Modelling spatial patterns of wildfire occurrence in South-Eastern Australia

Yang Zhang\textsuperscript{a}, Samsung Lim\textsuperscript{a} and Jason John Sharples\textsuperscript{b}

\textsuperscript{a}School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW, Australia; \textsuperscript{b}School of Physical, Environmental and Mathematical Sciences, The University of New South Wales, Canberra, ACT, Australia

\textbf{ABSTRACT}
This paper describes the development and validation of spatial models for wildfire occurrence at a broad landscape scale. The hotspots databases from the Moderate Resolution Imaging Spectroradiometer (MODIS) and logistic regression models are investigated for the comprehensive understanding of environmental and socioeconomic determinants regulating the spatial distribution of wildfires over the 11-year period 2003–2013. The probability of occurrence of at least one fire on a 1 km\textsuperscript{2} grid cell in a 1,030,000 km\textsuperscript{2} region located in South-Eastern Australia is studied for the prediction of future fire occurrence. Our research shows that wildfires are most likely to occur in mountainous areas, forests, savannas and lands with high vegetation coverage, and are less likely to occur on grasslands and shrublands. Wildfires also tend to occur in areas near human infrastructures. Environmental variables are strong individual predictors of fire occurrence while socioeconomic variables contribute more to the final model. The influence of environmental and socioeconomic conditions on wildfire occurrence and the spatial patterns of wildfires identified in this study can assist fire managers in implementing appropriate management actions in South-Eastern Australia. This paper also demonstrates the potential of applying the MODIS active fire product in wildfire occurrence studies.

\textbf{1. Introduction}
Wildfire is a major environmental and ecological issue across the world. Wildfires can alter the structure of ecosystems, affect ecological processes and functions, threaten human lives and increase fire-suppression costs (Bowman et al. 2009). Australia is recognized as one of the most flammable continents in the world. Wildfires in the southeastern region of the continent can occur at any time of the year, but mostly during the hotter months from October to March. Wildfires in a densely settled area require rapid, appropriate policy responses because they can cause massive loss of life and property (Russell-Smith et al. 2007). The limitation of resources and the requirement for quick responses demand accurate prediction of where fires will occur. Employing fire occurrence records within empirical models is essential to quantify the characteristics of fire activities to support planning and decision-making (Andrews & Finney 2007). These models can be used to identify fire-prone areas and help forest managers target suppression efforts (Pew & Larsen 2001; Syphard et al. 2008; Romero-Calcerrada et al. 2010; Wang & Anderson 2010; Renard et al. 2012).
Fire occurrence has been studied in many countries, mainly in North America (Preisler et al. 2004; Syphard et al. 2008; Preisler et al. 2011; Gracelwicz et al. 2012; Magnussen & Taylor 2012; Hawbaker et al. 2013), Europe (Catry et al. 2010; Chuvieco et al. 2010; Oliveira et al. 2012; Martínez-Fernández et al. 2013; Chuvieco et al. 2014) and Australia. Most of recent fire studies in Australia were focused on regional areas such as the Australian Capital Territory (McRae 1992), the mallee woodlands and heathlands of South-Eastern Australia (Krusel et al. 1993; Gibson et al. 2015), the entire Victoria (Dowdy & Mills 2012a, 2012b), the Sydney region (Bradstock et al. 2009; Penman et al. 2013) and the south-west Western Australia (Plucinski 2014; Plucinski et al. 2014). These studies are either temporally based (e.g. Dowdy & Mills 2012a) which emphasize the influences of dynamic meteorological variables on the temporal variation of fire incidences, or spatially based (e.g. Penman et al. 2013) which investigate the effects of geographic variables on the spatial pattern of fire occurrence (Plucinski 2011). None of the spatially based fire occurrence studies are conducted at a scale that covers a complete fire-prone state in Australia.

Wildfires are caused by both lightning incidences and human activities. The spatial distributions of both types of fires are usually regulated by top-down and bottom-up drivers across multiple scales (Parisien & Moritz 2009). At broad spatial scales dominated by top-down drivers, wildfires are selective with respect to different land cover types on account of their relationship with fuel loads and types (Gumming 2001; Mermoz et al. 2005; Moreira et al. 2009). Vegetation indexes have also proved to be useful for fire danger prediction due to their connection with fuel moisture content (Bisquert et al. 2011; Caccamo et al. 2012). At fine spatial scales dominated by bottom-up drivers, topography influences fire extent by affecting the rate and direction of fire spread (Rothermel 1983) and by creating microclimates that affect fuel moisture content and air temperature (Heyerdahl et al. 2001; Sharples 2009). Socioeconomic variables such as distance to roads are found to be significant in predicting fire (especially human-caused fire) locations at both regional (Romero-Calcerrada et al. 2008; Vilar et al. 2010) and national (Hawbaker et al. 2013; Rodrigues et al. 2014) scales.

Most fire occurrence studies are based on historical data which are sourced from fire management agency observation records (e.g. Bradstock et al. 2009). In the past few decades, the emerging space-borne sensors with short revisit times and high accuracy measurements make it possible to understand fire occurrence over a large scale (Pausas & Keeley 2009). In Australia, continental-scale spatial and temporal fire patterns have been statistically assessed based on Advanced Very High Resolution Radiometer (AVHRR) imagery (Craig et al. 2002; Russell-Smith et al. 2007). The Moderate Resolution Imaging Spectroradiometer (MODIS) on Terra (1999) and Aqua (2002) which has been found to be the most precise and reliable system regarding the accuracy and completeness of target detection (Justice et al. 2002) provides a better option. Studies on fire occurrence (e.g. Renard et al. 2012; Hawbaker et al. 2013; Yulianti et al. 2013) have been undertaken using spatial and temporal wildfire occurrence information provided by active fire products (Giglio et al. 2003) generated from MODIS data. However, none of these studies have explored the relationship between MODIS-based active fire locations and their determinants in Australia.

The choice of modelling method is dependent on the characteristic of the dependent variable. When using a very fine spatial resolution (e.g. 1 km), a binomial response is required because only presence/absence is recorded (Taylor et al. 2013). Logistic regression has been intensively used to model the probability of fire occurrence (Chou 1992; Pew & Larsen 2001; Vasconcelos et al. 2001; Syphard et al. 2008; Padilla & Vega-García 2011; Magnussen & Taylor 2012). The advantage of logistic regression is that it can relate locally observed independent variable to each fire record (Taylor et al. 2013) and the result is easy to interpret. Although other techniques, i.e. non-parametric methods may be theoretically superior to logistic regression, the issues of overfitting and lack of transparency are their drawbacks (Magnussen & Taylor 2012).

In this paper, we used 11 years (2003—2013) of MODIS active fire data in combination with environmental and socioeconomic explanatory data to build a binary logistic regression model in order to estimate the probability of at least one fire within a 1 km² grid cell, and to generate a fire
occurrence probability map over South-Eastern Australia. Our objective is to provide quantitative statistics of the environmental and socioeconomic factors contributing to the spatial distribution of wildfires at a broad landscape scale and provide practical guidance for fire management in this region.

2. Methodology

2.1. Study area

South-Eastern Australia defined in this research contains the mainland of New South Wales (NSW), Victoria (VIC) and Australian Capital Territory (ACT), covering an area of 1,030,000 km² (figure 1). The dominant land cover types (see Section 2.2.2 and figure 1) in this region are open shrublands (39%), croplands (26%), evergreen broadleaf forests (13%) and woody savannas (10%), as calculated by the authors. The climate in this region is temperate: cold and damp in winter, hot and dry in summer. Low-frequency, high-intensity fires occur in this area due to the latitudinal gradient in summer monsoon rainfall activity (Murphy et al. 2013).

Figure 1. Location and the land cover types of South-Eastern Australia. NSW, New South Wales; VIC, Victoria; ACT, Australian Capital Territory.
2.2. Data description

A range of data were collected and transformed for different purposes. Some of them were used for statistical analysis (see Table 1) while others, e.g. CLUM land use (see Section 2.2.1 for the details), were used as filters of fire occurrence points.

### 2.2.1. Land use

The Catchment scale Land Use of Australia Map (CLUM) was used as a filter of fire occurrence points in this study. The data-set was updated in March 2014 and published by the Department of Agriculture (DA). Collected by the Australian Collaborative Land Use and Management Program (ACLUMP) state and territory partners, land use in CLUM is mapped at a detailed catchment scale (1:25,000 or 1:100,000). CLUM data is classified according to the Australian Land Use and Management (ALUM) Classification version 7, which makes it a single dominant land use map for a given area based on the major management objective. The classification has six primary classes of land use: (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. Resampling of the 50-m resolution data to 1 km was performed using the majority algorithm to make it consistent with the resolution of MODIS hotspots.

### 2.2.2. Land cover

The MODIS 500 m Land Cover Type product (MCD12Q1) based on the classification system defined by the International Geosphere Biosphere Program (IGBP) was used for the fire occurrence points filtering and the statistical analysis. It identified 17 classes (11 natural vegetation classes, 3 developed and mosaic land classes and 3 non-vegetated land classes) at a global scale. We reclassified them into six primary classes to consider the influence of primary vegetation types on fire occurrence in the study area: (1) forests; (2) shrublands; (3) savannas; (4) grasslands; (5) permanent wetlands; (6) croplands, water bodies and others. The MCD12Q1 data in 2003 was chosen because it was the year that had the highest number of fire points in the study area. The data was also resampled from a resolution of 500 m to 1 km using the majority algorithm.

### Table 1.

| Variable                        | Source          | Description of original data                                                                 |
|---------------------------------|-----------------|-----------------------------------------------------------------------------------------------|
| Explanatory variables           |                 |                                                                                              |
| Land cover                      | NASA            | MODIS 500-m MCD12Q1, 2003 Six primary classes (forests; shrublands; savannas; grasslands; permanent wetlands; croplands, water bodies and others) |
| NDVI                            | NASA            | MODIS 1-km MYD13A3 NDVI, Collection 5, January 2003                                            |
| EVI                             | NASA            | MODIS 1-km MYD13A3 EVI, Collection 5, January 2003                                             |
| Elevation                       | NASA            | ASTER 30-m GDEM, V2                                                                          |
| Slope                           | NASA            | Derived from elevation grid                                                                  |
| Northwesst                      | NASA            | NW = \cos(\text{aspect}) = = -1, 0, \cos((\text{aspect} + 45) + \pi/180))                    |
| Distance to zero residual contours | NASA       | Derived from elevation grid                                                                  |
| Distance to primary road         | OSM             | Mean Euclidean distance                                                                      |
| Distance to secondary road       | OSM             | Mean Euclidean distance                                                                      |
| Distance to railway             | OSM             | Mean Euclidean distance                                                                      |
| Distance to WUI                 | OSM             | WUI, wildland–urban interface                                                              |
| Population density              | ABS             | LGA units, 2003                                                                               |
| Fire occurrence                 | NASA            | Derived from MODIS 1-km MCD14ML, Collection 5                                                 |
2.2.3. Vegetation
We used Collection 5 MODIS global monthly Vegetation Index product series (MYD13A3) as an indicator of fuel load in the study area. The MYD13A3 data is provided monthly at 1-km spatial resolution as a gridded level-3 product. Two types of vegetation index were used: Normalized Difference Vegetation Index (NDVI) and a new Enhanced Vegetation Index (EVI). NDVI is the most commonly used index to assess live fuel moisture content (Hardy & Burgan 1999; Chuvieco et al. 2004; Caccamo et al. 2012), but can experience saturation under high-density vegetation conditions (Sellers 1985). EVI minimizes canopy background variations and has an improved sensitivity in high biomass regions. We tested both NDVI and EVI in January 2003.

2.2.4. Topography
Our topographical variables include elevation, slope, transformed aspect index (northwestness) and distance to zero residual contours. The elevation layer was resampled at 1-km resolution from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 (GDEM V2) 30 m data. The resampling was carried out by calculating the mean of the cell values within each 1 km² rectangular block. Slope (in percentage) and aspect (in degrees) maps were derived from elevation data. Because the aspect is a circular variable that cannot be used in linear statistics, we cosine-transformed the aspect layer to a linear variable to obtain an index of ’northwestness’ which can better distinguish xeric exposures (high index values) from mesic exposures (low index values) (Franklin et al. 2000). Another topographical variable tested in this study is the distances to zero meso-scale elevation residual contours as suggested by McRae (1992). These contours were produced by generating a macro-scale surface using an ordinary Kriging interpolation method, and subtracting it from a finer resolution elevation surface to finally leave a zero meso-scale elevation surface. Because the resolution of the explanatory variable should not be finer than 1 km in this study, the meso-scale contours we generated are coarser than those in McRae’s study.

2.2.5. Socioeconomic data
Most wildfires are of anthropogenic origin, either deliberate or accidental, which indicates the potential relationship between fire occurrence and socioeconomic factors such as wildland-urban interface (WUI), distance to roads and railways and population density. The WUI in this study was defined as the boundary between wildlands and urban areas. Wildland (see Section 2.2.6) and urban residential areas were derived from the 50-m resolution CLUM map. A raster map representing the distance to the nearest WUI was generated at a 1-km resolution. We derived primary roads, secondary roads and railways from the database of OpenStreetMap (OSM), a collaborative project to provide open, freely available geographic data of the world (Neis et al. 2011). From the OSM database, 1-km resolution distance maps were generated based on the Euclidean distance to the nearest road. We used population density data in 2003 with spatial units of Local Government Area (LGA) from Australian Bureau of Statistics (ABS).

2.2.6. Fire occurrence
The dependent variable — wildfire occurrence — was originally derived from Collection 5 MODIS global monthly fire location product (MCD14ML) detected using a contextual algorithm (Giglio et al. 2003). The MCD14ML data is the combined Terra and Aqua MODIS Level 2 swath 5-min MOD14/MYD14 active fire product, so it contains a precise date and time of the active fire (hotspot) when MODIS passes over. The spatial resolution of MCD14ML is 1 km.

The validation of the MODIS active fire product based on the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) imagery shows a commission error between 2% and 3% globally (Morissette et al. 2005; Csiszar et al. 2006; Schroeder et al. 2008), even though high commission errors are found in low fire activity areas such as urban sites and agricultural locations (Hantson et al. 2013). Each fire activity is provided with a detection confidence level (low, medium
or high); however, hotspots of all confidence levels are included in our study because the low-confidence hotspots provide additional information which outweighs the problem of slightly higher commission errors (Hantson et al. 2013).

In addition to the existence of commission error, the MODIS active fire product has other limitations. First, it does not distinguish between fire causes (lightning and human), which makes it impossible to analyse each explanatory variable in its causality context. Also, an omission error of 18% related to the size of the fire patch is found (Hawbaker et al. 2008). Because the majority of human-caused fires may not be large enough to be detected by MODIS sensors, a bias towards natural-caused fires should be realized. Moreover, the fact that prescribed burning and other non-wildfires were also recorded makes the data filter process necessary. In addition, MODIS active fire points represent fire activities being recorded at given times and locations, which means that they contain information on both fire ignition and spread. Therefore, fire occurrence in our study should be defined as the occurrence of fire activity rather than ignition.

Our study period was from January 2003 to December 2013, including all the years when both MODIS Aqua and Terra data products were available. There were missing data in 2007 from mid-August, on part of 21 April 2009, and on 22 April 2009. The number of MODIS hotspots was 176,884.

As mentioned earlier, it is inappropriate to employ all the MODIS hotspots into our models because fires that occurred on non-wildland areas are recorded. A mask representing wildland areas in the study area was generated using both CLUM Land use and MODIS land cover data to filter the fire hotspots. The pixel value of the binary raster was 0 if located at land cover classes (6), as well as at land use classes (3), (4), (5) and (6) except for plantation forestry, or 1 if located somewhere else. All the hotspots on the non-wildland areas were removed. This process was also able to minimize the influence of commission errors of the product. We then generated a histogram representing the monthly distribution of fire hotspots (figure 2). The monthly profiles for fires show the

![Figure 2. Monthly distribution of fire hotspots from 2003 to 2013 for South-Eastern Australia.](image)
majority of fires occurring from November to February when fire danger levels are the highest, although these figures experience slight upper fluctuations in spring and autumn months, due to the influence of prescribed burning programmes that are generally conducted during cooler periods (autumn and early spring in our study area). We therefore only used fire data within the typical fire season (November to February) to mitigate the influence of prescribed burning.

We created a continuous 1-km spatial resolution density map by calculating the number of fire points occurring during the 11 fire seasons of 2003—2013 that fall within each cell of wildland areas. In total, there were 538,690 data points in the density map which included 28,761 presence points and 509,929 absence points. Data analysis was conducted based on sample statistics in order to reduce the data volume. Among many sampling methods, we utilized the stratified sampling method, i.e. the ratio of presence to absence is preserved. 10,000 samples (which include 596 presence and 9,404 absence points) were randomly selected. That is, 1.86% of the population is sampled. All the non-zero values were converted to 1 so that the value of each point denotes the presence (figure 3) or absence of at least one fire incidence within each cell.

2.3. Modelling approach

To estimate the probability, $P$, of at least one fire within a cell, we developed a multiple logistic regression model. Let the probability, $P_i$, of at least one fire in cell $i$, and $x_j$ be the value of the $j$th covariate in cell $i$, the logistic regression can be defined as:

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

(1)

where $\beta_0$ is an intercept, and $\beta_n$ are coefficients for explanatory variables, $x_{ni}$. We fitted the model using all the fire occurrence background points. The ratio of ones to zeros is 1:24.6.

Figure 3. Map showing the distribution of the dependent variable representing the presence of fire (black dots). The grey colour represents non-wildland areas in South-Eastern Australia.
We developed univariate logistic regression models for each explanatory variable to evaluate the independent influence of each variable on the fire occurrence. Following the suggestions of Serneels and Lambin (2001), we also tested the performance of quadratic or logarithmic versions of continuous variables. Our final model was chosen by implementing the Akaike Information Criterion (AIC) in a backwards stepwise algorithm (Venables & Ripley 1999).

To avoid the effects of multicollinearity, we used Spearman’s rank correlation to compare the correlations among continuous explanatory variables (table 2). A correlation above 0.6 (Wintle et al. 2005) among the explanatory variables was found between NDVI and slope ($r = 0.65, P < 0.001$), NDVI and population density ($r = 0.81, P < 0.001$), NDVI and EVI ($r = 0.97, P < 0.001$), EVI and population density ($r = 0.84, P < 0.001$), slope and elevation ($r = 0.67, P < 0.001$), distance to WUI and population density ($r = 0.62, P < 0.001$). Therefore, we excluded EVI, slope and population density in the further analysis. We also implemented a diagnostic procedure using the variance inflation factor (VIF), an measure that shows how much the variance of an coefficient is increased due to collinearity (Belsley et al. 1980). Instead of comparing the correlation between pairs of variables, VIF calculates the linear relationship between one variable and all the other variables. VIFs ranging from 1.01 to 8.29 indicate that there is no evidence of multicollinearity, as recommended by Belsley et al. (1980).

We computed the receiver operating characteristics (ROC) curve of our models to determine the optimal discrimination threshold for predicting fire occurrence. By plotting true positive rate (sensitivity) against the false positive rate (specificity), we were able to evaluate the accuracy of the model. The area under the curve (AUC) of the ROC curve is used to measure the model fit, with a measure of 0.9—1 excellent, 0.8—0.9 good, 0.7—0.8 fair, 0.6—0.7 poor and 0.5—0.6 fail (Swets 1988). To validate the performance of the final model, we used a 10-fold cross-validation approach (Breiman et al. 1984). The procedure was to randomly partition the original samples into 10 groups of approximately equal size, use a single group as the validation data and the remaining 9 groups as training data. The process was repeated 10 times and each time the AUC of the validation data was calculated. A statistical summary of the 10 AUCs was calculated to measure the uncertainty related to the location of the presence or absence. The mean of the AUCs is the cross-validated AUC estimate.

The contribution of each variable was calculated by conducting a jackknife procedure based on the change in AUC (Bar Massada et al. 2013), which was the 10-fold cross-validated AUC in this study. The approach consists of removing explanatory variables from the full model one at a time, and calculating the cross-validated AUC. The difference of AUC values between the full-model and the model without the variable denotes the loss of explanatory power of the model in absence of a given factor. In addition, the AUC values of univariate models were computed, and the variables were ranked accordingly.

All statistical analyses in this study were conducted using R package version 3.1.1 (R Development Core Team 2014). Logistic regression was fitted using a generalized linear model (GLM). VIFs, ROCs and cross-validated AUCs were computed using R modules fmsb (Nakazawa 2014), pROC (Robin et al. 2011) and cvAUC (LeDell et al. 2014), respectively.

3. Results

According to the univariate logistic regression modelling results (table 3), all the explanatory variables are statistically significant ($P \leq 0.05$), except for northwestness ($P = 0.37$). In the group of MODIS land cover, wildfires are most likely to occur on forests and savannas, while least likely to occur on shrublands and grasslands. Permanent wetlands is not a significant predictor ($P = 0.95$) of fire occurrence due to its small sample size (four observations). NDVI also shows the expected positive relationship to the response variable. The occurrence of fire is positively related to elevation, and is negatively related to distance to zero meso-scale elevation residual contours. Fire occurrence is negatively related to all the socioeconomic variables (distance to primary road, distance to...
Table 2. Correlation matrix for continuous variables.

|        | NDVI  | Elevation | Slope | Northwestness | Distance to zero residual contours | Distance to primary road | Distance to secondary road | Distance to railway | Distance to WUI | Population Density |
|--------|-------|-----------|-------|---------------|-----------------------------------|--------------------------|----------------------------|--------------------|-----------------|---------------------|
| EVI    | 0.97  | 0.52      | 0.59  | 0.01          | -0.30                            | -0.32                    | -0.33                      | -0.45              | -0.55           | 0.84                |
| NDVI   | 0.56  | 0.65      | 0.002 | -0.32         | -0.29                            | -0.31                    | -0.44                      | -0.52              | 0.81            |                     |
| Elevation | 0.67  | 0.03      | -0.35 | -0.25         | -0.14                            | -0.22                    | -0.35                      | 0.41               |                |                     |
| Slope  | -0.03 | -0.45     | -0.01 | -0.004        | 0.03                             | 0.002                    | 0.02                       | 0.002              |                |                     |
| Northwestness | -0.01 | -0.004   | 0.15  | 0.11          | 0.12                             | 0.26                     | -0.26                      |                    |                |                     |
| Distance to zero residual contours | -0.01 | 0.03     | 0.06  | 0.45          | 0.39                             | -0.39                    | -0.34                      | -0.34              |                |                     |
| Distance to primary road               |        |          |        |                |                                  |                          |                           |                    |                |                     |
| Distance to secondary road              |        |          |        |                |                                  |                          |                           |                    |                |                     |
| Distance to railway                      |        |          |        |                |                                  |                          |                           |                    |                |                     |
| Distance to WUI                          |        |          |        |                |                                  |                          |                           |                    |                |                     |
| Distance to WUI                          |        |          |        |                |                                  |                          |                           |                    |                |                     |


secondary road, distance to railway and distance to WUI), which suggests that fires are more likely to occur close to human facilities and urban areas.

The final model for fire occurrence includes four environmental variables (land cover, NDVI, elevation, northwestness) and three socioeconomic variables (distance to primary road, distance to secondary road and distance to WUI). The AUC values corresponding to the 10-fold procedure range from 0.858 to 0.906, with a standard deviation of 0.014 (table 4). This means that, although the AUC value can be affected by the location of the points, the variability of the AUC value is acceptable. The cross-validated AUC estimate is 0.886, which indicates that our model performs well in predicting fire occurrence.

According to the AUC values of the univariate models (figure 4), NDVI shows the strongest predictive power, followed by elevation and land cover. The jackknife estimate of variable importance (figure 4) shows slightly different results. Elevation contributes the most in wildfire occurrence prediction, which coincides with its performance in the univariate model. Distance to WUI and land cover rank the second and the third, respectively. The remaining variables are ordered as follows: distance to primary road, NDVI, distance to secondary road and northwestness.

A 1-km resolution fire occurrence probability map was generated by applying the coefficients of the final model to raster layers corresponding to the explanatory variables (figure 5(a)). Relatively high wildfire probabilities (greater than 0.2) were found in forestry, high elevation areas close to urban areas, while low probabilities (less than 0.2) existed inland away from the coast and mountainous areas. The largest area of high wildfire probability existed in the Great Dividing Range along the coastline of the study area, with the most flammable areas located in the forestry areas of the Australian Alps extending across eastern VIC, southeastern NSW and the ACT, followed by the wildland-urban interface areas at New England Range in northeastern NSW and Blue Mountains above the Sydney Basin. Natural conservation regions such as Mount Kaputar National Park in northeastern NSW and Grampians National Park in western VIC are also more likely to burn. In addition, 10 prediction maps corresponding to the 10-fold procedure were generated and converted into a standard deviation map (figure 5(b)), which would provide insights into the spatial

| Table 3. Univariate logistic regression results for variables explaining wildfire occurrence in South-Eastern Australia. |
| Explanatory variables | Coefficient | Standard error | P value | 95% confidence interval |
|-----------------------|-------------|----------------|---------|------------------------|
| Land cover            |             |                |         |                        |
| Forest (Intercept)    | −1.44       | 0.05           | <0.001  | −1.54 −1.34            |
| Shrublands            | −3.73       | 0.18           | <0.001  | −4.10 −3.40            |
| Savannas              | −0.83       | 0.12           | <0.001  | −1.07 −0.60            |
| Grasslands            | −2.35       | 0.36           | <0.001  | −3.14 −1.71            |
| Permanent Wetlands     | −12.12      | 267.7          | NA      | 18.57                  |
| Vegetation index      |             |                |         |                        |
| NDVI                  | 0.00060     | 0.00002        | <0.001  | 0.00055 0.00064        |
| Topography            |             |                |         |                        |
| Elevation             | 0.0033      | 0.0001         | <0.001  | 0.0031 0.0036          |
| Northwestness         | 0.05        | 0.059          | 0.37    | −0.06 0.17             |
| Distance to zero residual contours | −0.00047 | 0.00005        | <0.001  | −0.00056 −0.00038     |
| Anthropogenic data    |             |                |         |                        |
| Distance to primary road | −0.000017 | 0.000002       | <0.001  | −0.000021 −0.000014   |
| Distance to secondary road | −0.000031 | 0.000003       | <0.001  | −0.000037 −0.000025   |
| Distance to railway   | −0.0000993 | 0.0000099      | <0.001  | −0.00000076            |
| Distance to WUI       | −0.000042  | 0.000002       | <0.001  | −0.000047 −0.000037   |

Table 4. Summary of the AUCs of the 10-fold cross-validation.

| Min  | Max  | Median | Mean  | Standard deviation | Variance | Range  |
|------|------|--------|-------|-------------------|----------|--------|
| 0.858 | 0.906 | 0.885  | 0.886 | 0.014             | 0.0002   | 0.048  |

Note: The mean of AUC values is the cross-validated AUC estimate.
distribution of the uncertainty of the final model. The variability was distributed primarily in high fire probability areas, e.g. the Australian Alps.

4. Discussion

Our model is able to describe the spatial pattern of wildfire occurrence in South-Eastern Australia over the 11-year period from 2003 to 2013. Wildfire locations in our study area are found to be significantly influenced by land cover types. Forests are most susceptible to fire due to the dominance of the fire-prone eucalyptus-related vegetation and heavy fuel loads. Savannas rank second probably due to their inherent features relating to ease of ignition (Murphy et al. 2013). Shrublands are least susceptible to burning compared to other land cover types, due to the lower predominance of grass components in these areas (Murphy et al. 2013). Our result is slightly inconsistent with the findings in other landscapes (e.g. Mermoz et al. 2005; Oliveira et al. 2014), possibly because of the low-level shrub canopy cover (<60%) in most shrub areas of South-Eastern Australia.

As expected, fires are more likely to occur in areas with high values of vegetation index on account of its strong relationship with fuel flammability (Caccamo et al. 2012). Fires are also more
likely to occur in areas with higher elevation. This can be explained by the fact that in our study area, the spatial distribution of the vegetation corresponds to that of elevation, which means there is more vegetation to be burnt at higher elevation. When it comes to the influence of aspect on fire occurrence, some researchers in the northern hemisphere have found that a southerly aspect (or a northerly aspect in the southern hemisphere) is more flammable because south-facing slopes receive longer and more direct solar exposure, decreasing the fuel moisture content and enhance its flammability (Mouillot et al. 2003; Mermoz et al. 2005), while others found that northern slopes are more fire prone because much more water is available and results in heavier fuel loads (Carmo et al. 2011; Oliveira et al. 2014). However, in this study northwesterly aspect is not able to predict fire occurrence. This may be because the coarse spatial resolution fails to provide sufficient or correct information on solar exposure and fuel load. Fires tend to distribute in areas near the zero meso-scale elevation residual contour, which is generated by removing micro- and macro-scale variation of elevation, leaving only the meso-scale residual and generating contours accordingly. 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. Fires are located in areas near human infrastructures (roads and railways) and WUI is consistent with the results of other studies at small landscape scales (e.g. Penman et al. 2013).

Most environmental factors (NDVI, elevation and land cover) are informative when analysed independently, which is expected because their spatial patterns correspond with the distribution of wildfires at such a broad spatial scale. On the other hand, some of them (e.g. land cover) have low variable contributions because much of the information they provide is included in a relatively more influential variable (e.g. elevation).

Socioeconomic variables do not exhibit good predictive power in the univariate analysis. There are likely several reasons for this. First, the indistinguishability of ignition sources of MODIS active fire product that may hinder the contribution of the socioeconomic variables. Second, the fact that not all human-caused fire occurrences are retrieved also indicates that the independent predictive powers of socioeconomic factors are possibly being underestimated in this study. Third, our fire occurrence points contain information for both ignition and spread. Although fire ignitions in Australia have been proven to be strongly influenced by human activities (Willis 2005), fire spread is fundamentally a function of fuel, climate and terrain (Pyne et al. 1996). However, the contribution of variables to the final model tells a slightly different story. Distance to WUI contributes more predictive power compared with NDVI and land cover, which supports the influence of human activities in fire occurrence. Therefore, socioeconomic variables should not be ignored in fire risk assessments at broad landscape scales. Moreover, the association between human activity and fire occurrence indicates the threat of wildfire to human lives and assets. Reducing the fuel load near densely settled areas close to fire-prone bushlands is therefore an essential issue in the perspective of wildfire management.

The fire probability map produced from the final model illustrates the most fire-prone locations in South-Eastern Australia. The distribution of relative high uncertainty in high fire occurrence areas and the limited predictive capacity of logistic regression in these areas (Rodrigues & de la Riva 2014) means that fire probability has possibly been underestimated. Nevertheless, the prediction map still provides useful information on areas where environmental and socioeconomic conditions meet the requirement for enhanced likelihood of fire occurrence. According to this map, firefighting resources and fire prevention activities should be allocated close to mountainous areas, forests and savannas, as well as lands with heavy fuel loads. Areas close to WUI and transport networks should also be emphasized.

We use MODIS data rather than historical records in this study because the former is globally accessible, which makes it possible to conduct a study at a broad spatial scale, apply the method to other regions of the world, assess the suitability of the model and explore the variation of spatial patterns in different study areas. This is especially true in data-poor regions. However, researchers should bear in mind the inherent drawbacks of the MODIS active fire product such as the existence
of commission error (which can be minimized by introducing controlling factors), the indistinguishability of ignition sources and the bias towards natural-caused fires. Another defect of our model is that the explanatory variables only represent the environmental and socioeconomic conditions of the study area for a short period of time even though the active fire data covers 11 years. To reduce the impact of the temporal mismatch, we chose data recorded at the most appropriate time (e.g. January 2003). Nevertheless, the overall goodness of fit of our final model is satisfactory. Further improvement of the model would be obtained by the introduction of precise occurrence data that has lower omission error or can identify ignition types. Adding more explanatory variables, especially climate variables, would also be helpful.

5. Conclusion

In this paper, we used logistic regression in combination with land cover, vegetation index, topographic and socioeconomic information to characterize the spatial pattern of the fire occurrence at least one fire on a 1 km² grid in South-Eastern Australia over the period 2003—2013. Our models and the final map suggest that mountainous areas, forests, savannas and lands with high vegetation coverage would be most fire-prone, while grasslands and shrublands can be less vulnerable to wildfire in the study area. Wildfires also tend to occur in areas near human infrastructures and the WUI. Environmental variables are powerful in predicting fire occurrence when analysed individually while socioeconomic variables contribute more to the final model. The results show the extended knowledge about the influence of environmental and socioeconomic conditions on wildfire occurrence and the spatial pattern of wildfire in the mainland NSW, VIC and ACT, which can help fire agencies in these three regions better arrange their limited resources and target management activities. Our study also demonstrates that the MODIS active fire product is a useful data source to study environmental and socioeconomic controls on the distribution of wildfire although attention should be paid to the data manipulation procedure and the interpretation of the modelling result.

Acknowledgments

We would like to thank the reviewers for their insightful suggestions that greatly improved the manuscript. We also thank Mr. Hossein Asefi for his input to the mathematical approach presented in this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Yang Zhang is a PhD candidate in the School of Civil and Environmental Engineering at the University of New South Wales, Sydney, Australia. With a background in geographic information system (GIS) and remote sensing (RS) applications, Yang research interest includes understanding environmental and anthropogenic drivers regulating spatial patterns of wildfire occurrence at a broad scale in south-eastern Australia.

Samsung Lim is an associate professor in the School of Civil and Environmental Engineering at the University of New South Wales, Sydney, Australia. His main research aim is to apply the geospatial information approach to real-world problems and help decision-making in the fields of asset/utility management, natural resources exploration, intelligent transportation, natural disaster and emergency management, and public health research.

Jason John Sharples is an associate professor in the School of Physical, Environmental and Mathematical Sciences (PEMS) at the Canberra campus of the University of New South Wales, Australia. His main research interests include the mathematical modelling of combustion, and bushfire and associated processes.
References

Andrews P, Finney M. 2007. Predicting wildfires. Sci Am. 297:47–55.

Bar Massada A, Syphard AD, Stewart SI, Radeloff VC. 2013. Wildfire ignition-distribution modelling: a comparative study in the Huron—Manistee National Forest, Michigan, USA. Int J Wildland Fire. 22:174–183.

Belsley DA, Kuh E, Welsch RE. 1980. Regression diagnostics: identifying influential data and sources of collinearity. Hoboken (NJ): Wiley.

Bisquert MM, Sánchez JM, Caselles V. 2011. Fire danger estimation from MODIS Enhanced Vegetation Index data: application to Galicia region (north-west Spain). Int J Wildland Fire. 20:465–473.

Bowman DMJS, Balch JK, Artaxo P, Bond WJ, Carlson JM, Cochrane MA, D’Antonio CM, DeFries RS, Doyle JC, Harrison SP, et al. 2009. Fire in the earth system. Science. 324:481–484.

Bradstock RA, Cohn JS, Gill AM, Bedward M, Lucas C. 2009. Prediction of the probability of large fires in the Sydney region of south-eastern Australia using fire weather. Int J Wildland Fire. 18:932–943.

Breiman L, Friedman J, Stone CJ, Olshen RA. 1984. Classification and regression trees. Boca Raton (FL): Chapman and Hall/CRC Press.

Caccamo G, Chisholm LA, Bradstock RA, Puotinen ML, Pippen BG. 2012. Monitoring live fuel moisture content of heathland, shrubland and sclerophyll forest in south-eastern Australia using MODIS data. Int J Wildland Fire. 21:257–269.

Carmo M, Moreira F, Casimiro P, Vaz P. 2011. Land use and topography influences on wildfire occurrence in northern Portugal. Landsc Urban Plan. 100:169–176.

Catry FX, Rego FC, Baçio FL, Moreira F. 2010. Modeling and mapping wildfire ignition risk in Portugal. Int J Wildland Fire. 18:921–931.

Chou YH. 1992. Management of wildfires with a geographical information system. Int J Geogr Inf Syst. 6:123–140.

Chuvieco E, Aguado I, Jurdao S, Pettinari M, Yebra M, Salas J, Hantson S, De La Riva J, Ibarra P, Rodrigues M. 2014. Integrating geospatial information into fire risk assessment. Int J Wildland Fire. 23:606–619.

Chuvieco E, Aguado I, Yebra M, Nieto H, Salas J, Martín MP, Vilà L, Martínez J, Martín S, Ibarra P, et al. 2010. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. Ecol Model. 221:46–58.

Chuvieco E, Cocero D, Riano D, Martin P, Martínez-Vega J, de la Riva J, Pérez F. 2004. Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. Remote Sens Environ. 92:322–331.

Craig R, Heath B, Raisbeck-Brown N, Steber M, Marsden J, Smith R. 2002. The distribution, extent and seasonality of large fires in Australia, April 1998-March 2000, as mapped from NOAA-AVHRR imagery. In: Russell-Smith J, Craig R, Gill AM, Smith R, Williams J, editors. Australian fire regimes: contemporary patterns (April 1998–March 2000) and changes since European settlement. Canberra: Department of the Environment and Heritage; p. 1–77.

Csizsar IA, Morisette JT, Giglio L. 2006. Validation of active fire detection from moderate-resolution satellite sensors: the MODIS example in Northern Eurasia. ITGRS. 44:1757–1764.

Dowdy AJ, Mills GA. 2012a. Atmospheric and fuel moisture characteristics associated with lightning-attributed fires. J Appl Meteor Climatol. 51:2025–2037.

Dowdy AJ, Mills GA. 2012b. Characteristics of lightning-attributed wildland fires in south-east Australia. Int J Wildland Fire. 21:521–524.

Franklin J, McCullough P, Cary C. 2000. Terrain variables used for predictive mapping of vegetation communities in southern California. In: Wilson JP, Gallant JC, editors. Terrain analysis: principles and applications. New York (NY): Wiley; p. 331–353.

Gibson RK, Bradstock RA, Penman T, Keith DA, Driscoll DA. 2015. Climatic, vegetation and edaphic influences on the probability of fire across Mediterranean woodlands of south-eastern Australia. J Biogeogr. 42:1750–1760.

Giglio L, Descoïtres J, Justice CO, Kaufman YJ. 2003. An enhanced contextual fire detection algorithm for MODIS. Remote Sens Environ. 87:273–282.

Gravelcic NJ, Nelson TA, Wulder MA. 2012. Spatial and temporal patterns of wildfire ignitions in Canada from 1980 to 2006. Int J Wildland Fire. 21:230–242.

Gumming SG. 2001. Forest type and wildfire in the Alberta boreal mixedwood: what do fires burn? Ecol Appl. 11:97–110.

Hantson S, Padilla M, Corti D, Chuvieco E. 2013. Strengths and weaknesses of MODIS hotspots to characterize global fire occurrence. Remote Sens Environ. 131:152–159.

Hardy CC, Burgan RE. 1999. Evaluation of NDVI for monitoring live moisture in three vegetation types of the western United States. Photogramm Eng Remote Sens. 65:603–610.

Hawbaker TJ, Radeloff VC, Stewart SI, Hamer RB, Keuler NS, Clayton MK. 2013. Human and biophysical influences on fire occurrence in the United States. Ecol Appl. 23:565–582.

Hawbaker TJ, Radeloff VC, Syphard AD, Zhu Z, Stewart SI. 2008. Detection rates of the MODIS active fire product in the United States. Remote Sens Environ. 112:2656–2664.
Heyerdahl EK, Brubaker LB, Agee JK. 2001. Spatial controls of historical fire regimes: a multiscale example from the Interior West, USA. Ecology. 82:660—678.

Justice CO, Townsend JRG, Vermote EF, Masuoka E, Wolfe RE, Saleous N, Roy DP, Morisette JT. 2002. An overview of MODIS Land data processing and product status. Remote Sens Environ. 83:3—15.

Krusel N, Packham D, Tapper N. 1993. Wildfire activity in the Mallee Shrubland of Victoria, Australia. Int J Wildland Fire. 3:217—227.

LeDell E, Petersen M, van der Laan M. 2014. cvAUC: Cross-validated area under the ROC curve confidence intervals [Computer software]. Version 1.1.0. LeDell E; [updated 2014 Dec 7; cited 2015 Jul 8]. Available from: https://github.com/ledell/cvAUC

Magnussen S, Taylor SW. 2012. Prediction of daily lightning- and human-caused fires in British Columbia. Int J Wildland Fire. 21:342—356.

Martínez-Fernández J, Chuvieco E, Koutsias N. 2013. Modelling long-term fire frequency factors in Spain by accounting for local variations with geographically weighted regression. Natl Hazards Earth Syst Sci. 13:311—327.

McRae RHD. 1992. Prediction of areas prone to lightning ignition. Int J Wildland Fire. 2:123—130.

Mermoz M, Kitzberger T, Veblen TT. 2005. Landscape influences on occurrence and spread of wildfires in Patagonian forests and shrublands. Ecology. 86:2705—2715.

Moreira S, Vaz P, Catry F, Silva JS. 2009. Regional variations in wildfire susceptibility of land-cover types in Portugal: implications for landscape management to minimize fire hazard. Int J Wildland Fire. 18:563—574.

Morissette JT, Giglio L, Csiszar I, Setzer A, Schroeder W, Morton D, Justice CO. 2005. Validation of MODIS active fire detection products derived from two algorithms. Earth Interact. 9:1—25.

Mouillot F, Ratte J-P, Joffre R, Moreno J, Rambal S. 2003. Some determinants of the spatio-temporal fire cycle in a Mediterranean landscape (Corsica, France). Landsc Ecol. 18:665—674.

Murphy BP, Bradstock RA, Boer MM, Carter J, Cary GJ, Cochrane MA, Fensham RJ, Russell-Smith J, Williamson GJ, Bowman DMJS. 2013. Fire regimes of Australia: a pyrogeographic model system. J Biogeogr. 40:1048—1058.

Nakazawa M. 2014. fmsb: Functions for medical statistics book with some demographic data [Computer software]. Version 0.5.1; [updated 2014 Sep 15; cited 2015 Jul 8]. Available from: http://CRAN.R-project.org/package=fmsb

Neis P, Zielstra D, Zipf A. 2011. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007—2011. Future Internet. 4:1—21.

Oliveira S, Moreira F, Boca R, San-Miguel-Ayan J, Pereira JMC. 2014. Assessment of fire selectivity in relation to land cover and topography: a comparison between Southern European countries. Int J Wildland Fire. 23:620—630.

Oliveira S, Oehler F, San-Miguel-Ayan J, Camia A, Pereira JM. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. For Ecol Manage. 275:117—129.

Padilla M, Vega-García C. 2011. On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human-caused fire occurrences in Spain. Int J Wildland Fire. 20:46—58.

Parisien MA, Moritz MA. 2009. Environmental controls on the distribution of wildfire at multiple spatial scales. Ecol Monogr. 79:127—154.

Pausas JG, Keeley JE. 2009. A burning story: the role of fire in the history of life. Bioscience. 59:593—601.

Penman TD, Bradstock RA, Price O. 2013. Modelling the determinants of ignition in the Sydney Basin, Australia: implications for future management. Int J Wildland Fire. 22:469—478.

Pew KL, Larsen CPS. 2001. GIS analysis of spatial and temporal patterns of human-caused wildfires in the temperate rain forest of Vancouver Island, Canada. For Ecol Manage. 140:1—18.

Plucinski MP. 2011. A review of wildfire occurrence research. Melbourne: BCR Centre.

Plucinski MP. 2014. The timing of vegetation fire occurrence in a human landscape. Fire Saf J. 67:42—52.

Plucinski MP, McCaw WL, Gould JS, Wotton BM. 2014. Predicting the number of daily human-caused bushfires to assist suppression planning in south-west Western Australia. Int J Wildland Fire. 23:520—531.

Preisler HK, Brillinger DR, Burgan RE, Benoît JW. 2004. Probability based models for estimation of wildfire risk. Int J Wildland Fire. 13:133—142.

Preisler HK, Westerling AL, Gebert KM, Munoz-Arriola F, Holmes TP. 2011. Spatially explicit forecasts of large wildland fire probability and suppression costs for California. Int J Wildland Fire. 20:508—517.

Pyne SJ, Andrews PL, Laven RD. 1996. Introduction to wildland fire. 2nd ed. New York: Wiley.

R Development Core Team. 2014. R: A language and environment for statistical computing [Computer software]. Version 3.1.1. Vienna: R Foundation for Statistical Computing; [updated 2014 Jul 10; cited 2014 Dec 10]. Available from: http://www.R-project.org/

Renard Q, Pélissier R, Ramesh BR, Kodandapani N. 2012. Environmental susceptibility model for predicting forest fire occurrence in the Western Ghats of India. Int J Wildland Fire. 21:368—379.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 12:1.

Rodrigues M, de la Riva J. 2014. An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environ Model Software. 57:192—201.
Rodrigues M, de la Riva J, Fotheringham S. 2014. Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. Appl Geogr. 48:52–63.

Romero-Calcerrada R, Barrio-Parra F, Millington JDA, Novillo CJ. 2010. Spatial modelling of socioeconomic data to understand patterns of human-caused wildfire ignition risk in the SW of Madrid (central Spain). Ecol Model. 221:34–45.

Romero-Calcerrada R, Novillo CJ, Millington JDA, Gomez-Jimenez I. 2008. GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (Central Spain). Landsc Ecol. 23:341–354.

Rothermel RC. 1983. How to predict the spread and intensity of forest and range fires. Ogden (UT): USDA Forest Service, Intermountain Forest and Range Experiment Station.

Russell-Smith J, Yates CP, Whitehead PJ, Smith R, Craig R, Allan GE, Thackway R, Frakes I, Criddle S, Meyer MCP, et al. 2007. Bushfires ‘down under’: patterns and implications of contemporary Australian landscape burning. Int J Wildland Fire. 16:361–377.

Schroeder W, Prins E, Giglio L, Csiszar I, Schmidt C, Morisette J, Morton D. 2008. Validation of GOES and MODIS active fire detection products using ASTER and ETM+ data. Remote Sens Environ. 112:2711–2726.

Sellers PJ. 1985. Canopy reflectance, photosynthesis and transpiration. Int J Wildland Fire. 6:1335–1372.

Syphard AD, Radeloff VC, Keuler NS, Taylor RS, Hawbaker TJ, Stewart SI, Clayton MK. 2008. Predicting spatial patterns of fire on a southern California landscape. Int J Wildland Fire. 17:602–613.

Taylor SW, Woolford DG, Dean CB, Martell DL. 2013. Wildfire prediction to inform fire management: statistical science challenges. Stat Sci. 28:586–615.

Vasconcelos MJP, Silva S, Tome M, Alvim M, Pereira JMC. 2001. Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. Photogramm Eng Remote Sens. 67:73–81.

Venables WN, Ripley BD. 1999. Modern applied statistics with S-PLUS. 3rd ed. New York: Springer.

Vilar L, Woolford DG, Martell DL, Martin MP. 2010. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. Int J Wildland Fire. 19:325–337.

Wang Y, Anderson KR. 2010. An evaluation of spatial and temporal patterns of lightning- and human-caused forest fires in Alberta, Canada, 1980–2007. Int J Wildland Fire. 19:1059–1072.

Willis M. 2005. Bushfires—how can we avoid the unavoidable? Global Environ Change B Environ Hazards. 6:93–99.

Wintle BA, Elith J, Potts JM. 2005. Fauna habitat modelling and mapping: a review and case study in the Lower Hunter Central Coast region of NSW. Austral Ecol. 30:719–738.

Yulianti N, Hayasaka H, Sepriando A. 2013. Recent trends of fire occurrence in Sumatra (analysis using MODIS hotspot data): a comparison with fire occurrence in Kalimantan. Open J For. 3:129–137.