Comparing four regression techniques to explore factors governing the number of forest fires in Southeast, China

Qianqian Cao\textsuperscript{a}, Lianjun Zhang\textsuperscript{a}, Zhangwen Su\textsuperscript{b}, Guangyu Wang\textsuperscript{c}, Shuaichao Sun\textsuperscript{d} and Futao Guo\textsuperscript{d}

\textsuperscript{a}Department of Sustainable Resources Management, College of Environmental Science and Forestry, State University of New York (SUNY-ESF), Syracuse, NY, USA; \textsuperscript{b}College of Forestry, Northeast Forestry University, Harbin, Heilongjiang, P.R. China; \textsuperscript{c}Asia Forest Research Centre, Faculty of Forestry, University of British Columbia, Vancouver, BC, Canada; \textsuperscript{d}College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, Fujian, P.R. China

\textbf{ABSTRACT}

Four regression techniques, including two global models (i.e., Poisson and negative binomial) and two geographically weighted regression (GWR) models (i.e., geographically weighted Poisson regression (GWPR) and geographically weighted negative binomial regression (GWNBR)) were used to explore which was the most suitable method for predicting the number of forest fires and to investigate the spatially varying relationships between forest fires and environmental factors in Fujian province, in the Southeast of China. Our results showed that the GWR models fitted the fire count data better than the global models, and yielded more realistic spatial distributions of model predictions. Particularly, GWNBR was superior for addressing overdispersion in the fire count data because it estimated the dispersion parameter at a local level. Additionally, our study indicated that more forest fires occurred in areas of lower elevation, flatter terrain, and higher population density. The global models showed that precipitation had positive impacts on fire occurrence in the study area. In contrast, the GWR models revealed that precipitation was positively related to the forest fires in the western regions of Fujian, but negatively related in the eastern coastal regions. Our study could provide better insight into forest fire management based on local environmental characteristics.

\textbf{ARTICLE HISTORY}

Received 30 July 2020  
Accepted 28 January 2021

\textbf{KEYWORDS}

geographically weighted Poisson regression; geographically weighted negative binomial regression; forest fire count; over-dispersion; spatial autocorrelation and heterogeneity

\textbf{1. Introduction}

Forest fires can spread across a large area of forested land and lead to significant, long-lasting impacts on ecological, social, and economic systems (Scott et al. 2013). In recent decades, forest fire occurrence has risen, along with the severity and damage they have inflicted (Stocks and Martell 2016; Guo et al. 2015).
Many statistical methods and models have been developed to understand the influencing factors for the occurrence and number of forest fires, and predict their potential threats to the environment, property, and lives (Costafreda-Aumedes et al. 2017). In recent years, advances in computer technology and software have led to wide applications of regression techniques in forest fire modeling, which demonstrate the dynamic relationships between forest fires and explanatory variables such as vegetation patterns, fuel moisture conditions, meteorological variables, and some socio-economic factors (e.g., population density, gross domestic product (GDP), education level, etc.). Poisson regression has been considered a reliable model for predicting the number of fires per region and time period since the 1970s (Cunningham and Martell 1973; Dayananda 1977; Gill et al. 1987; Mandallaz and Ye 1997; Wotton et al. 2003, 2010). Furthermore, negative binomial (NB) regression has also been applied in forest fire modeling when the fire data (count response variable) is dispersed. Studies have shown that the NB regression performs better in both model fitting and prediction performance (Cardille et al. 2001; Xiao et al. 2011; Chas-Amil et al. 2015).

The above Poisson and NB models can be classified as global models, meaning that a single model with one set of model parameters can be used to explain the entire study region. However, spatial autocorrelation and heterogeneity exist across forest ecosystems, and the relationships between forest fires and environmental factors are non-stationary across space. Geographically weighted regression (GWR) was introduced by Fotheringham et al. (2002) as an alternative approach for modeling spatially non-stationary fires across geo-locations. GWR considers varying relationships in space and takes local variations into account. With geographical information system (GIS) technology, the spatially varying regression coefficients of the GWR models can be visualized to identify local trends and spatial 'hotspots' (Fotheringham and Brunsdon 2010). In recent years, GWR has been extended to the framework of generalized linear models, such as geographically weighted Poisson regression (GWPR) and geographically weighted negative binomial regression (GWNBR) (Nakaya et al. 2005; Su et al. 2019). The new developments have provided an appropriate foundation for modeling spatially varying binary and count response variables in the field of forest fires (Koutsias et al. 2010; Rodrigues et al. 2014; Guo et al. 2016b).

However, GWPR is more challenging than a global Poisson model due to a common problem of overdispersion in spatial count data (Haining et al. 2009). Overdispersion causes the challenge due to the strict requirement of a Poisson distribution for the count response variable such that the mean is equal to the variance. Given the nature of rare events, the number (count) of forest fires usually has much larger variance than the mean (i.e., overdispersion), because the zero count tends to occur more often than higher numbers of fire events. To model the spatial count data with overdispersion, it may be more appropriate to use a negative binomial distribution instead of a Poisson distribution. Da Silva and Rodrigues (2014) proposed the geographically weighted negative binomial regression (GWNBR) for incorporating spatial count data with overdispersion. Including spatial effects into statistical models is valuable for understanding relationships geographically and identifying local 'hot spots' of high fire risks. In the past studies of forest fire modeling, the single model
method was usually used. However, a comparison study among different regression techniques was rarely conducted.

In this study, we applied two global models (i.e., Poisson and negative binomial) and two GWR models (i.e., GWPR and GWNBR) to develop forest fire prediction models and identify the driving factors of forest fire counts in a sub-tropical region of China. The results from each individual model were compared and the localized significant explanatory variables for prediction and prevention of forest fires were targeted. Hopefully, our study can improve the comprehensive understanding of the applicability of global and GWR models in forest fire studies.

2. Data and methods

2.1. Study area

Fujian province is located in a sub-tropical region of China with a total land area of 124,000 km$^2$ (Figure 1). It has the highest forest coverage in the nation (about 66% of Fujian province is covered by forests and vegetation), and experiences high annual forest fire incidences, with nearly 15,000 forest fires occurring from 2000 to 2010 (Guo et al. 2017). The dominant tree species in the province include Massoniana (Pinus massoniana Lamb.), Chinese fir (Cunninghamia lanceolata(Lamb.)), Casuarina (Casuarina equisetifolia L.), Moso bamboo(Phyllostachys heterocycle (Carr.) Mitford cv. Pubescens). Climate is warm and humid with an average annual rainfall of 1400 – 2000 mm and average temperature of 17 – 21 °C. Forest fire season typically spans from September to April (Guo et al. 2016a).

Figure 1. Location map and distribution of fire points of Fujian province, China. Source: Author.
2.2. Data preparation

2.2.1. Fire data (response variable)

In this study we used Medium Resolution Imaging Spectroscopy (MODIS) to record the spatial distribution of fire pixels in Fujian province from 2001 to 2016. This product is considered a reliable and suitable source for monitoring forest fires (Justice et al. 2002). We obtained a daily forest fire product (MOD14A1) with a resolution of 1 km, which has been widely used in recent forest fire studies (Amraoui et al. 2015; Guo et al. 2017). Since this product cannot distinguish forest fires from non-forest fires that occur in cities/towns, construction sites, agricultural lands, and other areas, we further processed the fire data by (1) removing the fire points in cities/towns, construction sites, and farmland based on an 1-km resolution land-use map (the map is provided by Resource and Environmental Data Cloud Platform (http://www.resdc.cn/Default.aspx)), and (2) extracting the fire points based on the time of fire occurrence within the fire season (September 15 to April 30 of the following year). All forest fire points were recorded using geographical coordinates. The Create Fishnet and Spatial Join tools in ArcGIS 10.2 (ESRI, 2010) were then used to divide Fujian province into $4 \times 4$ km grids (a total of 7433 grids). Thus, the response variable was the number of

![Frequency Distribution of Forest Fire Occurrences](image)

Figure 2. Frequency distribution of forest fire occurrences. Note: The x-axis represents the number of fires in each unit grid, and the y-axis represents the relative frequency of forest fire. Source: Author.
forest fires in each grid during a period of 16 years (2001 – 2016). Figure 2 shows the frequency distribution of forest fire counts: there is a strong mode at zero count and a heavy right tail, indicating the frequency of pixels with zero fire count is much higher, and the frequencies of other fire counts rapidly decreased. Figure 3a illustrates the spatial distribution of forest fire points.

**2.2.2. Potential driving factors (explanatory variables)**

A total of 15 explanatory variables were collected in this study as follows:

1. Topographic variables included elevation (km), slope (degree), and aspect. High resolution (25 m) Digital Elevation Model (DEM) data was collected from the National Administration of Surveying, Mapping and Geoinformation of China (http://www.gscloud.cn/sources). Using the 3D Analysis tool in ArcGIS 10.2, the slope and slope direction were derived from the DEM, and aspect was then converted into an aspect index using the following formula (Guo et al. 2017): Aspect index = \( \cos(\theta \times \pi/180) \), where \( \theta \) is the degree of slope generated in ArcGIS ranging from 0 - 360° so that the aspect index ranges from -1 to 1. An aspect index
value closer to 1 indicates higher potential solar radiation. Average elevation, slope, and aspect index were then extracted in ArcGIS 10.2 for each grid.

2. Meteorological variables included precipitation (mm/day), temperature (°C), and relative humidity (%), which were derived from the HadCM2 global climate model (Guo et al. 2016b), an ocean-air coupling model developed by the Hadley Centre. ArcGIS 10.2 was used to calculate annual average daily values of each meteorological variable for each grid.

3. Human factors included socioeconomic variables (per capita GDP and population density) and infrastructural variables. GDP and population data were obtained from the Resource and Environmental Data Cloud Platform (http://www.resdc.cn/Default.aspx). The data included the grid population density and per capita GDP for the years 2000, 2005, 2010, and 2015 at a 1-km resolution. Based on the raster population and GDP data, the Raster Calculator tool in ArcGIS 10.2 was used to calculate average annual growth rates of population and per capita GDP from 2000 to 2015. The average population and per capita GDP from 2000 to 2015 were calculated for each grid. Infrastructural variables included road density (km/km², ratio of road length to grid area) and water density. A 1:250,000 vector map of infrastructure was provided by the National Geomatics Center of China (http://www.ngcc.cn/) and ArcGIS 10.2 was used to evaluate the ratio of the length of road and area of water within each grid to the grid area.

4. Vegetable coverage is commonly used to indicate the total amount of live and dead fuels above the surface. One practical method to estimate this variable was the Normalized Difference Vegetation Index (NDVI). The NDVI data were derived from the MODIS NDVI product with a spatial resolution of 500 m, provided by the Geospatial Data Cloud (http://www.gscloud.cn/). Land use feature variables were estimated using the spatial distribution data of one million vegetation types in China from the Resource and Environmental Data Cloud Platform (http://www.resdc.cn/Default.aspx). Forest, shrub, grass, and crop were four major vegetation types considered, respectively covering 64.95%, 20.40%, 1.69%, and 12.6% of the total land area in this study. ArcGIS 10.2 was used to calculate the proportion of each land cover in each grid.

5. In this study, the variables were obtained and collected from different types of data (e.g., vector data and raster data), and the resolution of the raster data was also different. In order to ensure the accuracy of the analysis and avoid systematic errors in data analysis, we unified all variables to the same resolution scale. In addition, we set the resolution scale of the reanalyzed data to 4 × 4 km. The reason was that the area of Fujian province is about 12.4 × 10³ km². If the grid resolution was too small, the sample size would be too large so that the computation would be challenging. Meanwhile, a smaller pixel size would increase the number of ‘0’ pixels, which would lead to the discretization of data or dependent variables to be severely over dispersed.

2.2.3. Preliminary selection of variables
A multicollinearity analysis was performed before model fitting. The variance inflation factor (VIF) was used to detect multicollinearity problems among the
explanatory variables. In general, a VIF above ten indicates that the parameter estimate and its standard error of an explanatory variable are impacted by multicollinearity (Myers et al. 2002; Guo et al. 2017). We used a global Poisson model to test multicollinearity, which resulted in elimination of the socioeconomic variable GDP because its VIF was 18.56. Therefore, a total of 14 explanatory variables were used to fit both the global and GWR models in this study. The descriptive statistics of response and explanatory variables are shown in Table 1.

### 2.3. Models

#### 2.3.1. Poisson regression

A count response variable \( (Y_i) \) has non-negative integer values \((i = 1, 2, \ldots, n)\), and Poisson regression is often used to model a count response variable \( Y_i \) against the underlying predictor variables (McCulloch and Searle 2001; Myers et al. 2002). The probability density function \((pdf)\) of the Poisson distribution is

\[
P(Y_i) = \frac{e^{-\mu_i} \cdot \mu_i^{Y_i}}{Y_i!}
\]

where \( P(Y_i) \) is the probability that the number of events \((Y_i) \) occurred during a given time period, and \( \mu_i \) is the parameter representing the expected value of \( Y_i \). The Poisson distribution assumes equal mean and variance such that \( E(Y_i) = \mu_i \) and \( Var(Y_i) = \mu_i \). The explanatory variables are linked to the expected value \( \mu_i \) via a link function such as a natural logarithm:

\[
\ln (\mu_i) = \beta_0 + X_i \beta
\]

where \( X_i \) represents the explanatory variables, \( \beta_0 \) is the intercept coefficient, and \( \beta \) is the vector of the model slope coefficients. Thus, the expected value of \( \mu_i \) can be predicted by the inverse link function \( \hat{\mu}_i = e^{(\hat{\beta}_0 + X_i \hat{\beta})} \).
2.3.2. Negative binomial (NB) regression

Although Poisson regression is a common choice for modeling a count response variable, it is often criticized for its restrictive assumption of equal mean and variance. It is well known that the potential drawback of Poisson regression is the underestimation of standard errors of the model coefficients due to the overdispersion problem in the data. One way of dealing with this issue is to rescale the standard errors by the estimated dispersion parameter, while keeping the model coefficients unchanged. However, a better alternative for correcting the overdispersion problem is to use a negative binominal (NB) regression, which automatically builds in a dispersion parameter in its distribution function so that the estimation of both model coefficients and standard errors are corrected for overdispersion in the data (McCulloch and Searle 2001; Myers et al. 2002). The unconditional distribution of $Y_i$ can be written as follows:

$$P(Y_i) = \frac{\Gamma(Y_i + 1/k)}{\Gamma(Y_i + 1)\Gamma(1/k)} \left( \frac{1}{1 + \kappa \mu} \right)^{1/\kappa} \left( \frac{\kappa \mu}{1 + \kappa \mu} \right)^{Y_i}$$

where $\Gamma$ denotes the gamma function, $\kappa$ is the dispersion parameter, and the mean and variance of $Y_i$ are

$$E(Y_i) = \mu$$

$$V(Y_i) = \mu + \kappa \mu^2 = \mu(1 + \kappa \mu)$$

Thus, the Poisson model is the limiting model of the negative binomial model when $\kappa \to 0$. The common link function for the negative binomial regression is the same as the Poisson regression (Eq. [2]).

2.3.3. Geographically weighted regression (GWR)

To investigate the spatial variation or heterogeneity of a regression relationship, data must be collected with the location coordinates $(v_{xi}, v_{yi})$ for each observation $i$. This local information allows for estimation of the localized regression coefficients of the relationship of interest. When GWR was first developed, the Gaussian assumption was assumed for the model error term (Fotheringham et al. 1998), expressed as:

$$Y_i = \beta_0(v_{xi}, v_{yi}) + \sum_{k=1}^{p} \beta_k(v_{xi}, v_{yi})X_{ki} + \varepsilon_i$$

where $Y_i$ is the response variable, $X_k$ is a set of $p$ explanatory variables ($k = 1, 2, \ldots, p$), $\beta_0(v_{xi}, v_{yi})$, $\beta_1(v_{xi}, v_{yi})$, $\ldots$, $\beta_p(v_{xi}, v_{yi})$ are the regression coefficients for the $k$th predictor variable at the $i$th location, and $\varepsilon_i$ is the random error term whose distribution is assumed $N(0, \sigma^2I)$ with $I$ denoting an identity matrix. The aim of GWR is to obtain the estimates of these functions for each predictor variable and each geographic location $i$. The estimation procedure is as follows: (i) draw a circle of a given radius around one particular location $i$ (the center), (ii) compute a weight ($w_{ij}$) for
each neighboring observation \( j \) according to the distance \( d_{ij} \) between location \( j \) and center \( i \), and (iii) estimate the model coefficients using weighted least-squares regression such that:

\[
\hat{\beta}_i = (X'W_iX)^{-1}X'W_iY
\]  

where the weight matrix \( W_i \) is:

\[
W_i = \begin{pmatrix}
w_{i1} & 0 & \cdots & 0 \\
0 & w_{i2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & w_{in}
\end{pmatrix}
\]

The weighting function is defined by the kernel type and size of kernel (bandwidth), which determines the geographical weight of the \( j \)th neighboring observation at the \( i \)th regression point. The weight should decrease gradually as the distance between \( i \) and \( j \) increases, until it reaches a constant or zero. The model parameter estimates are highly related to the kernel size, so the choice of kernel is important in the modeling process in GWR.

### 2.3.4. Geographically weighted Poisson regression (GWPR)

The GWPR model is developed by adding geographical location into the standard Poisson regression. It uses a similar link function to Eq. [2] and is in the following form:

\[
\ln(\mu_i) = \beta_0(v_{xi}, v_{yi}) + \sum_{k=1}^{P} \beta_k(v_{xi}, v_{yi})X_{ki}
\]

where \( \beta_0 \) and \( \beta_k \) are the GWPR model parameters specifically describing the location of \( i \) with \( v_{xi} \) and \( v_{yi} \) coordinates.

### 2.3.5. Geographically weighted negative binomial regression (GWNBR)

GWNBR is an extension of the global model of NB regression that allows the spatial variation of parameters \( \beta_0 \), \( \beta_k \), and \( \kappa \). This local model is described as:

\[
\ln(\mu_i) = \beta_0(v_{xi}, v_{yi}) + \sum_{k=1}^{P} \beta_k(v_{xi}, v_{yi})X_{ki}, \quad \kappa_i(v_{xi}, v_{yi})
\]

where \( \beta_0 \) and \( \beta_k \) are the GWPR model parameters specifically describing the location of \( i \) with \( v_{xi} \) and \( v_{yi} \) coordinates, and \( \kappa_i \) is the local dispersion parameter.

In this study, we applied the same Gaussian kernel function and bandwidths to both GWPR and GWNBR models. It is known that the bandwidth has profound impacts on model fitting, the spatial distribution of model predictions, and the localized model coefficients (Fotheringham et al. 2002; Guo et al. 2008). If different bandwidths were used for GWPR and GWNBR, the modeling results would be less...
comparable between two models because we would not be able to distinguish if the model differences were due to the models *per se* or the bandwidths used.

### 2.4. Model evaluation and comparison

Overdispersion in the response variable is always a concern when modeling count data. One way to detect the problem is to divide the model deviance, which measures the discrepancy between the observed and fitted response variable, by its degrees of freedom (df). If this deviance/df ratio is close to 1, there is no concern of overdispersion; if it is greater than 1, an overdispersion problem may exist and some correction may be necessary (Myers et al. 2002).

In a GWR model, it is necessary to decide on an optimal bandwidth for model fitting (Fotheringham et al. 2002). In this study, we used the Akaike Information Criterion (AIC) to decide the optimal bandwidth and related kernel function for estimating each GWR model (Fotheringham et al. 2002; Guo et al. 2008). The estimated bandwidth was selected to be 112,410 meters, depending on the residuals of global models.

To evaluate spatial variation or heterogeneity in the model coefficients of GWPR and GWNBR, we followed the approach in Chen et al. (2012). The interquartile range (IQR) of the coefficient estimates computed by the GWR localized models was compared to the standard error of the global estimates derived from a traditional regression model. When IQR is twice as large as the standard error, it indicates that spatial non-stationary exists in the relationships between the response variable and its accompanying predictor variables.

Model fitting was evaluated using AIC and mean squared errors (MSE) (Burnham and Anderson 2004). Smaller AIC and/or MSE implies better model fitting performance. To evaluate the spatial autocorrelation of the residuals, the Geary’s contiguity ratio (Geary’s C) was calculated, as well as the Z-test for testing $H_0: C = 1$ and associated $p$-value. The closer to 1 the Geary’s C is, the lower the spatial dependence of the residual is, and hence, the model accounts for more spatial structure problems.

Because the generalized linear models (i.e., Poisson and NB) are estimated using maximum likelihood estimation, no sums of squares of total, regression and residuals are computed so that no meaningful $R^2$ can be calculated (Zheng and Agresti 2000). To assess the model prediction performance, we applied several measures: (1) Chi-square ($\chi^2$) goodness of fit was used to test the overall performance of model prediction among the four models (Terceiro 2003; Zhen et al. 2018); (2) a correlation measure between observed response variable ($Y$) and model prediction($\hat{Y}$) (Zheng and Agresti 2000):

$$\text{Cor}(Y, \hat{Y}) = \frac{\text{Cov}(Y, \hat{Y})}{\sqrt{\text{Var}(Y) \cdot \text{Var}(\hat{Y})}}$$  \hspace{1cm} [11]$$

where $\text{Cov}(Y, \hat{Y})$ is the covariance between $Y$ and $\hat{Y}$, $\text{Var}(Y)$ is the variance of $Y$, and $\text{Var}(\hat{Y})$ is the variance of ($\hat{Y}$). This measure is actually the positive square root...
of the average proportion of variance explained by the model, which equals the square root of $R^2$ in an OLS linear regression; (3) the prediction performance of the two GWR models were further compared by testing the differences of the mean squared prediction errors ($\text{MSPE} = (\text{observation} - \text{prediction})^2$) across six categories of fire counts (i.e., 0, 1, 2, 3–4, 5–10, and >10); and (4) the predicted number of forest fires from the four models and the spatial distribution of model coefficients for each explanatory variable were mapped using ArcGIS10.2.

### 3. Results

#### 3.1. Comparison of model fitting and prediction accuracy

##### 3.1.1. Overall comparison between global and GWR models

The statistics of model fitting are listed in Table 2, including AIC, MSE, and Geary’s C of model residuals with the corresponding Z-test and $p$-value. The AIC values were 53282.4 for the global Poisson model and 36446.6 for the global NB model, which were much greater than those of both GWPR (27066.3) and GWNBR (22886.7) models. The MSE of the GWR models (28.4 for GWPR and 31.3 for GWNBR) were smaller than those of the two global models (36.2 for Poisson and 36.7 for NB). Thus, it was evident that the GWR models fitted the fire count data better than the two global models. Additionally, the Geary’s C of the two global model residuals was significantly smaller than one, indicating a clustered spatial pattern (i.e., positive spatial autocorrelation in model residuals). In contrast, the Geary’s C of the GWNBR model residuals was not significantly different from one, representing a desirable random spatial pattern. The Geary’s C of the GWPR model residuals was significantly larger than one, indicating a dispersed/uniform spatial pattern (i.e., negative spatial autocorrelations in model residuals) (Table 2).

To compare the overall performances of model predictions, we used Chi-square ($\chi^2$) goodness of fit statistics (Table 2). The Chi-square statistics of the two GWR models were much smaller than those of the two global models, indicating that GWPR and GWNBR were better than Poisson and NB in terms of model prediction performance. Similarly, the model $\text{Cor}(Y, \hat{Y})$ indicated that the two GWR models predicted the fire counts much better than the two global models, while GWPR was slightly better than GWNBR (Table 2).

Further, the predictions of the two GWR models were closer to the observed forest fire counts than the global models. The predicted counts from the global models did not exceed 20, indicating that no grid across Fujian province had fire occur more
than 20 times during this time period (Figures 4(a) and (c)). In contrast, the predictions from the GWR models showed wider ranges, where some locations were predicted to have greater than 20 fire occurrences during this time period (Figures 4(b) and (d)). Regarding spatial distribution of model predictions, the GWR models showed that high incidences (>10) of forest fires were mainly concentrated in the northwest and center regions of the study area, with only a few high incidences of forest fires predicted by the global models randomly scattered across the study area (Figures 4(a) and (c)).

3.1.2. Comparison between global Poisson and NB models

In this study, the deviance/df ratio of the global Poisson model was 4.6593 (Table 2), showing the existence of the overdispersion problem. The deviance/df ratio of the global NB model was 1.1117, close to 1, indicating that the NB model was indeed a better choice than the Poisson model for handling overdispersion in the fire count.

Figure 4. Spatial distribution of model predictions from (a) global Poisson, (b) GWPR, (c) global NB, and (d) GWNBR. The predicted fire occurrence number was divided into six categories (0, 1, 2, 3–4, 5–10, >10). Source: Author.
data. Furthermore, the AIC of the two models confirmed that the global NB model (36446.6) was superior to the global Poisson model (53282.4) for fitting the fire count data (Table 2).

The predictions from both global Poisson and NB models were between 0 and 20, suggesting that 0 or at most 20 fires were expected in each grid during this time period. Both global models projected that the areas with frequent forest fires were mainly concentrated in the northwest and southeast of the study area (Figures 4(a) and (c)).

### 3.1.3. Comparison between GWPR and GWNBR models

Similar to the global models, the GWNBR model (22886.7) had a much smaller AIC than the GWPR model (27066.3). The Geary’s $C$ of GWNBR revealed that its model residuals were randomly distributed. The results of AIC and Geary’s $C$ implied that GWNBR was better than GWPR for fitting the fire count data. The spatial distributions of the fire predictions from the two GWR models were similar, i.e., the areas with high frequencies of forest fires (i.e., predicted fire occurrence $> 10$) were concentrated in the northwest and center of Fujian province. The Chi-square statistics indicated the GWNBR was slightly better than GWPR in terms of model prediction (Table 2).

The correlation measure $\text{Cor}(Y, \hat{Y})$ indicated that the overall model prediction of GWPR (0.5653) was slightly better than GWNBR (0.4926) (Table 2). To further compare the performance of model predictions between two GWR models, the observed forest fire counts were divided into six categories (i.e., 0, 1, 2, 3-4, 5-10, and $>10$) and the mean prediction error (MSPE) of GWPR and GWNBR was tested for each category (Table 3). The results showed that the MSPE of the GWNBR model was significantly smaller than that of GWPR in the category of 5-10 fire counts, which had the most fire occurrence cases (1735) of all categories, while the MPEs of other categories were similar between the two GWR models. It was evident that GWNBR performed better than GWPR for the major category of forest fire counts.

### 3.2. Variability of significant explanatory variables

#### 3.2.1. Overall comparison between global models and GWR models

The relationship between forest fire counts and explanatory variables estimated by the global models is assumed constant and stationary across the study area. On the other hand, the model coefficients of the GWR models are spatially varied from location to location. The results of the GWR models showed that the relationship

| Category of fire occurrence | Number of cases | GWNBR (MSPE) | GWPR (MSPE) | t-test | p-value |
|-----------------------------|----------------|--------------|-------------|--------|---------|
| 0                           | 1525           | 10.00        | 9.20        | 1.31   | 0.19    |
| 1                           | 1147           | 10.90        | 9.80        | 0.85   | 0.40    |
| 2                           | 1036           | 7.19         | 6.75        | 0.66   | 0.51    |
| 3-4                         | 1361           | 6.63         | 6.06        | 0.62   | 0.54    |
| 5-10                        | 1735           | 12.62        | 28.38       | -4.95  | <0.0001 |
| $>10$                       | 629            | 265.01       | 238.21      | 0.51   | 0.61    |

Table 3. Mean squared prediction error (MSPE) of GWNBR and GWPR models for six categories of fire occurrence.
between fire counts and explanatory variables was spatially non-stationary. Compared to the global models, the GWR models highlighted the spatial heterogeneity of the relationships between response and explanatory variables. The significance level was chosen at $\alpha = 0.05$ for this study. Both global models showed that elevation, slope, aspect, humidity, precipitation, population, NDVI, road density, water density, grass cover, and crop cover were significantly related to forest fire counts (Table 4). However, the significance of the GWR model coefficients may not be consistent across the study area, but vary from location to location. Figures 5 and 6 demonstrate the locally significant model coefficients.

### 3.2.2. Comparison of explanatory variables of global models

The estimated coefficients of the global Poisson and NB models are listed in Table 4. The relationship between significant explanatory variables and forest fire counts in the NB models had the same correlation as in the global Poisson model. Additionally, in both global Poisson and NB models, the estimated model coefficients of humidity, precipitation, road density, and grass cover were significantly positive, indicating that greater relative humidity and/or precipitation, denser roads and more grass cover may lead to more forest fires (see discussion section for details). In contrast, the estimated model coefficients of water density, population density, NDVI, crop cover, and two topographic factors (elevation and aspect) were significantly negative, indicating that forest fires were less likely to occur in developed areas, at relatively high elevations, with less sunshine radiation and/or in the areas with low crop cover.

### 3.2.3. Comparison of spatial variability of significant variables in GWR models

The GWR Poisson and NB model coefficients of all 14 explanatory variables were spatially varied because their IQRs were at least twice as large as the standard errors of the corresponding global model coefficients. Our results suggested that the relationships between forest fire counts and topographical, meteorological variables, human variables (e.g., population, water, and road density), and vegetation and land...
use cover were indeed heterogeneous across the study region (Table 5). Spatial distributions of the significant model coefficients of explanatory variables were similar between GWPR and GWNBR. Additionally, the coefficients of explanatory variables were only significant in some locations of the study area, and the influencing direction (the sign of coefficients) and power (estimated value of coefficients) were also spatially varied (Figures 5 and 6). For example, temperature was positively related to the forest fire counts in the north, but negatively related to the fire counts in the center of the province. Similarly, the precipitation and population density demonstrated negative relationships with the forest fire counts in the east and southeast of Fujian, but positive relationships in the north.

Furthermore, negative-binomial regression addresses the issue of overdispersion by including a dispersion parameter to accommodate the unobserved heterogeneity in different geo-locations. Figure 7 clearly indicates that, similar to the model

Figure 5. Spatial distribution of significant model coefficients of (a) Temperature, (b) Precipitation, (c) Relative humidity, (d) Water density, (e) Road density, (f) Population density, (g) Elevation, (h) Slope, (i) Aspect, (j) NDVI, (k) Forest cover, (l) Shrub cover, (m) Crop cover, and (n) Grass cover of the GWPR model. Grey pixels are non-significant coefficients. Source: Author.
coefficients, the localized dispersion parameters of the GWNBR model varies across the study area, with higher dispersion parameters from northeast to southern coastal areas, and smaller ones in the inland of Fujian province.

4. Discussion

4.1. Comparison of prediction models for the number of forest fires

Our results indicated that the global NB model fitted the data better than the global Poisson model (Table 2). More importantly, the global NB model incorporated the dispersion parameter to overcome the overdispersion problem in the fire count data. Our findings were similar to other studies in the literature (Guo et al., 2016d). Further, it was evident that the two GWR models performed much better than the two global models, which was indeed expected as the GWR models estimated the

Figure 6. Spatial distribution of significant model coefficients of (a) Temperature, (b) Precipitation, (c) Relative humidity, (d) Water density, (e) Road density, (f) Population density, (g) Elevation, (h) Slope, (i) Aspect, (j) NDVI, (k) Forest cover, (l) Shrub cover, (m) Crop cover, and (n) Grass cover of the GWNBR model. Grey pixels are non-significant coefficients. Source: Author.
Table 5. Coefficient estimates of explanatory variables of geographically weighted Poisson (GWPR) and negative binominal (GWNBR) models.

| Statistics | Model | \(b_{\text{Intercept}}\) | \(b_{\text{Elevation}}\) | \(b_{\text{Slope}}\) | \(b_{\text{Aspect}}\) | \(b_{\text{Humidity}}\) | \(b_{\text{Temperature}}\) | \(b_{\text{Precipitation}}\) | \(b_{\text{Population}}\) | \(b_{\text{NDVI}}\) | \(b_{\text{Forest cover}}\) | \(b_{\text{Shrub cover}}\) | \(b_{\text{Grass cover}}\) | \(b_{\text{Crop cover}}\) | \(b_{\text{Road density}}\) | \(b_{\text{Water density}}\) | \(b_{\text{Local dispersion}}\) |
|------------|-------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Mean       | GWPR  | -4.292          | -1.617          | 0.00569         | -0.3442         | 0.0286          | 0.3232          | -0.6032         | -3.966          | 1.3094          | -0.1415         | 0.0252          | -0.0137         | -5.6522         | -0.0375         | --             |
|            | GWNBR | -5.846          | -1.628          | 0.00003         | -0.3219         | 0.1005          | 0.0291          | 0.3119          | 0.7773          | -3.645          | 1.3393          | 0.0603          | 0.0107          | -0.0233         | -5.7854         | -0.1058         | 0.0565         |
| Median     | GWPR  | -5.659          | -1.760          | 0.00227         | -0.2787         | 0.1022          | 0.0501          | 0.3207          | -0.1382         | -3.880          | 1.2280          | 0.2621          | 0.0115          | 0.0000          | 0.0001          | --             |
|            | GWNBR | -5.902          | -1.707          | 0.00516         | -0.2473         | 0.1099          | 0.0507          | 0.2918          | -0.1055         | -3.667          | 1.3831          | 0.1507          | 0.0033          | 0.0278          | 0.0000          | -0.0995         | 0.0510         |
| Minimum    | GWPR  | -108.98         | -3.925          | 0.06112         | -2.0639         | -1.2975         | -0.5312         | -3.9636         | -5.6987         | -14.584         | -1.6778         | -2.3525         | -0.7401         | -4.2852         | -1.0281         | --             |
|            | GWNBR | -118.33         | -3.415          | 0.000643        | -1.6342         | -0.7575         | -0.4062         | -3.9332         | -2.4948         | -14.390         | -1.3371         | -0.2233         | -0.5752         | -12.345         | -0.8397         | 0.0190         |
| Maximum    | GWPR  | 111.264         | 2.809           | 0.0956          | 2.0505          | 1.1494          | 0.4068          | 3.0787          | 9.7859          | 4.926           | 5.4778          | 2.3102          | 0.8653          | 0.5377          | 301.45          | 1.0482         |
|            | GWNBR | 78.395          | 1.774           | 0.0863          | 1.0643          | 1.2201          | 0.4429          | 2.1058          | 11.4046         | 3.681           | 3.5797          | 1.7295          | 0.6013          | 0.4911          | 411.82          | 0.7017         | 0.1422         |
| The interquartile range (IQR) | GWPR  | 38.5508 | 0.9821 | 0.0313 | 0.6457 | 0.3947 | 0.2537 | 1.0017 | 1.3115 | 4.0899 | 1.7066 | 1.3304 | 0.2282 | 0.4003 | 3.2728 | 0.4229         |
|            | GWNBR | 31.5258 | 0.8528 | 0.0371 | 0.5888 | 0.3247 | 0.2188 | 0.9027 | 1.3347 | 3.6160 | 1.3479 | 1.0937 | 0.1580 | 0.3502 | 1.8188 | 0.4297 | 0.0248         |
| Standard deviation | GWPR  | 23.6224 | 0.9470 | 0.0286 | 0.5800 | 0.3486 | 0.1795 | 0.8532 | 2.2621 | 2.7867 | 1.4283 | 0.9207 | 0.1548 | 0.2734 | 29.089 | 0.3684         |
|            | GWNBR | 28.3133 | 0.7823 | 0.0300 | 0.4405 | 0.2816 | 0.1659 | 0.7859 | 2.3213 | 2.8390 | 1.1447 | 0.7916 | 0.2484 | 0.4541 | 0.3308 | 0.0217         |
| Standard error | GWPR  | 1.4903 | 0.0457 | 0.00203 | 0.0482 | 0.0157 | 0.00964 | 0.01767 | 0.0199 | 0.1553 | 0.0788 | 0.0415 | 0.0171 | 0.0232 | 0.056 | 0.0291 |
|            | GWNBR | 3.2821 | 0.088 | 0.00483 | 0.1086 | 0.0346 | 0.0215 | 0.036 | 0.0387 | 0.3788 | 0.183 | 0.0968 | 0.037 | 0.05 | 0.1314 | 0.0703 | 0.0204 |
| Status     | GWPR  | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS |
|            | GWNBR | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS |

†Note: Standard error was estimated from the global Poisson and NB models; NS indicates spatially nonstationary.
local parameters of each location and effectively explained the spatial variability of the response variable (Wu and Zhang 2013; Guo et al. 2016b, Rodrigues et al. 2014).

The spatial distribution of the local dispersion parameters is shown in Figure 7. The larger dispersion parameters were clustered along the southeast coast of Fujian province, where the forest fires occurred less than 10 times (Figure 3(a)). The local dispersion parameters then gradually decreased in the northwest areas. The existence of spatially varying overdispersion across the study area clearly suggested the needs to correct the model standard errors location by location. Although the overall prediction performances of the two GWR models were similar (Table 2), GWNBR was more appropriate for modeling the dispersed fire count data across the study area due to its lower mean prediction error (MPE) for the fire count categories 5–10 (Table 3).

4.2. Influence of drivers on forest fire occurrence in Fujian

Based on the global models, the results indicated that the number of forest fires in Fujian province increased at lower elevation, on flatter terrain, and near denser roads.
These findings are generally supported by the GWR models. The human activities more likely occurred in the areas of low elevation, flatter terrain, and dense road nets, which may lead to higher fire risks. However, according to the global models, it was found that areas with high population density had fewer forest fires, where human activity is also very frequent. This finding was consistent with the GWR models that showed a negative relationship between population density and forest fires in the eastern coastal areas. High population density tends to be concentrated in cities or developed areas with developed industry and low forest coverage, where the combustible material is not as continuous as the vegetation in remote forests and firefighting and prevention measures are more concentrated (Garcia et al. 1995; Maingi and Henry 2007; Miranda et al. 2012; Guo et al. 2016a, 2016c).

In addition, both global and GWR models similarly indicated that the aspect index (value from $-1$ to $1$) was negatively associated with forest fires. The smaller the aspect index was, the closer the region was to the south and southwest, with stronger sunshine, higher temperatures, and faster evaporation. These features lead to dryer fuel making it more likely for forest fires to occur (Hu 2005). Some research indicated that a higher NDVI meant a higher fuel load and higher probability and frequency of forest fires, but our study in Fujian province showed the opposite effect. One explanation for this is that in the areas with high vegetation coverage, the surface temperature is low, which makes it more difficult for soil moisture to evaporate, leading to higher fuel moisture content, and thus lower likelihood of burning (Huang et al. 2015).

The GWR models illustrated that the model coefficients of crop cover were positive in the north, but negative in the south. One possible explanation for this is that agricultural land is mainly distributed in the southeastern part of Fujian province. The forest coverage rate in these areas was relatively low, so the number of forest fires was negatively correlated with the crop area. Additionally, slope was negatively associated with fire occurrence in the north and had an opposite impact in the south. This may be due to higher grassland distribution in southern Fujian, as when the slope is large, grass becomes drier and may lead to more fires. As opposed to in the north, where there is more forested area, and since fewer human activities occur in high-slope areas, there are fewer human-caused fire sources.

The important of meteorological factors was also confirmed by this study. According to the global models, precipitation and relative humidity have positive impacts on fire occurrence over the study area. Although these findings seem contradictory to the general understanding that as rainfall increases, fewer fires will occur, one explanation for this is that more rainfall and humidity are beneficial to the growth of ground-cover vegetation, and increased amounts of surface fuel load, which increases the risk of forest fire occurrence. In the Kruger National Park of South Africa, van Wilgen et al. (2000) observed a strong positive correlation between precipitation rates and fire activity. Spessa et al. (2005) and Randerson et al. (2005) also found a similar positive association between precipitation and fire activity in north Australia using different satellite data sets. In contrast, the GWR models provided more specific spatial relationships between meteorological factors and fire occurrence. They found precipitation to be positively correlated with fire occurrence in the west.
of Fujian province, but negatively correlated in eastern coastal regions. Since there is plenty of rainfall in the coastal areas, the role of precipitation is more likely a limiting factor on fire occurrence rather than a promoting factor of fuel load, which may positively impact fire counts. The correlation between relative humidity and fire occurrence is also spatially varying, positive in the north and negative in the center of Fujian (Figures 5 and 6).

4.3. Research implementation and future research

Our study revealed that the GWR models, particularly GWNBR, were prominent for explaining more on the spatially varying relationships between forest fires and drivers, and those drivers were not consistently positively or negatively influencing the number of forest fires across Fujian province. Hopefully, our results will provide new insight into forest fire mapping, prevention, and management based on local environmental characteristics.

In this study, we identified and selected 14 explanatory variables as the significant driving factors which impacted on the number of forest fires in Fujian province. However, the phenomenon of forest fire occurrence is much more complex than that represented by the four statistical models, because those driving factors (including topographic, meteorological, human and vegetation variables) interact with each other in the nature, while a statistical model only represents a simplified, mathematically-formalized approximation for the complex relationships between forest fires and driving factors. Thus, given the regression models and results in this study, further studies using different statistical analyses and/or models may be considered to take the complexity of the forest fire occurrence and counts into account. For example, the interaction terms between some of those factors may be included and tested in the models. Multivariate analysis such as path analysis may be applied to study the correlations between forest fires and driving factors, as well as among these driving factors to identify dominant factors for influencing the occurrence and number of forest fires. Structural equation modeling (SEM) is a combination of factor analysis and multiple regression analysis which may be applied to analyze the structural relationships between forest fires and multiple driving factors. Based on our present study, the combined analyses of spatial models and SEM may be worthy of considering in the future.

5. Conclusion

In this study, the spatial varying relationships between forest fire occurrence and driving factors in Fujian province from 2001 – 2016 were evaluated using two global models (Poisson and NB) and two geographically weighted generalized linear models (GWPR and GWNBR). The results indicated that, compared to the global Poisson and NB models, the GWR generalized linear models had better performance in model fitting, predictions, and spatial distributions of model predictions, as well as detecting the impact hotspots of the predictor variables. Further, we compared the performance of GWPR and GWNBR and found that although the overall prediction performance
of the two GWR models were similar, GWNBR was more effective for fitting spatially over-dispersed forest fire count data because it accounted for overdispersion at different locations across the study area.

In addition, we determined the drivers and spatial distribution of subtropical forest fires in Fujian province, China based on the above methods. Two types of models (global and GWR) similarly indicated that the number of forest fire occurrences in Fujian increases at lower elevation, with flatter terrain, and near denser roads. Compared to the global models, the GWR models can indicate more specific spatial relationships between drivers and fire occurrence. For example, precipitation and population density had different impacts on fire occurrence in coastal areas than in other regions of the study area. In summary, GWPR and GWNBR were more effective and appropriate methods for analyzing the occurrence of spatially varied fire occurrence. GWNBR was particularly suitable for dealing with the over-dispersed forest fires. This type of model can clearly indicate the key fire drivers and their influence, and can provide reliable insight into forest fire mapping, prevention, and management based on local characteristics.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the National Key R&D Plan of Strategic International Scientific and Technological Innovation Cooperation Project (grant no. 2018YFE0207800); the AsiaPacific Network for Sustainable Forest Management and Rehabilitation (APFnet Climate Research Project) Phase II.

References

Amraoui M, Pereira MG, DaCamara CC, Calado TJ. 2015. Atmospheric conditions associated with extreme fire activity in the Western Mediterranean region. Sci Total Environ. 524-525: 32–39.

Burnham KP, Anderson DR. 2004. Multimodel inference: Understanding AIC and BIC in model selection. Sociol Methods Res. 33(2):261–304.

Cardille JA, Ventura SJ, Turner MG. 2001. Environmental and social factors influencing wildfires in the upper Midwest, United States. Ecol Appl. 11(1):111–127.2.0.CO;2

Chas-Amil ML, Prestemon JP, McClean CJ, Touza J. 2015. Human-ignited wildfire patterns and responses to policy shifts. Appl Geogr. 56:164–176.

Costafreda-Aumedes S, Comas C, Vega-Garcia C. 2017. Human-caused fire occurrence modelling in perspective: a review. Int J Wildland Fire. 26:983–998.

Chen VY, Deng WS, Yang TC, Matthews SA. 2012. Geographically weighted quantile regression (GWQR): an application to U.S. mortality data. Geogr Anal. 44(2):134–150.

Cunningham AA, Martell DL. 1973. A stochastic model for the occurrence of man-caused forest fires. Can J for Res. 3(2):282–287.

Da Silva AR, Rodrigues TCV. 2014. Geographically weighted negative binomial regression incorporating overdispersion. Stat Comput. 24(5):769–783.

Dayananda PWA. 1977. Stochastic models for forest fires. Ecol Modell. 3(4):309–313.
Fotheringham AS, Brunsdon C. 2010. Local forms of spatial analysis. Geogr Anal. 31(4): 340–358.
Fotheringham AS, Brunsdon C, Charlton M. 2002. Geographically weighted regression. New York: Wiley.
Fotheringham AS, Charlton ME, Brunsdon C. 1998. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. Environ Plan A. 30(11): 1905–1927.
Gill AM, Christian KR, Moore PHR, Forrester RI. 1987. Bush fire incidence, fire hazard and fuel reduction burning. Austral Ecol. 12(3):299–306.
Guo L, Ma Z, Zhang L. 2008. Comparison of bandwidth selection in application of geographically weighted regression: a case study. Can J for Res. 38(9):2526–2534.
Guo F, Innes LJ, Wang G, Ma X, Sun L, Hu H, Su Z. 2015. Historic distribution and driving factors of human-caused fires in the Chinese boreal forest between 1972 and 2005. JPECOL. 8(5):480–490.
Guo F, Su Z, Wang G, Sun L, Lin F, Liu A. 2016a. Wildfire ignition in the forests of southeast China: Identifying drivers and spatial distribution to predict wildfire likelihood. Appl Geogr. 66:12–21.
Guo F, Selvalakshmi S, Lin F, Wang G, Wang W, Su Z, Liu A. 2016b. Geospatial information on geographical and human factors improved anthropogenic fire occurrence modeling in the Chinese boreal forest. Can J Res. 46(4):582–594.
Guo F, Wang G, Su Z, Liang H, Wang W, Lin F, Liu A. 2016c. What drives forest fire in Fujian, China? Evidence from logistic regression and Random Forests. Int J Wildland Fire. 25(5):505–519.
Guo F, Wang G, Innes J L, Ma Z, Liu A, Lin Y. 2016d. Comparison of six generalized linear models for occurrence of lightning-induced fires in northern Daxing’an Mountains, China. J FOR Res. 27(2):379–388. doi:10.1007/s11676-015-0176-z.
Guo F, Su Z, Wang G, Sun L, Tigabu M, Yang X, Hu H. 2017. Understanding fire drivers and relative impacts in different Chinese forest ecosystems. Sci Total Environ. 605-606:411–425.
Haining R, Law J, Griffith D. 2009. Modeling small area counts in the presence of overdispersion and spatial autocorrelation. Comput Stat Data Anal. 53(8):2923–2937.
Hu HQ. 2005. Wildfire ecology and management. 2nd ed. Beijing: China Forestry Publishing House.
Huang B, Zhang H, Sun Z, Zhou L. 2015. Forest fire danger factors and their division in Shandong based on GIS and RS. Trans Chinese J Ecol. 34(5):1464–1472.
Justice CO, Giglio L, Korontzi S, Owens J, Morissette JT, Roy D, Descoitres J, Alleaume S, Petitcolin F, Kaufman Y. 2002. The MODIS fire products. Remote Sens Environ. 83(1-2): 244–262.
Koutsias N, Martínez-Fernández J, Allgöwer B. 2010. Do factors causing wildfires vary in space? Evidence from geographically weighted regression. GISci Remote Sens. 47(2): 221–240.
Maingi JK, Henry MC. 2007. Factors influencing wildfire occurrence and distribution in eastern Kentucky, USA. Int J Wildland Fire. 16(1):23–33.
Mandalaz D, Ye R. 1997. Prediction of forest fires with Poisson models. Can J Res. 27(10): 1685–1694.
McCulloch CE, Searle SR. 2001. Generalized, linear and mixed models. New York: John Wiley & Sons, Inc.; p. 1–184.
Miranda BR, Sturtevant BR, Stewart SI, Hammer RB. 2012. Spatial and temporal drivers of wildfire occurrence in the context of rural development in northern Wisconsin, USA. Int J Wildland Fire. 21(2):141–154.
Myers RH, Montgomery DC, Vining GG. 2002. Generalized linear models. New York: John Wiley & Sons; p. 100–194.
Nakaya T, Fotheringham AS, Brunsdon C, Charlton M. 2005. Geographically weighted Poisson regression for disease association mapping. Stat Med. 24(17):2695–2717.
Randerson JT, van der Werf GR, Collatz GJ, Giglio L, Still CJ, Kasibhatla P, Miller JB, White JWC, DeFries RS, Kasischke ES. 2005. Fire emissions for C3 and C4 vegetation and their influence on interannual variability of atmospheric CO2 and \( \delta^{13}\text{CO}_2 \). Global Biogeochem Cycle. 19:GB2019.

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.

Scott JH, Thompson MP, Calkin DE. 2013. A wildfire risk assessment framework for land and resource management. General Technical Report RMRS-GTR-315. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, p. 83.

Su Z, Tigabu M, Cao Q, Wang G, Hu H, Guo F. 2019. Comparative analysis of spatial variation in forest fire drivers between boreal and subtropical ecosystems in China. For Ecol Manage. 454:117669.

Spessa A, McBeth B, Prentice C. 2005. Relationships among fire frequency, rainfall and vegetation patterns in the wet-dry tropics of northern Australia: an analysis based on NOAA-AVHRR data. Global Ecology and Biogeography. 14(5):439–454.

Stocks BJ, Martell DL. 2016. Forest fire management expenditures in Canada: 1970–2013. For Chron. 92(03):298–306.

Terceiro M. 2003. The statistical properties of recreational catch rate data for some fish stocks off the northeast US coast. Fishery Bullet. 101:653–672.

van Wilgen BW, Biggs HC, O'Regan SP, Mare N. 2000. A fire history of the savanna ecosystems in the Kruger National Park, South Africa, between 1941 and 1996. South Africa J Sci. 96(4):167–178.

Garcia CV, Woodard PM, Titus SJ, Adamowicz WL, Lee BS. 1995. A logit model for predicting the daily occurrence of human caused forest fires. Int J Wildland Fire. 5(2):101–111.

Wotton BM, Martell DL, Logan KA. 2003. Climate change and people-caused forest fire occurrence in Ontario. Clim Change. 60(3):275–295.

Wotton BM, Nock CA, Flannigan MD. 2010. Forest fire occurrence and climate change in Canada. Int J Wildland Fire. 19(3):253–271.

Wu W, Zhang L. 2013. Comparison of spatial and non-spatial logistic regression models for modeling the occurrence of cloud cover in northeastern Puerto Rico. Appl Geogr. 37:52–62.

Xiao Y, Ju H, Zhang X, Ji P. 2011. Relationship between fire-danger weather and forest fire in Qiannan area. Scient Silvae Sin. 47:128–133 (in Chinese).

Zhen Z, Shao L, Zhang L. 2018. Spatial hurdle Models for predicting the number of children with lead poisoning. IJERPH. 15(9):1792.

Zheng B, Agresti A. 2000. Summarizing the predictive power of a generalizedlinear model. Statist Med. 19(13):1771–1781.