  $\beta_1, \beta_2, \dots, \beta_n$  are coefficients (model parameters). To reduce the size of the dataset being analysed, we randomly selected 16,315 control points from the non-fire locations. The ratio of ones (ignitions) to zeros (sampled non-ignitions) is 1:3.97. The sampling rate,  $\pi$ , is approximately 0.016. The

probability of fire occurrence was calculated by incorporating an offset term,  $-\log(\pi)$ , following the method demonstrated by Preisler *et al.* (2004).

Spearman's rank correlation was used to compare the correlations among continuous explanatory variables to avoid multicollinearity. NDVI, EVI, January maximum temperature, July minimum temperature, elevation, slope and population density were removed because their correlations with other explanatory variables are above 0.6 (Wintle *et al.*, 2005). The association between categorical variables (CLUM land use and MODIS land cover) was calculated using Pearson's Chi-Squared test. CLUM land use was removed because high association ( $P \leq 0.001$ ) was found. To evaluate the influence of each explanatory variable on the probability of fire incidence, we developed univariate logistic regression models. The performance of square or log transformed continuous variables were also tested. Dummy variables were created for the MODIS land cover, with a reference class (intercept) of the forests; the coefficients of other classes are relative to the value of the intercept. We also calculated the 95% confidence interval for each variable using profiled log-likelihood function. Our final model was chosen using a backwards stepwise algorithm based on the Akaike Information Criterion (AIC) (Venables & Ripley, 1999). The receiver operating characteristics (ROC) curves of all the univariate models and the final model were computed to evaluate their performance in fire occurrence prediction. The area under the curve (AUC) of the ROC curve was used to measure the model fit (0.9-1: excellent, 0.8-0.9: good, 0.7-0.8: fair, 0.6-0.7: poor, 0.5-0.6: fail) (Swets, 1988). Significance of each explanatory variable was determined using the likelihood ratio test.

All the statistical analyses were computed using the R package version 3.1.1 (R Development Core Team, 2014). Logistic regressions were fitted using generalized linear models (GLMs). ROCs were calculated with the R module *pROC* (Robin *et al.*, 2011).

### 3. RESULTS

According to the results of univariate logistic regression analysis, all of the continuous explanatory variables except 'northwestness' are found to be significant ( $P \leq 0.05$ ) (Table 2). Fire occurrence is positively related to the mean annual precipitation. The distance to zero meso-scale elevation residual contour positively affects bushfire ignition. All the anthropogenic variables representing distances to human facilities are negatively related to fire occurrence, which suggests that bushfires are more likely to occur in the areas close to human activities.

The overall  $P$ -value of MODIS land cover indicates that it is significant in predicting fire occurrence as a whole (Table 2). Although the contrasts of savannas and permanent wetlands with the reference class (forests) are not significant, the smallest lower 95% confidence limit value of the two classes is -0.07, less than zero and greater than the upper limits of all the remaining classes. Therefore, forests, savannas and permanent wetlands are equally the most fire-prone. The sample size within the class of permanent wetland is 7, which leads to a wide confidence interval. There is significant evidence that grasslands are moderately susceptible to fire, with shrublands and croplands the least likely to ignite.

**Table 2.** Univariate regression results for variables explaining bushfire occurrence in South-Eastern Australia

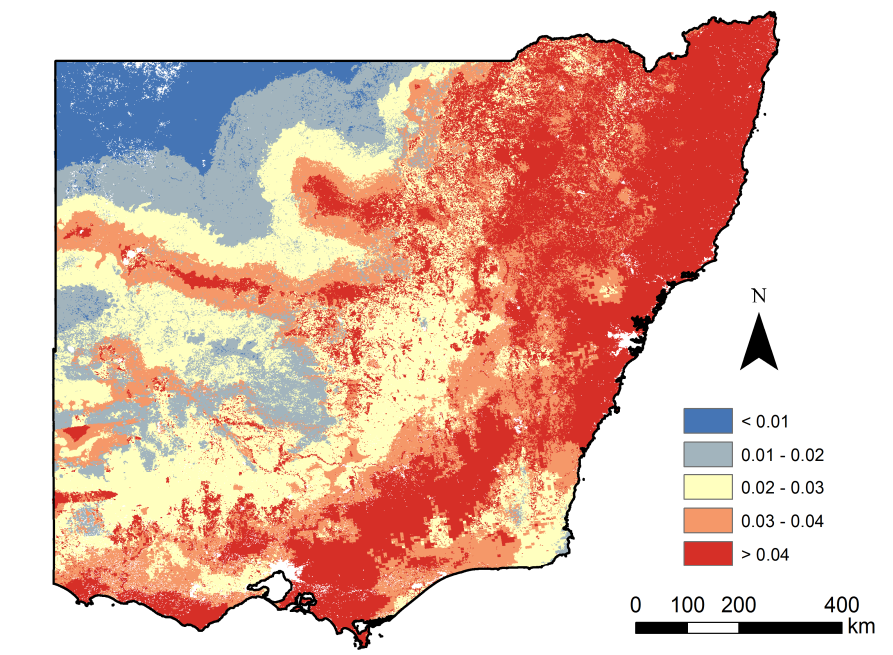
| Variable   | Coefficient | Standard Error | $P$ -value | 95% Confidence Interval |             |
|--|-------------|----------------|------------|-------------------------|-------------|
|  |             |                |            | Lower Bound             | Upper Bound |
| Climate  |             |                |            |                         |             |
| Mean annual precipitation                              | 0.001670    | 0.000050       | <0.0001    | 0.001573                | 0.001768    |
| MODIS land cover                                       |             |                |            |                         |             |
| Forests (Intercept)                                    | -0.836074   | 0.038641       | <0.0001    | -0.912198               | -0.760708   |
| Shrublands   | -0.897672   | 0.049979       | <0.0001    | -0.995594               | -0.799660   |
| Savannas   | 0.029822    | 0.055991       | 0.5943     | -0.079961               | 0.139542    |
| Grasslands   | -0.576073   | 0.073764       | <0.0001    | -0.721701               | -0.432474   |
| Permanent wetlands                                     | 1.123756    | 0.764739       | 0.1417     | -0.390200               | 2.750080    |
| Croplands  | -0.848992   | 0.054908       | <0.0001    | -0.956773               | -0.741514   |
| Topography   |             |                |            |                         |             |
| Northwestness  | -0.023449   | 0.024736       | 0.3366     | -0.071261               | 0.024403    |
| Distance to zero meso-scale elevation residual contour | -0.000022   | 0.000007       | 0.0010     | -0.000035               | -0.000009   |
| Anthropogenic variables                                |             |                |            |                         |             |
| Distance to primary road                               | -0.000008   | 0.000001       | <0.0001    | -0.000010               | -0.000007   |
| Distance to secondary road                             | -0.000014   | 0.000001       | <0.0001    | -0.000016               | -0.000011   |
| Distance to railway                                    | -0.000008   | 0.000000       | <0.0001    | -0.000009               | -0.000007   |



The equation that has the smallest AIC value is selected to be the final model. It includes five explanatory variables: mean annual precipitation, MODIS land cover, distance to zero meso-scale elevation residual contour, distance to secondary road and distance to railway (Table 3). The AUC of the final model is 0.71. Although the prediction power of our model is not perfect, we still obtain useful information from the modelling result. A fire occurrence probability map is generated by applying the coefficients of the final model to raster layers corresponding to the explanatory variables (Figure 4).

**Table 3.** Variables retained in the final model explaining bushfire occurrence in South-Eastern Australia

| Variable                                       | Degree of Freedom | P-value |
|--|-------------------|---------|
| Mean annual precipitation                      | 1                 | <0.0001 |
| MODIS land cover                               | 5                 | <0.0001 |
| Distance to zero meso-scale elevation residual | 1                 | <0.0001 |
| Distance to secondary road                     | 1                 | <0.0001 |
| Distance to railway                            | 1                 | 0.0003  |



**Figure 4.** Predicted probability of bushfire occurrence in South-Eastern Australia

#### 4. DISCUSSION

As is anticipated, bushfires in the study area are significantly influenced by both environmental and anthropogenic variables. The mean annual precipitation positively influences the fire incidence in our model. Although at a small scale, rain increases the moisture contents of both dead and live fuels, reducing the dry fuels available to burn, and consequently affecting fire danger; it is not the case at a broad scale. In the study area, the semi-arid regions in Australia’s continental interior lack the fuels necessary for a fire to start while the coastal regions with abundant rain provide ample fuels for fire ignition.

Fire probabilities are different regarding various land cover types. Forests are most likely to burn because they are covered by heavy fuel loads. Savannas are equivalently fire-prone because they are fundamentally easy to ignite (Murphy *et al.*, 2013). It is out of expectation that permanent wetlands are also most susceptible to fire. It may be due to the impact of climate change and urban expansion on the coastal fringe where wetlands are located (Schneider & Sutherland, undated). Shrublands are less fire-prone possibly because of the low-level shrub canopy cover (<60%) in this area.

Fires tend to distribute in areas near the zero meso-scale elevation residual contour. This finding is consistent with McRae (1992)’s finding about natural ignitions in the ACT area. This unobvious pattern is able to provide practical information to fire risk mapping. Although human-caused fires are not analysed specifically in this study, fires are still found in areas near human infrastructures because fire ignitions are strongly affected by human activities (Vilar *et al.*, 2010).

## 5. CONCLUSION

In this paper, we incorporated climate, land cover, topographic and anthropogenic information into logistic regressions to study the distribution of fire occurrence in South-Eastern Australia over the period 2003-2013. From the information provided by the modelling results and the fire probability map, bushfires in the study area mostly occurred in coastal and mountainous areas close to human infrastructures and zero meso-scale elevation residual contours, as well as on permanent wetlands, savannas and forests, while they rarely occurred inland. The results extend knowledge about the influence of environmental and anthropogenic conditions on bushfire incidence and the spatial pattern of bushfire in the NSW, VIC and ACT areas, which can help fire agencies in these three jurisdictions better target their management activities.

## ACKNOWLEDGMENTS

J.J. Sharples is supported by the Australian Research Council (IN130100038, IN140100011).

## REFERENCES

- Bradstock, R. A., Cohn, J. S., Gill, A. M., Bedward, M., & Lucas, C. (2009). Prediction of the probability of large fires in the Sydney region of south-eastern Australia using fire weather. *International Journal of Wildland Fire*, 18(8), 932-943.
- Giglio, L., Descloitres, J., Justice, C. O., & Kaufman, Y. J. (2003). An Enhanced Contextual Fire Detection Algorithm for MODIS. *Remote Sensing of Environment*, 87(2-3), 273-282. doi: [http://dx.doi.org/10.1016/S0034-4257\(03\)00184-6](http://dx.doi.org/10.1016/S0034-4257(03)00184-6)
- Gumming, S. G. (2001). Forest type and wildfire in the Alberta boreal mixedwood: What do fires burn? *Ecological Applications*, 11(1), 97-110.
- Loboda, T. V., & Csiszar, I. A. (2007). Reconstruction of fire spread within wildland fire events in Northern Eurasia from the MODIS active fire product. *Global and Planetary Change*, 56(3-4), 258-273. doi: <http://dx.doi.org/10.1016/j.gloplacha.2006.07.015>
- McRae, R. H. D. (1992). Prediction of Areas Prone to Lightning Ignition. *International Journal of Wildland Fire*, 2(3), 123-130.
- Murphy, B. P., Bradstock, R. A., Boer, M. M., Carter, J., Cary, G. J., Cochrane, M. A., . . . Bowman, D. M. J. S. (2013). Fire regimes of Australia: a pyrogeographic model system. *Journal of Biogeography*, 40(6), 1048-1058. doi: 10.1111/jbi.12065
- Preisler, H. K., Brillinger, D. R., Burgan, R. E., & Benoit, J. W. (2004). Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire*, 13(2), 133-142.
- R Development Core Team. (2014). R: A language and environment for statistical computing (Version 3.1.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12.
- Russell-Smith, J., Yates, C. P., Whitehead, P. J., Smith, R., Craig, R., Allan, G. E., . . . Gill, A. M. (2007). Bushfires 'down under': patterns and implications of contemporary Australian landscape burning. *International Journal of Wildland Fire*, 16(4), 361-377.
- Schneider, G., & Sutherland, S. (undated). Fire in coastal wetlands - implications for management: NSW Land and Property Management Authority.
- Sharples, J. J. (2009). An overview of mountain meteorological effects relevant to fire behaviour and bushfire risk. *International Journal of Wildland Fire*, 18(7), 737-754.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857), 1285-1293.
- Syphard, A. D., Radeloff, V. C., Keuler, N. S., Taylor, R. S., Hawbaker, T. J., Stewart, S. I., & Clayton, M. K. (2008). Predicting spatial patterns of fire on a southern California landscape. *International Journal of Wildland Fire*, 17(5), 602-613.
- Venables, W. N., & Ripley, B. D. (1999). *Modern Applied Statistics with S-PLUS* (3rd ed.). New York: Springer.
- Vilar, L., Woolford, D. G., Martell, D. L., & Martn, M. P. (2010). A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. *International Journal of Wildland Fire*, 19(3), 325-337.
- Wintle, B. A., Elith, J., & Potts, J. M. (2005). Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. *Austral Ecology*, 30(7), 719-738. doi: 10.1111/j.1442-9993.2005.01514.x