Applied Geography 66 (2016) 12–21

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Wildfire ignition in the forests of southeast China: Identifying drivers and spatial distribution to predict wildfire likelihood

Futao Guo [a], Zhangwen Su [a], Guangyu Wang [b, *], Long Sun [c], Fangfang Lin [d], Aiqin Liu [a, **]

[a] College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, 350002, China
[b] Sustainable Forest Management Laboratory, Faculty of Forestry, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[c] College of Forestry, Northeast Forestry University, Harbin, 150040, China
[d] College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

ARTICLE INFO

Article history:
Received 25 May 2015
Received in revised form
29 October 2015
Accepted 20 November 2015
Available online xxx

Keywords:
Human-caused fire
Fire occurrence
Spatial distribution
Drivers of forest fire
Logistic regression

ABSTRACT

Understanding the spatial distribution and driving factors of forest fire facilitates local forest fire management planning and optimization of resource allocation for fire prevention geographically. In this study, we analyzed the spatial pattern and drivers of forest fire in Fujian province, southeastern China, during 2000–2008 using Ripley's *K*-function and logistic regression (LR) model. The likelihood of fire occurrence was mapped based on the resultant model. The data regarding fire ignitions, weather conditions, vegetation, topography, infrastructure, and socioeconomic factors were extracted from ArcGIS environment. The study revealed that fire ignition was mainly clustered in space due to the comprehensive influence of different factors. Elevation, daily precipitation, and daily relative humidity were negatively associated with fire ignitions, whereas distance to settlement, population density, and per capita gross domestic product (GDP) impacted fire occurrence positively. The spatial distribution of fire occurrence likelihood was highly variable in Fujian: high fire likelihood was prevalent in the northern and southeastern parts of Fujian, whereas it was relatively low in the western province. Fire risk may be underestimated in some areas of Fujian according to the spatial patterns of the model residual, which should be paid more attention to in the forest fire management practice.

1. Introduction

Forest fire is an important ecological factor, which has a significant impact on forest regeneration and succession (Podur, Martell, & Csillag, 2003), and from economic and safety perspectives, results in a loss of forest resources and threatens the safety of human life and property (Flannigan, Stocks, & Wotton, 2000). Forest fires mainly fall into two categories: human-ignited or *anthropogenic* fires, versus fires that are not a direct consequence of human action – *naturally induced* fire. Worldwide, human activities are responsible for most wildfire ignitions – for example, more than 95% of all fires in southern Europe (San-Miguel-Ayanz & Camiá, 2009), and 60% in Alaska over the period of 1950–2005 were anthropogenic (Todd & Jewkes, 2006). The causes of forest fires differ between South and North China: in Daxing'an Mountains, in the north, fires were identified as originating equally from humans as by lightning (Guo et al. 2015), while in southern regions, the majority of fires were attributable to human activity. In the southeastern province of Fujian, human-caused forest fires reached 95% in the past decade (He, Liu, Zhao, & Zhou, 2013).

Understanding spatial distribution and primary factors that influence fire occurrence is crucial for forest management and allocation of fire prevention and suppression resources. For example, fire towers, inspection stations, fire patrols and firebreaks should be allocated around fire-prone zones, which can reduce economic expense and improve the efficiency of forest fire management. In the past decades, numerous studies have been conducted to identify spatial patterns and drivers of fire occurrence (Hu & Zhou, 2014; Martínez, Vega-Garcia, & Chuvieco, 2009; Syphard et al. 2008; Zhang, Zhang, Li, Xu, & Zhou, 2013). This early research tended to consider primarily meteorological factors. More recent research has begun to include a comprehensive analysis of vegetation, terrain, human activity, socioeconomic influences, and other biophysical and ecological factors (Chas-Amil, Prestemon, McClean,

* Corresponding author.
** Corresponding author.
E-mail addresses: guofutao@126.com (F. Guo), 463781229@qq.com (Z. Su), guangyu.wang@ubc.ca (G. Wang), sunlong365@126.com (L. Sun), 873410288@qq.com (F. Lin), fjlaq@126.com (A. Liu).

http://dx.doi.org/10.1016/j.apgeog.2015.11.014

& Touza, 2015; Fry & Stephens, 2006; Romero-Calcerrada, Barrio-Parra, Millington, & Novillo, 2010). Findings from these studies have revealed the essentiality of considering different types of potential influences, in order to identify key variables or driving factors (Catry, Rego, Bação, & Moreira, 2009; Loboda & Csiszar, 2007; Martínez et al. 2009; Nunes et al. 2005; Syphard et al. 2008).

Forest fire prediction research is a developing field in China, and has focused mainly on the Chinese boreal forest in northern China. Fujian, located southeast of China, is one of China's four major forest management regions, ranking the highest in terms of forest coverage. It is also an area with a high annual forest fire incidence. Although increased fire prevention efforts have reduced the number of annual forest fires in recent years in Fujian, the total area of forest burned has increased (He et al. 2013). Despite these increases, compared to the northern forest regions in China, studies focused on forest fire drivers and forecasting in Fujian province are insufficient to inform forest fire management in this region. For example, only a few variables have been used so far to perform fire danger classification and fire forecasting, and methods of analysis have not been especially sophisticated (He et al. 2013), which may miss out on important nuances, such as anthropogenic influences or interactions among variables. Research has shown that forest fire occurrence, especially anthropogenic forest fire, was affected by many factors (Oliveira, Oehler, San-Miguel-Ayanz, Camia, & Pereira, 2012; Zhang, Zhang, & Zhou, 2010) in which socioeconomic indicators and human activity were found to be indispensable considerations; nonetheless, these indicators have not been considered in the fire-prediction studies of Fujian. In the past decades, forest-related socio-economic activities, including tourism, have become potentially meaningful influences in Fujian due to the abundance of forest resources and increasing interest in these types of activities. This has the potential to increase the complexity of relationships between fire occurrence and local factors affecting risk of ignition, as well as the unique spatial distribution of fire occurrence.

The objectives of the present research are to (1) identify the spatial distribution of fire ignitions in Fujian, China, (2) understand the comprehensive and individual effects of ignition factors on fire occurrence, and (3) produce spatially explicit statistical models and maps predicting patterns of fire ignitions in Fujian, China, using a combination of biophysical and human variables. Results can provide the necessary guidance for local forest fire management in terms of fire resource allocation, reducing the economic burden of fighting fires, and improving the efficiency of forest management strategies in the forests of southeastern China. Findings from this case study also have the potential to be implemented in other areas of southeastern China, which have many shared variables, such as fire frequency, climate conditions, forest resources, and socioeconomic factors.

2. Materials and methods

2.1. Study area

Fujian is a province in southeastern China (Fig. 1a). The total land area of Fujian is 124,000 km^2, which accounts for 1.3% of China's total land area. The climate of Fujian is warm, humid subtropical monsoon, which is affected by the monsoon circulation and topography. Average annual rainfall is 1400–2000 mm, and average temperature 17–21 °C. The current forest coverage of Fujian is around 66%. Dominant tree species include *Pinus massoniana* Lamb., *Cunninghamia lanceolata*, *Casuarina equisetifolia* L., *Phyllostachys heterocycla*, and others.

Fujian has a relatively high forest fire frequency compared to other regions in China that have high forest coverage such as Chinese boreal forest. The fire season is from approximately September 15 until April 30 of the following calendar year. From 1951 to 1998, forest fires occurred on average 1385 times annually and more than 95% fires are caused by human activities (Zheng et al. 2001).

2.2. Spatial distribution analysis

K-function proposed by Ripley (1976) is a useful tool to describe how the interaction or spatial dependence between events varies through space. Ripley's *K*-function is defined as follows:

$$K(d) = \frac{1}{\lambda} E\,(\text{number of other events within d distance of an arbitrary event})$$

where λ is the density (number per unit area) of events, and $E(\bullet)$ is the expectation operator. It has been widely used in spatial point pattern analysis and spatial point process modeling (Dissing & Verbyla, 2003; Podur et al. 2003). Theoretically, for a homogeneous Poisson process, known as "complete spatial randomness" (CSR), $K(d) = \pi d^2$. For $d \geq 0$, Ripley's *K*-function can be used as a formal statistic to test the null hypothesis of CSR. The values of $K(d)$ less than πd^2 indicate regularity, whereas aggregation is indicated when $K(d)$ is greater than πd^2. There are three basic edge correction methods for Ripley's *K*-function and we used "The guard area correction" in this study. SpPack software was used to perform the *K*-function (Perry, 2004), and confidence envelopes were set to 95%, based on 499 replicates.

2.3. LR model

In recent years, many scholars have used LR to predict and analyze forest fire occurrence (Chang et al. 2013; Martínez et al. 2009; Oliveira et al. 2012; Rodrigues, de al Riva, & Fotheringham, 2014; Saefuddin, Setiabudi, & Fitrianto, 2012; Vega Garcia, Woodard, Titus, Adamowicz, & Lee, 1995). In the analysis, forest fire occurrence was assigned a value of 1 ($y = 1$), while "zero occurrence" was 0 ($y = 0$). Furthermore, we assumed that the probability of occurrence of forest fire ($y = 1$) was *P*, and the probability of no forest fires ($y = 0$) was $(1 - P)$. This allowed us to use LR to model the probability of occurrence of forest fire in association with each variable. The specific expression was

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m \tag{1}$$

The formula for estimating the probability of forest fire occurrence converted using *Logit* was

$$p = 1\Big/\left(1 + e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_m x_m)}\right) \tag{2}$$

In Eq. (2), *P* is the probability of forest fire occurrence, *m* is the number of covariates, is $(\beta_1, \beta_2, \ldots \beta_{\rm m})$ is the correlation coefficients for each variable using the LR model, and $(x_1, x_2, \ldots, x_m)$ are the respective variables which influenced the occurrence of forest fires.

2.3.1. Dependent variable

Binomial LR model requires that the data are in a binomial distribution. A certain percentage of random points (non-fire points) were created to satisfy the requirements of the binomial LR model. The forest fire data used in this study were Fujian 2000–2008 satellite fire point data provided by the Forestry Science Data Center (http://www.cfsdc.org/indexAction.action?classId=1). There were 13,185 forest fires that occurred in Fujian during 2000–2008. Data points also provided the geographic coordinates, time, and other information of the forest fires. There was

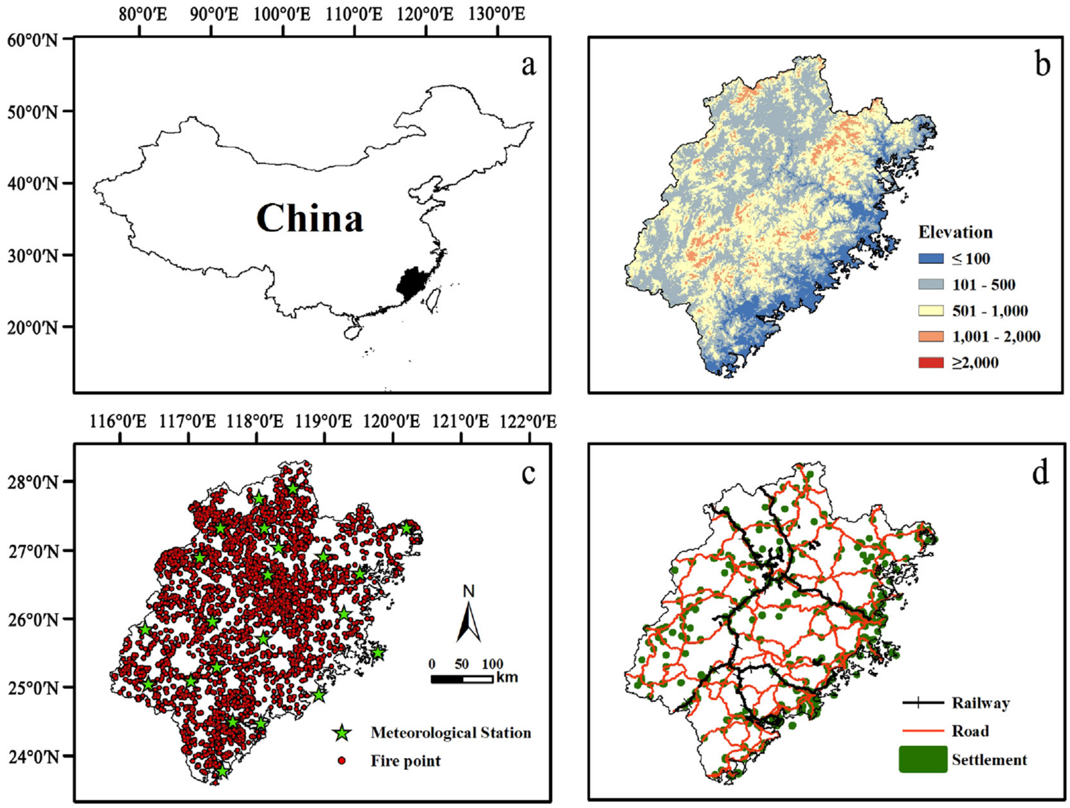

Fig. 1. Study area showing the bounds of Fujian province in China (a); elevation (b); fire points (ignitions) and meteorological stations (c); and railway, road, and settlement (d).

no uniform requirement for the number of random points selected. We randomly generated 14,965 random points, about the same number as the ignition points.

2.3.2. Independent variables

In this paper, independent variables included five different aspects: topography, vegetation, weather, infrastructure, and social and economic data. Detailed descriptions for each variable are provided in Table 1.

2.3.2.1. Topography. Topography data included elevation, slope, and aspect, extracted from a 1:50,000 digital elevation model (DEM) dataset, which was built in 2002 (http://www.gscloud.cn/). We extracted elevation and slope values for each point from the DEM dataset using ArcGIS. Since aspect was a character (*i.e*., not numerical), these data could not be directly used to fit model; consequently, we transformed these data using the following steps: (1) calculate the proportion of each aspect in the total aspect; (2) determine the aspect of each point (fire or non-fire point); and (3) use this value (*i.e*., proportion of aspect) to represent the aspect to which each point (fire or non-fire point) belongs (Chang et al. 2013).

2.3.2.2. Vegetation data. The vegetation data for this study were based on a vegetation structure and type map from the Cold and Arid Regions Environmental and Engineering Research Institute of Chinese Academy of Sciences (http://westdc.westgis.ac.cn/). The data were gathered in 2000, and spatial resolution was reported to be 1 km. Using this map, we grouped polygons of vegetation into four categories: Needleleaf evergreen trees (40.9% cover of Fujian province), broadleaf evergreen tree (13.2% cover), broadleaf deciduous shrub (11.8% cover), grass, and crop (34.1% cover). The forest type distributions are shown in Fig. 2.

Forest type is also character data. Vegetation extraction steps were similar to those used to transform aspect data: (1) calculate the proportion of plant type in the total plant types; (2) determine the plant type of each point (fire or non-fire point); (3) use this value (*i.e*., proportion of plant types) to represent the plant type to which each point (fire or non-fire point) belongs.

2.3.2.3. Meteorological data. The meteorological data were derived from the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn/), and covered daily meteorological data from 22 national meteorological stations (Fig. 1c) in the Fujian province. Daily weather data included 22 meteorological factors that were preprocessed to remove missing data points (*e.g*., due to equipment damage), and the remaining factors included average surface temperature (°C), daily maximum surface temperature (°C), average wind speed (m·s−1), precipitation (mm/24 h), sunshine hours, average temperature (°C), minimum relative humidity (%), and others (Table 1), for a total of 15 meteorological variables. The “Extract Values to Points” function in ArcGIS was used to extract meteorological data to fire and random points.

2.3.2.4. Infrastructure and salient geographic features. Geographic distribution of infrastructure and salient geographic features were taken from the National Administration of Surveying, Mapping and Geoinformation of China (http://218.244.250.78/NgccDigitalHall/). The data were collected in 2000, based on a 1:250,000 precision vector map. Data included spatial distribution of railways, roads, rivers, settlements, and other anthropogenic landscape structures. We converted infrastructure data using Arcgis10.0, and then calculated the Euclidean distance from each point (fire and random) to all the types of infrastructure, including the distance from each point to railways, roads, rivers, residential areas, and other infrastructure within the study area.

Table 1
Predictor variables included in forest fire model development for the Fujian forest region.

Variable type	Variable name	Code	Resolution/scale	Description	Source/reference
Topographic	Elevation	Elev.	Raster/25 m	Elevation of each fire point and control extracted from a raster map of study area	<National Administration of Surveying, Mapping and Geoinformation of China>, 2002
	Slope	Slope		Slope of each fire point and control extracted from a raster map of study area	
	Aspect	Aspect		Proportion of each aspect class (N,NE, E, SE, S, SW, W, NW) in the study area	
Vegetation	Forest type	Forest type	Raster/1 km	Proportion of each forest type in the study area	The Cold and Arid Regions Science Data Center, China (http://wesdc.westgis.ac.cn/)
Climatic	Daily precipitation	Da_preci	Daily/0.01	Corresponding daily climate factors of each fire point and control point based on five national weather stations	China Meteorological Data and Sharing Network (http://cdc.cma.gov.cn/)
	Sunshine hours	SSD			
	Daily mean wind speed	Da_wind			
	Daily maximum wind speed	Da_maxwind			
	Daily average ground surface temperature	GST_avg			
	Daily maximum ground surface temperature	GST_max			
	Daily minimum ground surface temperature	GST_min			
	Daily mean site pressure	Da_spre			
	Daily maximum site pressure	Da_maxspre			
	Daily minimum site pressure	Da_minspre			
	Daily mean temperature	Da_temp			
	Daily mean relative humidity	Da_RH			
	Daily minimum relative humidity	Da_minRH			
	Daily maximum temperature	Da_maxtemp			
	Daily minimum temperature	Da_mintemp			
Infrastructure	Distance to the nearest railway	Dis_railway	Vector/1:100,000	The straight distance between a fire point or a control point and the nearest railway	<National Administration of Surveying, Mapping and Geoinformation of China>2002
	Distance to the nearest river	Dis_river		The straight distance between a fire point or a control point and the nearest river	
	Distance to the nearest road	Dis_road		The straight distance between a fire point or a control point and the nearest road	
	Distance to the nearest settlement	Dis_sett		The straight distance between a fire point or a control point and the nearest settlement	
Socioeconomic	Per capita GDP	CGDP	Grid/1 km	Per capita GDP of the study area	Data Sharing Infrastructure of Earth System Science (http://www.geodata.cn/Portal/index.jsp), 2010
	Density of population	Den_Pop		The annual population density of the study area	

2.3.2.5. Demographic and socioeconomic data. Gridded demographic and socioeconomic data were taken from the Data Sharing Infrastructure of Earth System Science (http://www.geodata.cn/Portal/index.jsp) with 1-km resolution, including the grid population density and per capita GDP (CGDP) from 2000, 2003, 2005, and 2010. In addition, we calculated the average population and GDP growth rates in the intervals of 2000–2003, 2003–2005, and 2005–2010 and generated grid data for the annual population and GDP from 2000 to 2008. These data were then correlated with fire point and random points, using the ArcGIS "Raster Extraction tool" (Oliveira et al. 2012).

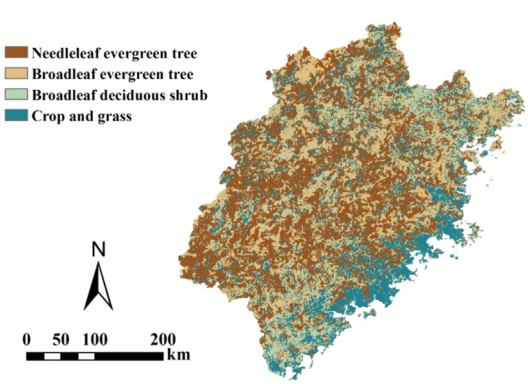

Fig. 2. The distribution of forest types in Fujian province.

2.3.3. Test for multicollinearity

Multicollinearity refers to the correlation between each explanatory variable in a linear regression model. This correlation may distort the model estimation or interfere with accurate estimation. This study used VIF (variance inflation factor) to test for multicollinearity, and variables with significant collinearity were gradually removed. Generally, this study considered VIF equal to 10 as the benchmark. If the VIF was >10, this indicated significant collinearity between variables which were gradually removed (Wu & Zhang, 2013).

2.4. Calibration of LR model

In order to evaluate model performance, we used receiver operator characteristic (ROC) curve test method; this has been used to evaluate the goodness of fit of LR models by other scholars (Chang et al. 2013; Del Hoyo, Isabel, & Vega, 2011). The evaluation is based on the value of area under the curve (AUC). Typically, the AUC should range from 0.5 to 1: higher AUC values indicate a better goodness of fit of the regression model. It is generally suggested that AUC equal to 0.5 indicates a totally random prediction, AUC in the range of 0.5–0.7 poor fit, and AUC of 0.7–0.9 normal fit, while an AUC of 0.9–1 indicates high goodness of fit. At present, the ROC test method is applied to forest fire prediction (Del Hoyo et al. 2011; Jiménez-Valverde et al. 2012).

In LR, judgment about the cutoff point is critical to estimate the value of the model's predicted probability. Many studies in the past used the system default value of 0.5 (Deng, Li, Feng, & Zhang, 2012); however, in recent years, some scholars pointed out that this method might result in large deviation (Chang et al. 2013). In an effort to address this problem, scholars began using sensitivity and specificity of ROC in the calculation of "Youden index (sensitivity + specificity −1)," then judging the "cut-off point" (best cutoff value) to classify probability. This method has also been used to predict forest fire (Martínez et al. 2009). If the predicted probability of the model is higher than the cutoff point, it is predicted that forest fire will occur; however, a lesser value indicates no forest fire.

2.5. Selection of model variables

In this study, all 28,150 fires and random non-fire points (Sect. 2.4.1) were assigned either to a validation set (60%) or a calibration set (40%). In order to reduce the influence of a random division of samples on the selection of model parameters (variables), the division and model fitting were performed three times. Three intermediate models were generated. Variables that were significant in at least two of the three intermediate models were selected to be used in analysis of the complete dataset (Oliveira et al. 2012; Rodrigues et al. 2014).

2.6. Mapping the likelihood of fire occurrence and fire risk

The probability of forest fire and random points were predicted using Eq. (3), and maps of fire occurrence likelihood were created using the Kriging method using ArcGIS, based on the predicted and residual values in the LR model. In addition, the map of fire risk was categorized into three classes based on the fire occurrence likelihood and the cutoff value (0.404) of a complete dataset: low (0–0.404), medium (0.404–0.5), and high (>0.5) (Chang et al.2013).

3. Results

3.1. Spatial analysis

We computed a *K*-function for each year from 2000 to 2008, and also the combined fire ignitions of 9 years (Fig. 3 and Fig. 4). As we can see, there is a clear departure from CSR toward clustering, in most years between 2000 and 2008, and across the 9 years of the dataset. The amount of clustering for a given year changed also as a function of distance – clustering started at 0.1° (~10 km), which suggests that there were more fires within any specified distance (>10 km) than expected under complete spatial randomness, where an identified fire is the starting point. The degree of clustering seemed to increase with radius. For the years 2004 and 2007, clustering ended at 0.6 and 1° (~60 and 100 km), respectively, and tended toward more regular distribution after their clustering-end distances. Fire ignitions in 2008 were regularly distributed in space at varying distances.

3.2. Test for multicollinearity

The multicollinearity test selected a total of 18 variables for model fitting, including ten non-meteorological variables ("distance to railway," "distance to road," "distance to river," "distance to settlement," "elevation," "slope," "aspect," "plant functional types," "population density," and "per capita GDP") and eight meteorological variables ("daily maximum globe surface temperature," "daily minimum surface temperature," "daily maximum wind speed," "20:00–20:00 precipitation," "daily maximum air pressure," "sunshine hours," "daily average relative humidity," and "daily minimum relative humidity").

3.3. Analysis of model fitting results

3.3.1. Selection of model parameters

Table 2 shows that 11 variables in the intermediate models met the stated requirements (significant in at least two of the three intermediate models) and were used in the final stage of model development.

3.3.2. Evaluation of model performance

Fig. 5 shows ROC curves for the three subsamples and the complete sample size. The AUC values (*i.e.*, area under the ROC curve) were between 0.8 and 0.9 ($p < 0.0001$). This suggests an acceptable goodness of fit (Liu & Yang, 2013), such that the model is a good candidate for predicting forest fire occurrence in Fujian Province. According to the selected variables in the model, combined with cutoff values for the three subsamples (0.445, 0.420, and 0.442), results showed that the predictive accuracy of the three models derived from the three subsamples was higher and was close to each other (Table 3). These variables were then used to assess model fit for all samples. Result of ROC testing showed an AUC value of 0.843, indicating a high goodness of fit when the full dataset was tested using our test model. The associated cutoff value – 0.404 – was then used to guide classification (*i.e.*, fire occurrence/nonoccurrence), and the predictive accuracy of the final prediction model using all sample data was found to be 76.4% (Table 3, Fig. 5).

3.3.3. Spatial distribution of fire ignition likelihood

Based on the final model of parameter estimation using LR (Table 4), the probability model for forest fire in Fujian province was calculated as

$$p = 1\Big/\left(1 + e^{-(-0.1389-0.0019x_1+0.2937x_2+0.0449x_3+0.0058x_4-0.0173x_5-0.0053x_6+0.0132x_7-0.0148x_8+0.0001x_9+0.0001x_{10})}\right) \tag{3}$$

The various '*x*' variables represent elevation (*x*1), distance to road (*x*2), distance to settlement (*x*3), maximum ground surface temperature (*x*4), daily minimum globe surface temperature (*x*5), daily precipitation (*x*6), sunshine hours (*x*7), daily relative humidity (*x*8), population density (x9), and per capita GDP (CGDP) (*x*10).

The fire ignition likelihood map (Fig. 6a) indicated that areas with high probability of forest fire were distributed throughout most of Fujian, but were especially concentrated in the north and northeast, covering most of the municipal areas of Nanping, Fuzhou, and Zhangzhou, and at the junctions of Sanming, Ningde, Fuzhou, and Nanping. The model showed that the western area of Fujian had a low probability of forest fire occurrence. The map of fire risk (Fig. 6b) was also generated based on fire occurrence likelihood and cutoff values. The distribution of high-risk zones was similar to Fig. 6a, which indicates that high and medium fire risk areas were evenly distributed throughout the north and northeast

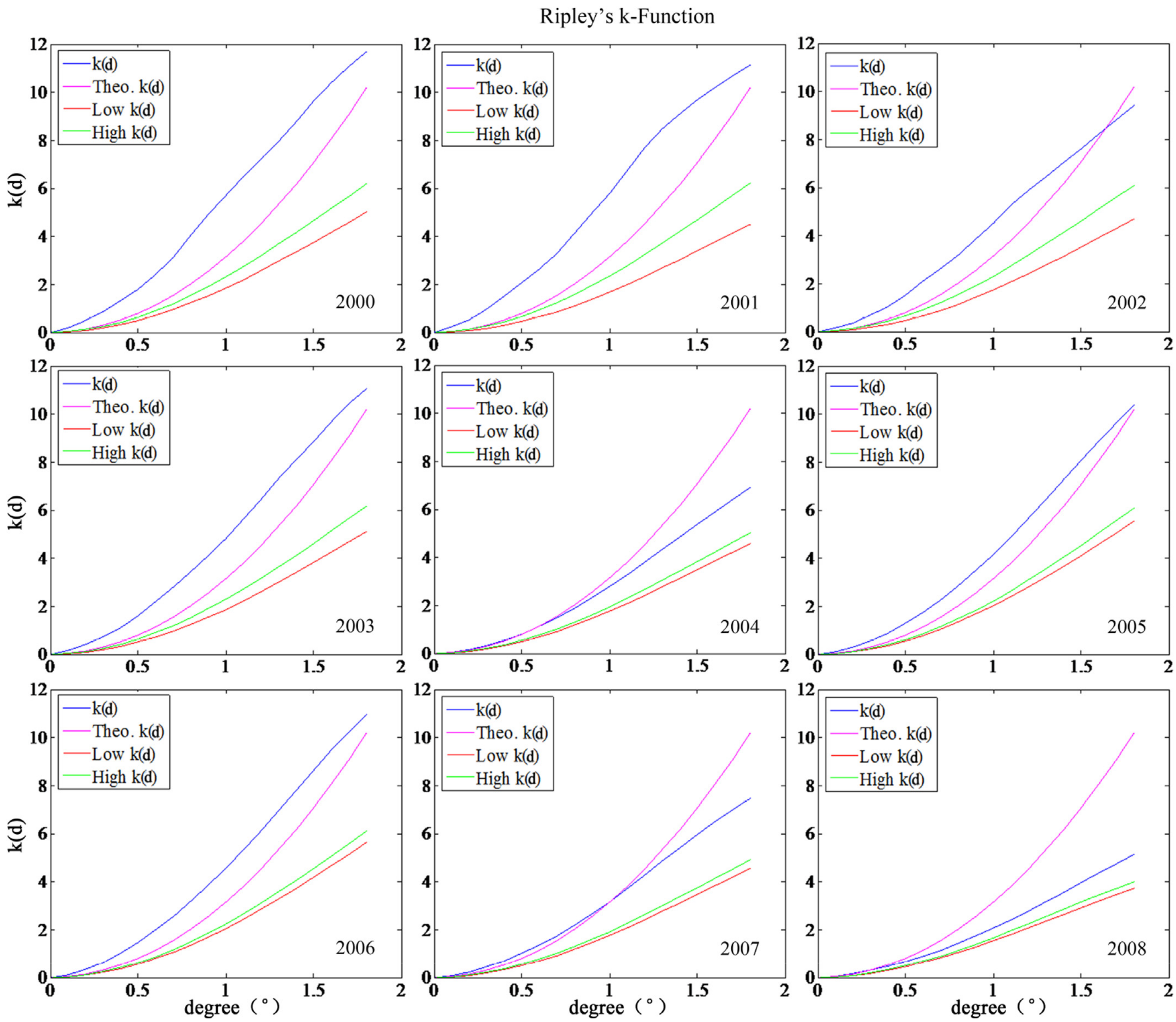

Fig. 3. Calculated Ripley's *K*-functions for the density of fire ignitions each year from 2000 to 2008 in the study area (Fujian, China), compared with the theoretical Ripley's *K*-function, representing complete spatial randomness (CSR). The pink line represents the empirical *K*-function under CSR with green and red lines, which are 95% confidence envelopes. X-axis represents the distance to the nearest ignition from a given ignition point and the unit is degree. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

of Fujian. The cities of Nanping, Zhangzhou, and Putian had areas of high and medium fire risk.

In order to better evaluate the goodness-of-fit model and compare the predicted probability with the observed values in our model, model residuals and residual plots were generated, assessed, and mapped (Fig. 7). We found that the model overestimated forest fire occurrence rate in the northeastern Fujian, and some areas of southwestern Fujian, while it underestimated the probability of forest fire occurrence in most areas of Nanping, Sanming, and western Zhangzhou, and the middle area of Quanzhou.

4. Discussion

The fire ignition pattern in our study area was mainly clustered in space, which is consistent with other previous studies (Genton, Butry, Gumpertz, & Prestemon, 2006; Miranda, Sturtevant, Stewart, & Hammer, 2012; Mundo, Wiegand, Kanagaraj, & Kitzberger, 2013). Our findings indicated that the spatial heterogeneity (pattern) of fire ignition may be caused by the comprehensive effect of topography, human activity, and meteorological factors, which were consistent with other studies (Genton et al. 2006; Schoennagel, Veblen, & Romme, 2004). Proximity to transportation corridors (*e.g.*, roads) has been identified by others as a main driver of fire ignition (Butry & Prestemon, 2005). In contrast to Stephens (2005) and Romero-Calcerrada et al. (2010), railways were not identified as drivers in the present study. In this study, most railways were located in non-forest areas and, consequently, did not influence forest fire occurrence. Owing to the relative concentration and intensity of human activity at low altitudes, the likelihood of human-caused fire ignition was unsurprisingly higher (Chang et al. 2013; Miranda et al. 2012; Oliveira et al. 2012; Syphard

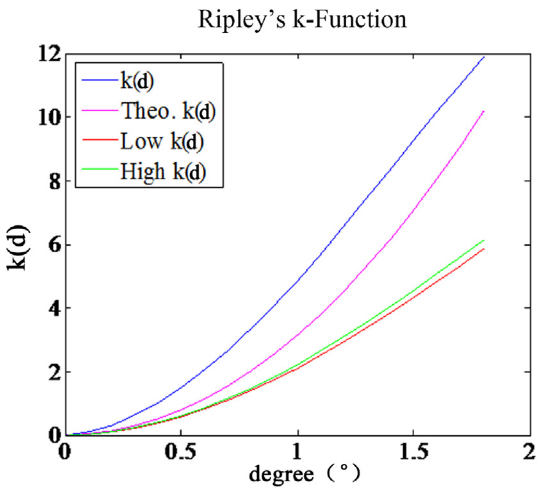

Fig. 4. Calculated Ripley's *K*-functions for the density of fire ignitions during 2000–2008 in the study area (Fujian, China), compared with the theoretical Ripley's *K*-function representing complete spatial randomness (CSR). The pink line represents the empirical *K*-function under CSR with green and red lines, which are 95% confidence envelopes. X-axis represents the distance to the nearest ignition from a given ignition point and the unit is degree. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

et al. 2008), The factors that can reflect the local socioeconomic activity such as population density, CGDP, and distance to settlement were found to be positively related to fire occurrence in Fujian, China. Our results are in agreement with other published works (Miranda et al. 2012; Syphard et al. 2007). Higher frequency of socio-economic activity associated with increased probability of fire occurrence in this study may indicate that in Fujian, many of the socio-economic activities may be forest-related. Forest type may not have emerged as crucial to fire occurrence because the species composition of forests in Fujian is relatively simple in space, which also agreed with the findings of other studies (Wu, He, Yang, Liu, & Liang, 2014).

Over the past decades, climatic factors have emerged as important for predicting fire ignition risk (Liu, Yang, Chang, Weisberg, & He, 2012; Preisler, Brillinger, Burgan, & Benoit, 2004; Syphard et al. 2008; Wotton, Martell, & Logan, 2003). In this study, we identified five important climatic drivers of fire ignition in Fujian: daily maximum temperature and sunshine hours were positively related to fire probability, while daily minimum temperature, daily precipitation, and daily mean relative humidity were found to be negatively correlated with fire occurrence. High temperatures and abundant sunshine are variables that likely influence each other, and either alone or in combination can contribute to increased evaporation from plants, as well as decrease moisture content of potential fire fuels (*e.g.*, downed woody material), leading to increased probability of fire occurrence (Chuvieco et al. 2004). Precipitation and relative humidity are two important factors for indicating the moisture of fuel in the forest. High precipitation and relative humidity contribute to fuel moisture, which in turn decreases the possibility of fire ignition (Zumbrunnen et al., 2011). It is worth noting that some variables such as temperature and relative humidity do not necessarily have a consistently linear relationship with fire occurrence (Castro, Tudela, & Sebastiá, 2003; Wu, He, Yang, & Liang, 2015); consequently, it is important that even where apparently linear relationships exist between certain variables and fire occurrence, unanticipated or subtle thresholds may be present and we must focus on discussions regarding the same.

Table 2
Significant ($p < 0.05$) variables selected by intermediate models using logistic regression. The direction of association between each predictor and dependent variable is also presented (+, positive association; −, negative association).

Variables	*p*-value min	*p*-value max	No. significant samples	Direction
Elev	<0.0001	<0.0001	3	−
Aspect	0.0201	0.1202	1	+
Slope	0.0360	0.1860	1	−
Forest type	0.0110	0.0850	1	+
Dis_road	<0.0001	<0.0001	3	−
Dis_river	0.0383	0.2130	1	+
Dis_sett	<0.0001	<0.0001	3	+
Da_maxtemp	<0.0001	<0.0001	3	+
Da_mintemp	<0.0001	<0.0001	3	−
Da_wind	0.0042	0.0931	1	−
Da_preci	<0.0001	<0.0001	3	−
SSD	<0.0001	<0.0001	3	+
Da_RH	<0.0001	<0.0001	3	−
Den_Pop	<0.0001	<0.0001	3	+
CGDP	<0.0001	<0.0001	3	+

Note: Elev is elevation; Dis_road is distance from ignition point to the nearest road; Dis_river is distance to the nearest river; Dis_sett is distance to the nearest settlement; Da_maxtemp is daily maximum globe surface temperature; Da_mintemp is daily minimum globe surface temperature; Da_wind is daily mean wind speed; Da_precipitation is daily precipitation; SSD is sunshine hours; Da_RH is daily mean relative humidity; Den-Pop is density of population; and CGDP is per capita GDP.

Fig. 6a illustrates that much of the province could be classified as having a high probability of forest fire, with exception of the western area of Fujian. Severe soil erosion may offer one likely explanation for the low probability of fire in the western area of Fujian (Chen, Chen, & Chen, 2011), whereby forest coverage was reduced and possibly even prevented from reestablishing. The fire risk map (Fig. 6b) shows that fire-risk zones, especially those with high fire risk, are distributed across administrative divisions; consequently, it would be more efficient and reasonable if fire prevention strategies and management plans were organized based on fire-risk zones, rather than on separate administrative divisions, as it is currently managed.

One potential caveat is that the distribution of model residuals indicates a degree of error in the model's predictive capacity that could be improved upon. The model may underestimate or overestimate the probability of occurrence of forest fires in Fujian spatially. The most likely source for model error is that the logistic model did not consider spatial correlation among fire points. Several studies have shown that the consideration of data spatial correlation could improve the predictive capacity of the model (Koutsias, Martínez-Fernández, & Allgöwer, 2010; Kupfer & Farris, 2007), and the spatial correlation should be considered in future studies. Particular attention should be paid to the areas where the fire risk has been underestimated during the fire season and more resources should allocated to high fire risk regions in order to improve the efficiency and level of fire prevention and management in Fujian, China.

5. Conclusions

The spatial patterns and likelihood of forest fire occurrence in Fujian province, China were investigated using geospatial information, weather conditions, and socioeconomic factors for the period 2000–2008. The computed model was used to map the likelihood of fire distribution within the study area. The findings showed that fire ignitions were mostly spatially clustered during the study period due to the comprehensive influence of elevation, weather conditions, infrastructure, and socioeconomic factors. The spatial and geographical variations among factors also lead to the spatial heterogeneity of the likelihood of fire occurrence. The forest

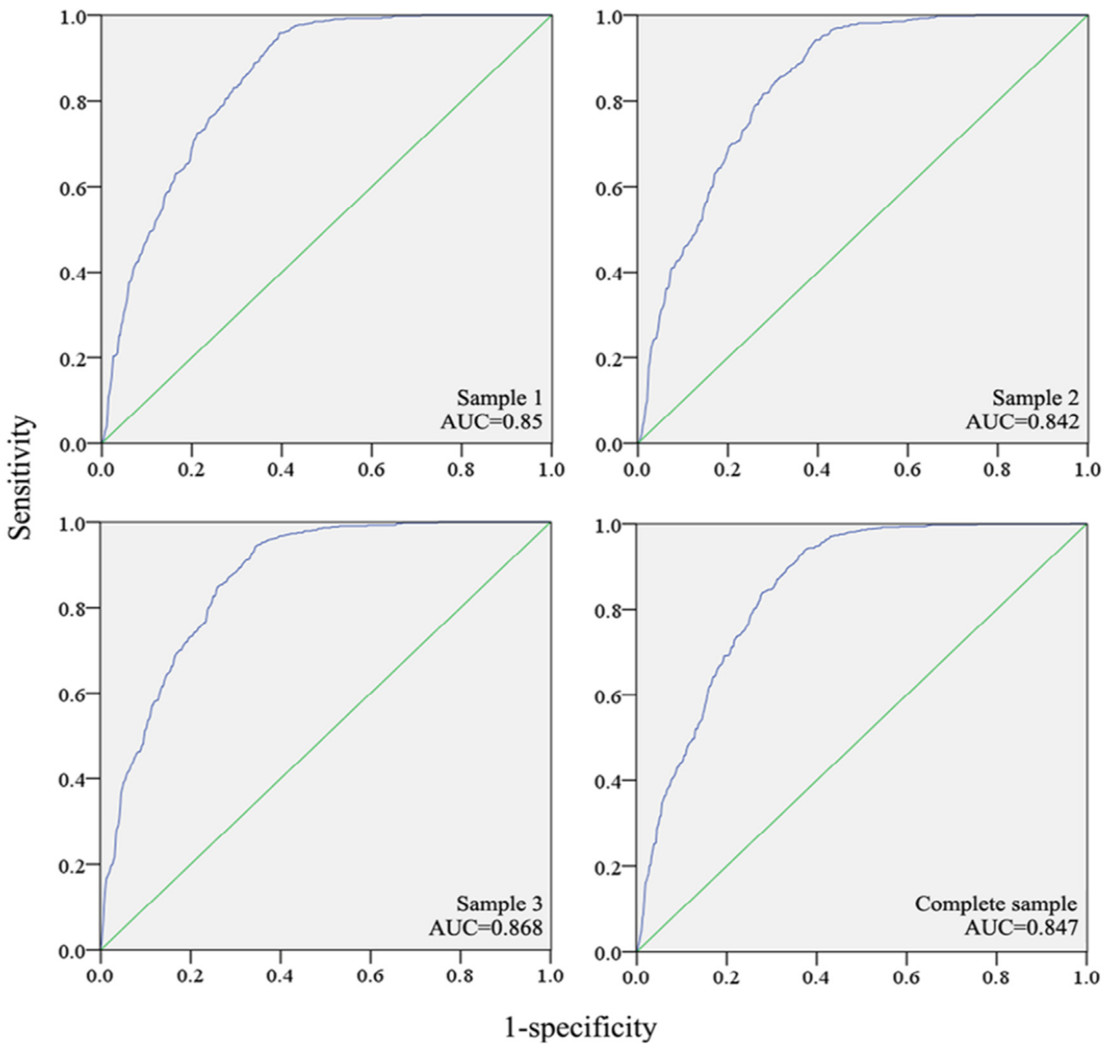

Fig. 5. Receiver operator characteristic (ROC) curves for logistic regression model fitting for three subsamples and complete sample.

fires were more likely to occur in areas with low elevations and high frequency of human activity. In addition, the majority of Fujian was threatened by high and medium fire risk, especially the northern and southeastern regions, but relatively low fire risk was identified in the western part of Fujian mainly due to the low forest coverage caused by soil erosion. Overall, the findings provide

Table 3
Model prediction performance of the intermediate models created with logistic regression, including the cutoff.

Sample	Cutoff	Observed		Predicted		
				Fire		Percentage correct
				0	1	
Sample 1	0.445	Fire	0	6494	2476	72.4
			1	1417	6502	82.1
		Overall percentage				76.9
Sample 2	0.420	Fire	0	6280	2676	70.1
			1	1222	6711	84.6
		Overall percentage				76.9
Sample 3	0.442	Fire	0	6546	2486	72.5
			1	1423	6434	81.9
		Overall percentage				76.9
Complete dataset	0.404	Fire	0	10,252	4713	68.5
			1	1930	11,255	85.4
		Overall percentage				76.4

Table 4
Variables selected using stepwise regression for the complete dataset.

Variables	Coefficients	Standard error	Wald test	*P*-value
Constant	−0.1389−	0.1460	0.9053	0.3413
Elev.	−0.0019	0.00007	855.7574	<0.0001
Dis_road	0.2937	0.0206	203.2626	<0.0001
Dis_settlement	0.0449	0.0024	338.4808	<0.0001
GST_max	0.0058	0.0003	502.4062	<0.0001
GST_min	−0.0173	0.0004	2290.5652	<0.0001
Da_prec	−0.0053	0.0006	89.9448	<0.0001
SSD	0.0132	0.0007	391.6069	<0.0001
Da_RH	−0.0148	0.0016	89.6316	<0.0001
Den_Pop	0.0001	0.00002	31.4403	<0.0001
CGDP	0.0001	0.00001	164.5007	<0.0001

Note: Elev is elevation; Dis_road is distance to the nearest road; Dis_sett is distance to the nearest settlement; Da_maxtemp is daily maximum globe surface temperature; Da_mintemp is daily minimum globe surface temperature; Da_precipitation is daily precipitation; SSD is sunshine hours; Da_RH is daily mean relative humidity; Den-Pop is density of population; and CGDP is per capita GDP.

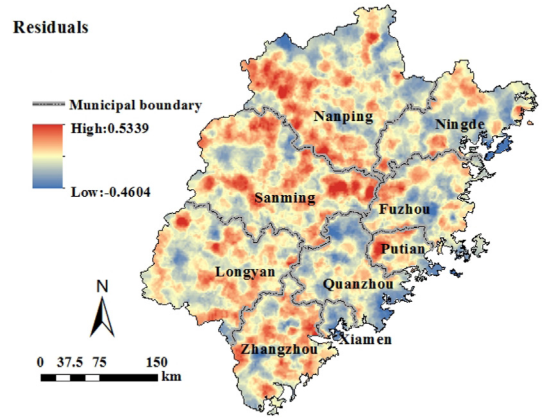

Fig. 7. Distribution of residuals obtained from the developed fire prediction model, overlaid by the eight main administrative areas of the study area (Fujian, China).

valuable insights that will guide planning and resource allocation for effective prevention of forest fire risk in Fujian province. Future efforts to prevent forest fire should focus on geographic patterns rather than administrative units as the fire-risk zones appear across all administrative regions in Fujian province.

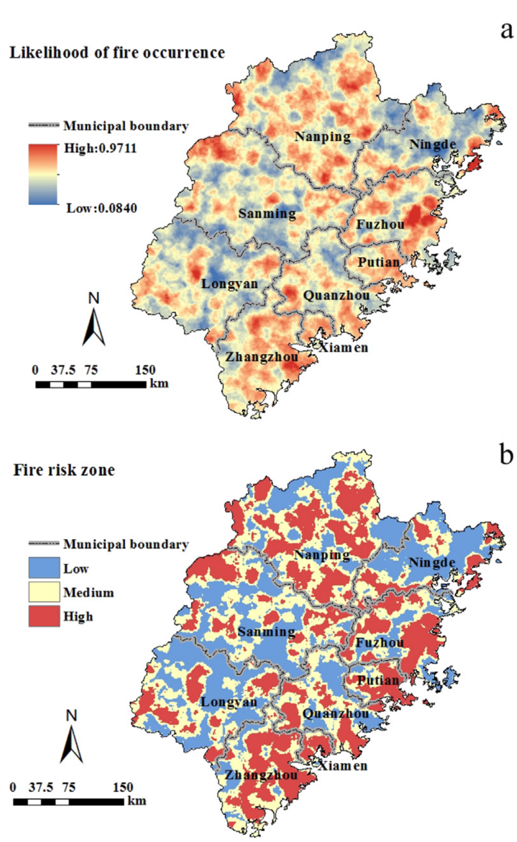

Fig. 6. (a, b) Distribution of fire occurrence likelihood based on prediction model within the eight administrative areas of Fujian, China. (a). Fire risk is represented by three classes: low, medium, and high fire risk (b).

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant No. 31400552), the Natural Science Foundation of Fujian Province (Grant No. 2015J05049), and Special funding for Universities by Fujian Education department (JK2014012). We also thank the anonymous reviewers for their valuable comments, which improved the paper significantly.

References

Butry, D. T., & Prestemon, J. P. (2005). *Spatio-temporal wildland arson crime functions*. Paper presented at the Annual Meeting of the American Agricultural Economics Association, 26–29 July, Providence, Rhode Island.
Castro, F. X., Tudela, A., & Sebastiá, M. A. (2003). Modeling moisture content in shrubs to predict fire risk in Catalonia (Spain). *Agricultural and Forest Meteorology, 116*, 49–59.
Catry, F. X., Rego, F. C., Bação, F. L., & Moreira, F. (2009). Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire, 18*(8), 921–931.
Chang, Y., Zhu, Z. L., Bu, R. C., Chen, H. W., Feng, Y. T., Li, Y. H., et al. (2013). Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. *Landscape Ecology, 28*(10), 1989–2004.
Chas-Amil, M. L., Prestemon, J. P., McClean, C. J., & Touza, J. (2015). Human-ignited wildfire patterns and responses to policy shifts. *Applied Geography, 56*, 164–176.
Chen, Z. Q., Chen, Z. B., & Chen, M. H. (2011). Geostatistical analysis on soil and water loss in Fujian province. *Chinese Journal of Natural Resources, 26*(8), 1394–1400.
Chuvieco, E., Cocero, D., Riaño, D., Martin, P., Martínez-Vega, J., de la Riva, J., et al. (2004). Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. *Remote Sensing of Environment, 92*, 322–331.
Del Hoyo, V. L., Isabel, M. P. M., & Vega, F. J. M. (2011). Logistic regression models for human-caused wildfire risk estimation: analysing the effect of the spatial accuracy in fire occurrence data. *European Journal of Forest Research, 130*, 983–996.
Deng, O., Li, Y. Q., Feng, Z. K., & Zhang, Y. D. (2012). Model and zoning of forest fire risk in Heilongjiang province based on spatial logistic. *Transactions of the CSAE, 28*(8), 200–205.
Dissing, D., & Verbyla, D. L. (2003). Spatial patterns of lightning strikes in interior Alaska and their relations to elevation and vegetation. *Canadian Journal of Forest Research, 33*, 770–782.
Flannigan, M. D., Stocks, B. J., & Wotton, B. M. (2000). Climate change and forest fires. *The Science of the Total Environment, 262*, 221–229.
Fry, D. L., & Stephens, S. L. (2006). Influence of humans and climate on the fire history of a ponderosa pine-mixed conifer forest in the southeastern Klamath Mountains, California. *Forest Ecology and Management, 223*, 428–438.
Genton, M. G., Butry, D. T., Gumpertz, M. L., & Prestemon, J. P. (2006). Spatio-temporal analysis of wildfire ignitions in the St Johns river water management district, Florida. *International Journal of Wildland Fire, 15*, 87–97.
Guo, F. T., Innes, J. L., Wang, G. Y., Ma, X. Q., Sun, L., Hu, H. Q., et al. (2015). Historic distribution and driving factors of human-caused fires in the Chinese boreal forest between 1972 and 2005. *Journal of Plant Ecology, 8*(5), 480–490.
He, Z. H., Liu, X. D., Zhao, H., & Zhou, J. Q. (2013). Forest flammability and temporal and spatial distribution rules of Forest fire in Jiangle county of Fujian province. *Hunan Agricultural Sciences, 07*, 108–111.

Hu, T. Y., & Zhou, G. S. (2014). Drivers of lightning- and human-caused fire regimes in the Great Xing'an Mountains. *Forest Ecology and Management, 329*, 49–58.

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modeling. *Global Ecology and Biogeography, 21*, 498–507.

Koutsias, N., Martínez-Fernández, J., & Allgöwer, B. (2010). Do factors causing wildfires vary in space? Evidence from geographically weighted regression. *GISci Remote Sensing, 47*, 221–240.

Kupfer, J. A., & Farris, C. A. (2007). Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. *Landscape Ecology, 22*, 837–852.

Liu, S. J., & Yang, J. (2013). Modeling spatial patterns of forest fire in Heilongjiang province using generalized linear model and maximum entropy model. *Chinese Journal of Ecology, 32*(6), 1620–1628.

Liu, Z., Yang, J., Chang, Y., Weisberg, P. J., & He, H. S. (2012). Spatial patterns and drivers of fire occurrence and its future trend under climate change in a boreal forest of northeast China. *Global Change Biology, 18*, 2041–2056.

Loboda, T., & Csiszar, I. A. (2007). Assessing the risk of ignition in the Russian far east within a modeling framework of fire threat. *Ecological Applications, 17*, 791–805.

Martínez, J., Vega-Garcia, C., & Chuvieco, E. (2009). Human-caused wildfire risk rating for prevention planning in Spain. *Journal of Environmental Management, 90*, 1241–1252.

Miranda, B. R., Sturtevant, B. R., Stewart, S. I., & Hammer, R. B. (2012). Spatial and temporal drivers of wildfire occurrence in the context of rural development in northern Wisconsin, USA. *International Journal of Wildland Fire, 21*, 141–154.

Mundo, I. A., Wiegand, T., Kanagaraj, R., & Kitzberger, T. (2013). Environmental drivers and spatial dependency in wildfire ignition patterns of northwestern Patagonia. *Journal of Environmental Management, 123*, 77–87.

Nunes, M. C. S., Vasconcelos, M. J., Pereira, J. M. C., Dasgupta, N., Alldredge, R. J., & Rego, F. C. (2005). Land cover type and fire in Portugal: do fires burn land cover selectively? *Landscape Ecology, 20*(6), 661–673.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. C. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. *Forest Ecology and Management, 275*, 117–129.

Perry, G. L. W. (2004). SpPack: spatial point pattern analysis in Excel using Visual Basic for Applications (VBA). *Environmental Modelling & Software, 19*, 559–569.

Podur, J., Martell, D. L., & Csillag, F. (2003). Spatial patterns of lightning-caused Forest fires in Ontario, 1976-1998. *Ecological Modelling, 164*, 1–20.

Preisler, H. K., Brillinger, D. R., Burgan, R. E., & Benoit, J. W. (2004). Probability based models for estimating wildfire risk. *International Journal of Wildland Fire, 13*, 133–142.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability, 13*(2), 255–266.

Rodrigues, M., de al Riva, J., & Fotheringham, S. (2014). Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Applied Geography, 48*, 52–63.

Romero-Calcerrada, R., Barrio-Parra, F., Millington, J. D. A., & Novillo, C. J. (2010). Spatial modeling of socioeconomic data to understand pattern of human-caused wildfire ignition risk in the SW of Madrid (central Spain). *Ecological Modelling, 221*, 34–45.

Saefuddin, A., Setiabudi, N. A., & Fitrianto, A. (2012). On comparison between logistic regression and geographically weighted logistic regression: with application to Indonesian poverty data. *WASJ, 19*(2), 205–210.

San-Miguel-Ayanz, J., & Camiá, A. (2009). Forest fires at a glance: facts, figures and trends in the EU. Discussion paper. In Y. Birot (Ed.), *Living with wildfires: What science can tell us. A contribution to the science-policy dialogue* (Vol. 15, pp. 11–18). European Forest Institute.

Schoennagel, T., Veblen, T. T., & Romme, W. H. (2004). The interaction of fire, fuels, and climate across rocky mountain forests. *BioScience, 54*(7), 661–676.

Stephens, S. L. (2005). Forest fire causes and extent on United States forest servce lands. *International Journal of Wildland Fire, 14*, 213–222.

Syphard, A. D., Radeloff, V. C., Keely, J. E., Hawbaker, R. J., Clayton, M. K., Stewart, S. I., et al. (2007). Human influence on California fire regimes. *Ecological Applications, 17*, 1388–1402.

Syphard, A. D., Radeloff, V. C., Keuler, N. S., Taylor, R. S., Hawbaker, T. J., Stewart, S. I., et al. (2008). Predicting spatial patterns of fire on a southern California landscape. *International Journal of Wildland Fire, 17*, 602–613.

Todd, S. K., & Jewkes, H. A.. (2006). Wildland fire in Alaska: a history of organized fire suppression and management in the last frontier. Agricultural and Forestry Experiment Station Bulletin No.114 University of Alaska Fairbanks March.

Vega Garcia, C., Woodard, P. M., Titus, S. J., Adamowicz, W. L., & Lee, B. S. (1995). A logistic model for predicting the daily occurrence of human caused forest fires. *International Journal of Wildland Fire, 5*(2), 101–111.

Wotton, B. M., Martell, D. L., & Logan, K. A. (2003). Climate change and people-caused forest fire occurrence in Ontario. *Climatic Change, 60*, 275–295.

Wu, Z. W., He, H. S., Yang, J., & Liang, Y. (2015). Defining fire environment zones in the boreal forests of northeastern China. *Science of the Total Environment, 518–519*, 106–116.

Wu, Z. E., He, H. S., Yang, J., Liu, Z. H., & Liang, Y. (2014). Relative effects of climatic and local factors on fire occurrence in boreal forest landscapes of northeastern China. *The Science of the Total Environment, 493*, 472–480.

Wu, W., & Zhang, L. J. (2013). Comparison of spatial and non-spatial logistic regression. *Applied Geography, 37*, 52–62.

Zhang, Z. X., Zhang, H. Y., Li, D. X., Xu, J. W., & Zhou, D. W. (2013). Spatial distribution pattern of human-caused fires in Hulunbeir grassland. *Acta Ecological Sinica, 33*, 2023–2031.

Zhang, Z. X., Zhang, H. Y., & Zhou, D. W. (2010). Using GIS spatial analysis and logistic regression to predict the probabilities of human-caused grassland fires. *Journal of Arid Environments, 74*, 386–393.

Zheng, H. Q., Chen, J. P., Zhang, X., Zhang, C. A., Zhang, C. G., & Chen, H. (2001). Study on the forecast system of Forest fire weather ranks in Fujian, 38–44+50 *Chinese Journal of Agrometeorology, 03*.

Zumbrunnen, T., Pezzatti, G. B., Menéndezd, P., Bugmann, H., Bürgi, M., & Conedera, M. (2011). Weather and human impacts on forest fires: 100 years of fire history in two climatic regions of Switzerland. *Forest Ecology and Management, 261*, 2188–2199.