Prediction And Evaluation of Forest Fire In Yunnan of China Based On Geographically Weighted Logistic Regression Model

Wenhui Li
Yunnan Normal University

Quanli Xu (✉️ go2happiness@163.com)
Yunnan Normal University

Junhua Yi
Kunming Metallurgy College

Jing Liu
Yunnan Normal University

Research Article

Keywords: forest fire, Forest fire forecasting, Logistic regression model, Geographically weighted Logistic regression, Spatial heterogeneity

Posted Date: November 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-997415/v1

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Prediction and Evaluation of Forest Fire in Yunnan of China Based on Geographically Weighted Logistic Regression Model

Wenhui Li¹,³,⁴,⁵, Quanli Xu¹,³,⁴,⁵*, Junhua Yi²*, Jing Liu¹,³,⁴,⁵

¹*Faculty of Geography, Yunnan Normal University, Kunming 650500, China
²*Geomatics Engineering Faculty, Kunming Metallurgy College, Kunming 650033, China
³*GIS Technology Engineering Research Centre for West-China Resources and Environment of Educational Ministry, Kunming 650500, China
⁴*Center for Geospatial Information Engineering and Technology of Yunnan Province, Kunming 650500, China
⁵*Key Laboratory of Resources and Environmental Remote Sensing for Universities in Yunnan, Kunming 650500, China

*Corresponding author(s). E-mail(s): go2happiness@163.com; yjh20k@163.com

Abstract
Establishing an effective forest fire forecasting mechanism is the premise of scientific planning and management of forest fires and forest fire prevention. In recent years, the forest fire prediction mechanism has been one of the key areas of concern for the government forestry management departments and forestry researchers. One of them, is logistic regression (LR). It is a relatively frequent prediction probability model used in forest fire prediction and prediction in China and abroad for the past few years. However, with the gradual deepening of research, it is found that the logistic regression model fails to fully consider the spatial non-stationary relationship between forest fires and driving factors, which leads to poor fitting effect and low prediction accuracy of the model. But its extended counterpart, the Geographically weighted logistic regression (GWLR) model, takes into account the spatial correlation between model variables, and effectively improves the fitting ability and prediction accuracy of the model. Therefore, this paper compares the ability of the logistic regression model and the geographically weighted logistic regression model in terms of fitting ability and prediction accuracy in order to obtain the ability of the two models to predict forest fires in Yunnan Province. In this paper, the samples were divided into 60% training samples and 40% test samples, and repeated sampling was carried out 5 times for training. Variables that appeared in the training model for 3 or more times were used to construct the final LR and GWLR models. Finally, the models with better fitting ability and higher prediction accuracy were used to classify the fire risks in Yunnan Province. The results show that the geographically weighted logistic regression model is superior to the logistic regression model in terms of fitting effect and accuracy. The geographically weighted logistic regression model is more suitable for the data structure of forest fires in Yunnan Province and has better prediction ability. The AUC value of the geographically weighted logistic regression model is 0.902, and the prediction accuracy is 82.7%: The AUC value of logistic regression model was 0.891, and the prediction accuracy was 80.1%. Fully considering the spatial heterogeneity among model variables can, to some extent, predict forest fires more accurately. The fitting of the two models shows that the relative humidity, temperature, air pressure, sunshine hours, daily precipitation, wind speed, and other meteorological factors; Vegetation type; terrain factor; Population density, road network and other human activity factors become the cause of forest fires in Yunnan Province.

Keywords: forest fire; Forest fire forecasting; Logistic regression model; Geographically weighted Logistic regression; Spatial heterogeneity

1. Introduction
Forest fires are a particularly important ecological disturbance factor in the ecosystem of a forest. It not only further affects the renewal and succession of a forest, but also threatens the safety of human life as well as property. At present, with the drastic global climate change, extreme weathers are gradually increasing, resulting in frequent forest fires with increasingly large affected areas and difficult fire control (Marlon et al. 2012, Pan et al. 2018a, Westerling 2016, Yue et
Since the 1920s, western European countries began to study and forecast forest fires, among which Canada, the United States and the former Soviet Union achieved rapid development in the early forest fire prediction and forecast research (Shu & Tian 1997). Earlier papers, based on historical forest fire data to establish forest fire prediction model. Poisson regression model and negative binomial regression model were first used in forest fire prediction, and these two models were gradually applied in forest fire research in many countries and regions (Boubeta et al. 2015, Crosby 1954, Cunningham & Martell 1973, Dayananda 1977, Wotton & Martell 2005). China was relatively late to join this endeavor, which was mainly after the founding of New China in the 1950s. In terms of methods and models, it more or less borrowed that of the former Soviet Union and European countries, which it gradually developed. In the subsequent research on forest fires in China, Shu Lifu et al. divided forest fire prediction and prediction into fire risk weather forecast, forest fire occurrence forecast and forest fire behavior forecast, and also divided forest fire prediction and prediction methods into empirical method, mathematical statistics method and physical method (Shu et al. 1998). After the 1980s, forest fire prediction in China developed rapidly (Wu et al. 2018).

At present, after nearly 100 years of systematic research, the probability prediction model of forest fire occurrence has been constantly improved toward practical application. At present, the models established by scholars to predict and further assess the risk of forest fires are mainly divided into two categories, one of which is the mathematical statistical model. It mainly mines historical fire data by using mathematical and statistical methods, and establishes a spatio-temporal prediction model between the occurrence of forest fires and their driving factors. It is the most commonly used method of forest fire occurrence prediction. For example, Poisson regression model (Guo et al. 2010b), negative binomial regression model (Guo et al. 2010a), Logistic regression model (Guo et al. 2015a), geographically weighted logistic regression model (Zhang et al. 2014); The other is machine learning. Compared with traditional linear regression model methods, machine learning methods do not need to specify model structure in advance. It can deal, most often with a higher explanatory power, with unknown interactions and nonlinear functions. For example, random forest model (Pan et al. 2018b), artificial neural network (Bisquert et al. 2012), bayesian network model (Chen et al. 2021, GAO & LIAO 2017), maximum entropy model (Parisien et al. 2012, Renard et al. 2012, Yang et al. 2021). However, machine learning method cannot provide a clear relationship between the occurrence of forest fires and its driving factors. There is no clear expression. Therefore, machine learning has not been widely applied and promoted in forest fire prediction in China, which made the application of mathematical statistics model become the main method in forest fire prediction research in China. Among them, logistic regression model is the most widely used. For example, CAI Qijun analyzed and determined forest fire driving factors in Zhejiang Province based on Logistic regression model (Cai et al. 2020). Ma Wenyuan et al. compared logistic regression model and random forest model to determine forest fire driving factors and the spatial distribution pattern in Shanxi Province (Ma et al. 2020). Chen Dai et al. predicted the probability of forest fires occurrence in the Greater Khingan Mountains region and classified fire risk levels based on logistic regression model (Chen 2019). But in recent years, with more in-depth studies, researchers found some limitations to that model. It assumes that the spatial relationship between the dependent variable and the independent variable is stable, that is, the model parameter is a constant in the whole study area, and a parameter is applied to the whole study area. However, the spatial relationship between the occurrence of forest fires and the driving factors is not stable, so the global logistic regression model cannot solve the spatial instability relationship between the two. But geographically weighted regression (GWR) model, which can solve the problem of spatial nonstationarity, is beneficial to reduce the differences between the model and to some extent, improve the accuracy of the model. Based on this, foreign scholars began to expand geographically weighted regression model that is geographically weighted logistic regression model (GWLR) applied in the prediction of forest fire (Monjarás-Vega et al. 2020, Rodrigues et al. 2014). The application verification shows that the geo-weighted logistic regression model further improves the prediction accuracy (Guo et al. 2016a, Martínez-Fernández et al. 2013, Oliveira et al. 2014). Therefore, geographically-weighted logistic regression model is gradually applied in domestic studies on regional forest fire prediction, forest fire classification, determination of forest fire drivers and other aspects (Guo et al. 2017, Liang 2016, Liang et al. 2017, Peng et al. 2021). The question then becomes; is geographically weighted Logistic regression model applicable to any region? And should the fitting accuracy and prediction accuracy be higher than the traditional global logistic regression model? Based on the forest fire data of Yunnan Province from 2010 to 2020, this study combined the driving factors of forest fire such as topography, meteorology, vegetation and population density. The applicability of global logistic regression model and local geographically weighted logistic regression model to forest fire prediction in Yunnan Province and their respective prediction ability were discussed from the aspects of model fitting ability and prediction accuracy.

2. Materials and Methods
2.1 Overview of the study area

Yunnan province is located in 21°8’32”~29°15’8” N, 97°31’39”~106°11’47” E, located in Yunnan-Guizhou Plateau. The terrain presents a high northwest and low southeast ladder distribution. It belongs to the low latitude and high altitude area. The highest altitude is 6740 m, and the lowest is only 76.4 m. The temperature change caused by the huge altitude difference constitutes a drastic climate variation. Different climate environment is conducive to the growth of a variety of vegetation, which makes the forest resources in Yunnan Province extremely rich. The forest coverage rate is 55.7 %, which is the key forest area in China. At the same time, Yunnan Province is also the focus of forest fires in China. There are mainly Pinus yunnanensis, Pinus armandii, Cunninghamia lanceolata, Pinus kesiya and other dominant species in most areas of the province. These dominant species belong to coniferous forests, which are flammable species. In recent years, the base timber forest, aerial forest and Yangtze River shelter forest planted artificially are mostly pine, fir and eucalyptus forests, and they are also flammable species. In addition, the monsoon climate in Yunnan Province is obvious. During the same period of rain and heat, dry and wet seasons are distinct. The precipitation in dry season (November-April next year) accounts for only 15 % of the annual precipitation, and the annual precipitation is unevenly distributed in time and space. Spring and winter are the fire-prone seasons in most areas of Yunnan Province. The above two conditions, namely tree species and climate, combine makes Yunnan Province prone to frequent forest fires, and due to the diversity and complexity of inflammable vegetation, inflammable tree species dominant, topography, climate environment, forest distribution and production and living fires in Yunnan Province. It constitutes the characteristics of forest fire prevention in Yunnan Province, such as difficulty in forest fire prevention, arduous fire prevention work, and severe fire prevention situation.

2.2 Data source and processing

The main data of this study include Yunnan Province 2010-2020 satellite monitoring forest fire data points, Land use, vegetation type data, population density data, residential area data, road network data and daily meteorological data of Yunnan Province from 2010 to 2020, Yunnan Digital Elevation Model (DEM) terrain data.

2.2.1 Forest fire data

The forest fire data of Yunnan Province comes from the VIIRS375 meters active fire products from NASA, including the latitudes and longitudes of the fire point, the date and time of the fire point, the confidence of the fire point and the type of fire point. Through comprehensive fire point confidence and use of fire point and land use type data overlay analysis of two conditions to screen out the final actual forest fire data, a total of 4021 items are obtained.

In the establishment of forest fire probability model in Yunnan Province, a certain proportion of non-fire points are needed to participate in the modeling. According to previous research experience, the proportion of fire points and non-fire points is set to 1 : 1, and random creation of non-fire points must follow two rules: non-fire points must fall on the area of forest land use type. Non-fire points must be random in time and space. Finally, 8042 forest fire point data sum
of fire point and non-fire point) in Yunnan Province are formed. In this paper, the total number of samples is randomly divided into 60% (4826 forest fire sample data) training samples to construct LR model and GWLR model and 40% (3216 forest fire sample data) test samples to test the goodness of fit of the two models. At the same time, in order to avoid the influence of random distribution samples on the construction of the model, the samples were randomly divided into five groups and the experiment was repeated five times. Finally, the significant variables that appear at least three times or more in five experiments were selected as the full sample data for model fitting. This study will use spss25.0, spssa and MGWR2.2 software to fit the logistic regression model and the geographically weighted logistic regression model, and use ArcGIS10.2 software to visualize the parameter estimation coefficient and coefficient t test value of the geographically weighted logistic regression model.

2.2.2 Meteorological data

Meteorological data mainly come from the National Meteorological Science Data Center (http://data.cma.cn/), which includes 125 national meteorological stations in Yunnan Province and 27 surrounding provinces, including daily maximum temperature, daily minimum temperature and daily average temperature. Daily maximum surface temperature, daily minimum surface temperature and daily average surface temperature. Sunshine hours. 24 hours precipitation. Daily maximum pressure, daily minimum pressure, daily average pressure. The data of meteorological factors such as daily maximum wind speed, wind direction of daily maximum wind speed, daily average relative humidity, and minimum relative humidity of Eyre. By creating a Tyson polygon of meteorological station in ArcGIS10.2, the 8042 forest fire sample points were matched with the meteorological station points, and the sample points were corresponding to the particular meteorological station points based on the spatial position. Finally, Python was used to match the sample points with the meteorological data according to the meteorological station points and dates.

2.2.3 Topography and vegetation data

Terrain data mainly comes from geospatial data cloud (http://www.gscloud.cn/). Terrain data mainly includes digital elevation data, aspect and slope. Mainly through the elevation data obtained to calculate the slope and aspect data. The vegetation data used in the study is the vegetation type data of Yunnan Province, which is derived from the geospatial data cloud. According to the national classification standard, the secondary class of vegetation type data in Yunnan Province is reclassified into the first class by ArcGIS10.2, mainly including coniferous forest, broadleaf forest, shrub forest, grass, meadow, alpine vegetation, cultivated vegetation, coniferous and broad-leaved mixed forest.

2.2.4 The data of humanities

The humanistic data in this study mainly include the population density data, road network data, railway network data and residential area data of Yunnan Province from 2010 to 2020. The vector data points of each year’s fire samples and the corresponding population density raster data are superimposed and analyzed by ArcGIS10.2, and the population density values of each forest fire point are extracted by the point tool of value extraction. The residential area, highway network and railway network data are mainly calculated by using the nearest neighbor analysis tool in ArcGIS10.2 to calculate the nearest distance between each sample point and the residential area, railway and highway.

2.2.5 Data normalization

Since a wide range of data with different sources is involved in the study, the dimensions and grades used are also different, this will decrease the fitting degree of the model and make it impractical when the model is fitted later. So in order to eliminate the dimension between the data, the level of difference, and the difference between the data level, the numerical problems caused by the data must be normalized. However, the methods used by different data are different, including temperature, air pressure, ground temperature, sunshine hours, precipitation, wind direction, wind speed, and other meteorological factors. Population density, the distance between the fire and the highway, the distance from the railway, the distance from the residential area and other social infrastructure and human data. The vegetation type data and elevation data will be normalized by Formula (1). The slope data will be normalized by Formula (2). The relative humidity data will be normalized by Formula (3).

\[
X_i = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (1)
\]

In the formula, \(X_i\) represents the value after normalization, \(x\) represents the value before normalization, \(x_{\text{max}}\) and \(x_{\text{min}}\) represent the maximum and minimum values in this set of data.

\[
x_{\theta} = \sin \theta \quad (2)
\]

In the formula, \(x_\theta\) represents the normalized slope value, while \(\theta\) represents the slope value.
\[ x_σ = \frac{σ}{100} \]

In the formula, \( x_σ \) represents the normalized relative humidity, while \( σ \) represents the original relative humidity.

### 2.2.5 Multicollinearity test

Multicollinearity means that there is a certain degree or high correlation between explanatory variables in the linear regression model, which can lead to the loss of significance of variable significance test and a failure in the model prediction function. Therefore, the main purpose of the multicollinearity test is to determine the relevant driving factor model variables that lead to forest fires and determine the independent variables that eventually enter the model. Therefore, when the formula contains multiple independent variables and the correlation between variables has to be tested, multiple collinearity tests should be conducted on independent variables to exclude the factors with significant collinearity. In this paper, the variance inflation factor (VIF) is used to test the multiple collinearity of the driving factors of forest fire occurrence applied in the study. The multiple collinearity test method used in this study is the variance inflation coefficient. The calculation formula is as follows:

\[ VIF = \frac{1}{1 - R^2} \]

This coefficient is usually judged using 10 as the critical value. When VIF<10, there is no multicollinearity. When 10\(<VIF<100, there is strong multicollinearity. When VIF\(\geq100, severe multicollinearity exists (Chang et al. 2013, Guo et al. 2016b).

### 2.3 Research methods

#### 2.3.1 Binomial classification Logistic regression model

The Binary logistic regression (LR) refers to the binary logistic regression with dependent variables, which is the most widely used model in China and abroad to predict the probability of forest fires. In this model, the values of dependent variables are only 1 and 0, and the target prediction probability is between \([0, 1]\). In our paper, the binomial logistic regression model is used as the probability model of forest fire occurrence in Yunnan Province. Assuming that the probability of forest fire occurrence (\( Y = 1 \)) is \( P \), the probability of non-forest fire occurrence (\( Y = 0 \)) is \( 1 - P \), and the regression relationship between the probability of forest fire occurrence and the variables \( i = 1, 2, ..., n \) is as follows:

\[ p(Y=1)=\frac{e^{z}}{1+e^{z}} = \frac{1}{1+e^{-z}} \]

Among them,

\[ z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n \]

LR model obtained through logical transformation is as follows:

\[ \text{Logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n \]

Where, \( p \) is the probability of forest fire; \( \beta_0, \beta_1, \beta_2, ... \) is the parameter estimation coefficient of the model, which can be solved by maximum likelihood method (Del Hoyo et al. 2011, Garcia et al. 1995). \( X_1, X_2, ..., X_n \) is the independent variable affecting the occurrence of forest fire.

#### 2.3.2 Geographically weighted Logistic regression
Geographically Weighted Logistic Regression (GWLR) is an extension of the traditional logistic regression model, which considers spatial nonstationarity (spatial correlation and spatial heterogeneity). The model uses the weighted least squares method to estimate the parameters of each coordinate point. The estimation of the parameters is local rather than global, and each position has a corresponding parameter estimation coefficient (Guo et al. 2015b, Xiao et al. 2013).

The expression of geographically weighted logistic regression model (GWLR) is the same as that of the logistic model (LR). The probability of forest fire occurrence (Y = 1) in position i is p, and the probability of no forest fire occurrence (Y = 0) is (1 − p). The regression relationship between the probability of forest fire occurrence in position i and the variables (i = 1,..., n) is as follows:

\[
P(Y = 1) = \frac{\exp \left( \hat{\beta}_{0'u,v'} + \hat{\beta}_{1'u,v'} X_{1'i} + \hat{\beta}_{2'u,v'} X_{2'i} + ... + \hat{\beta}_{n'u,v'} X_{ni} \right)}{1 + \exp \left( \hat{\beta}_{0'u,v'} + \hat{\beta}_{1'u,v'} X_{1'i} + \hat{\beta}_{2'u,v'} X_{2'i} + ... + \hat{\beta}_{n'u,v'} X_{ni} \right)}
\]

(8)

Among them,

\[
z = \hat{\beta}_{0'u,v'} + \hat{\beta}_{1'u,v'} X_{1'i} + \hat{\beta}_{2'u,v'} X_{2'i} + ... + \hat{\beta}_{n'u,v'} X_{ni}
\]

(9)

GWLR model obtained by logical transformation is as follows:

\[
\text{Logit}(p) = \ln \left( \frac{p}{1-p} \right) = \beta_{0(u,v)} + \beta_{1(u,v)} X_{1i} + \beta_{2(u,v)} X_{2i} + ... + \beta_{n(u,v)} X_{ni}
\]

(10)

Where, \((u_i,v_i)\) in Formula (8) is the geographic coordinates of fire point I; \(X_{1i}, X_{2i}, ..., X_{ni}\) is the independent variable; The estimated coefficient of the local regression model for position I can be calculated using weighted least square method.

2.3.3 Model prediction accuracy evaluation

This paper evaluates the fitting ability and prediction accuracy of the LR model and the GWLR model according to the receiver operating characteristic (ROC) curve analysis (Fielding & Bell 1997). The area under the curve (AUC) is the measurement standard of the fitting ability and prediction accuracy of the model. The larger the AUC, the better the fitting effect of the model (Del Hoyo et al. 2011, Zhou et al. 2009). In addition, according to the sensitivity and specificity calculated by ROC curve analysis method, the best critical value (Haden index) can be obtained, that is, Haden index = sensitivity value + specificity value - 1. If the probability of forest fires calculated by the model is greater than the optimal critical value, it is judged as Forest fires likely to happen, and if it is less than the critical value, it is judged as Forest fires unlikely to happen (Catry et al. 2009). Finally, the number of forest fires calculated is compared with the actual value, and the accuracy of logistic regression model (LR) and geographically weighted logistic regression model (GWLR) for judging whether forest fires occur is calculated.

2.3.4 Classification of fire risk classes

According to the predicted probability values of LR model and GWLR model, the spatial distribution of forest fire probability in Yunnan Province was visualized by the empirical Bayesian Kriging interpolation method in ArcGIS10.2 software, and the spatial distribution map of forest fire probability in Yunnan Province was obtained. Secondly, according to previous studies, the threshold for classification of fire risk grade is 0.5 (Zhang et al. 2013). In this paper, the classification of fire risk grade in Yunnan region is mainly based on the same threshold (0.5) determined by previous studies and the Harden index (optimal critical value, cut-off) calculated by ROC curve. The main classification rules are as follows: (1) The predicted probability value (P) of the model is <0.5, indicating low grade fire danger zone; (2) 0.5<P< Cut-off, Medium grade fire danger zone; (3) P> cut-off, high grade fire danger zone. Finally, the spatial distribution of fire hazards of different grades in Yunnan province is divided according to this rule.

3. Results and analysis

3.1 Multicollinearity test results

According to the results of multicollinearity diagnosis, VIF of seven meteorological variables, including daily mean pressure, daily minimum pressure, daily mean air temperature, daily minimum air temperature, daily mean surface temperature, daily minimum surface temperature, daily mean wind speed and so on, are all greater than 10, indicating that they all have a collinearity relationship. After removing these eight variables, what remains is average relative humidity, minimum relative humidity, sunshine time, daily highest temperature, daily maximum pressure, precipitation,
24 hours a day, the biggest manners, the daily maximum surface temperature, population density, slope, slope to fire recently, altitude, vegetation type, settlement to zero distance, railway, highway to point nearest to the point in recent distance 16 independent variables (table 1) LR model and GWLR model.

### Table 1 Multicollinearity diagnosis results of model variables

| Serial number | Model variables                        | Before eliminating variables | After eliminating variables |
|---------------|----------------------------------------|------------------------------|-----------------------------|
| 1             | Average daily relative humidity        | 8.796                        | 4.48                        |
| 2             | Minimum daily relative humidity        | 13.283                       | 5.915                       |
| 3             | 24 - hour sunshine hours               | 3.425                        | 3.062                       |
| 4             | Daily mean atmospheric pressure        | 45.065                       | -                           |
| 5             | Daily maximum pressure                | 2530.865                     | 2.649                       |
| 6             | Diurnal minimum                       | 2552.484                     | -                           |
| 7             | Daily mean temperature                | 117.726                      | -                           |
| 8             | Daily maximum temperature             | 39.719                       | 4.235                       |
| 9             | Daily minimum temperature             | 50.894                       | -                           |
| 10            | 24-hour precipitation                 | 1.287                        | 4.235                       |
| 11            | Daily mean wind speed                 | 2.922                        | -                           |
| 12            | Daily maximum wind speed              | 2.639                        | 1.109                       |
| 13            | Daily mean surface temperature        | 52.352                       | -                           |
| 14            | Daily maximum surface temperature     | 12.871                       | 4.055                       |
| 15            | Daily minimum surface temperature     | 36.910                       | -                           |
| 16            | Population density                   | 1.011                        | 1.01                        |
| 17            | Vegetation types                      | 1.009                        | 1.014                       |
| 18            | Slope                                 | 1.039                        | 1.027                       |
| 19            | Aspect                                | 1.004                        | 1.007                       |
| 20            | Altitude                              | 2.531                        | 2.494                       |
| 21            | Nearest distance from residential area to fire point | 1.150 | 1.263 |
| 22            | Nearest distance from road to fire point | 1.067 | 1.114 |
| 23            | Nearest distance from railway to fire point | 1.111 | 1.114 |

3.2 Logistic model fitting results

In this paper, the stepwise regression method and the Logistic model are used to fit and calculate the five groups of training samples. After fitting, five different subsets of characteristic variables are obtained (Table 2), and the independent variables that appear at least three times or more in the five groups of characteristic variables are selected to fit the full sample data of the Logistic model.

### Table 2. The subset results of characteristic variables of Logistic model

| Model independent variables | Training samples 1 group | Training samples 2 groups | Training samples 3 groups | Training samples 4 groups | Training samples 5 groups | Sample significance |
|-----------------------------|--------------------------|----------------------------|---------------------------|----------------------------|--------------------------|---------------------|
| Average daily relative humidity | ✓                        | ✓                          | ✓                         | ✓                          | ✓                        | 5                   |
Table 2 shows that 11 independent variables, such as daily average relative humidity, daily minimum relative humidity, sunshine hours, daily maximum temperature, daily maximum pressure, daily maximum surface temperature, population density, altitude, the nearest distance between residential area and fire point, the nearest distance between road and fire point, and the nearest distance between railway and fire point, will enter the final full sample model fitting stage (Table 3). Among them, 8 variables, such as daily average relative humidity, daily minimum relative humidity, sunshine hours, daily maximum pressure, altitude, daily maximum surface temperature, the nearest distance between residential area and fire point, and the nearest distance between railway and fire point, appear in 5 sample experiments. The highest temperature appeared four times a day. Population density and the nearest zero distance from road to fire occurred three times. In addition, the five variables of 24 - hour precipitation, maximum wind speed, slope, aspect and vegetation type will not enter the final full-sample fitting data, because they appear less than three times in the five sample experiments, among which 24 - hour precipitation once, maximum wind speed and slope twice, aspect and vegetation type 0 times.

### Table 3 Logistic model full sample data fitting results

| Model independent variables | Parameter coefficient | Standard error | Significance |
|----------------------------|-----------------------|----------------|--------------|
| Constant value             | 0.343                 | 0.112          | 0.002<0.01   |
| Average daily relative humidity | -0.009             | 0              | 0.000<0.01   |
| Minimum daily relative humidity | -0.007               | 0.001          | 0.000<0.01   |
| 24 - hour sunshine hours | 0.021                 | 0.002          | 0.000<0.01   |
| Daily maximum pressure | 0.001                 | 0              | 0.000<0.01   |
| Daily maximum temperature | 0.004                 | 0.001          | 0.003<0.01   |
| Daily maximum surface temperature | -0.003            | 0.001          | 0.000<0.01   |
| Population density | -0                    | 0              | 0.000<0.01   |
| Altitude | -0                    | 0              | 0.000<0.01   |
| Nearest distance from residential area to fire point | 2.856                | 0.374          | 0.000<0.01   |
| Nearest distance from road to fire point | -0.394              | 0.128          | 0.002<0.01   |
| Nearest distance from railway to fire point | 0.019                | 0.006          | 0.002<0.01   |

Table 3 shows that the variable appears in the model, and \( \times \) indicates that the variable is not in the model. It can be seen from Table 3 that five variables, including sunshine duration, daily maximum pressure, daily maximum temperature, the nearest distance from residential area to fire point, and the nearest distance from railway to residential area, are positively correlated with the occurrence of forest fire. In addition, six variables, including average relative
humidity, minimum relative humidity, daily maximum surface temperature, population density, altitude, and the nearest distance from highway to fire point, are negatively correlated with the occurrence of forest fire. Finally, using a p value of less that 0.001, it is determined that all these 11 variables have a significant impact on the occurrence of forest fire.

3.3 GWLR model fitting results

Through the application of variance expansion factor VIF test before the daily average relative humidity, daily minimum relative humidity, sunshine hours, daily maximum temperature, daily maximum pressure, 24-hour precipitation, maximum wind, daily maximum surface temperature, population density, slope, aspect, vegetation type, residential area to the nearest fire distance, road to the nearest zero distance, railway to the nearest fire distance and other 16 independent variables (Table 1) into the GWLR model fitting. The stability test of the sample assumes that the dependent variable and the independent variable have spatial stability characteristics. After fitting the five groups of training samples, the spatial non-stationarity of the spatial relationship between the dependent variable and the independent variable is tested. The main discriminant principle is that the quartile range of the estimation coefficient of an independent variable is greater than the ± 1 standard deviation range of the estimation coefficient of the independent variable in the global logistic regression model (LR) (Martínez-Fernández et al. 2013, Zhang et al. 2014). It is determined that the independent variable has significant spatial nonstationarity. The same principle is that if the variable has at least three or more spatial nonstationarity in five training samples, it enters the whole sample data fitting stage of the GWLR model. The test results show that the nearest distance from the residential area to the fire point in all the variables is spatially stable in the five training samples groups. The sunshine hours and the nearest distance from the railway to the fire point are otherwise spatially stable Only one time. The other samples are all spatially non-stationary variables in the five intermediate models (Table 4), which are all included in the full sample fitting data of GWLR.

| Model independent variables | Training sample 1 group | Training sample 2 group | Training sample 3 group | Training sample 4 group | Training sample 5 group | Sample significance |
|-----------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|---------------------|
| Average daily relative humidity | √                       | √                       | √                        | √                        | √                        | 5                   |
| Minimum daily relative humidity | √                       | √                       | √                        | √                        | √                        | 5                   |
| 24-hour sunshine hours | √                       | √                       | ×                        | √                        | √                        | 4                   |
| Daily maximum pressure | √                       | √                       | √                        | √                        | √                        | 5                   |
| Daily maximum temperature | √                       | √                       | √                        | √                        | √                        | 5                   |
| 24-hour precipitation | √                       | √                       | √                        | √                        | √                        | 5                   |
| Daily maximum wind speed | √                       | √                       | √                        | √                        | √                        | 5                   |
| Daily maximum surface temperature | √                       | √                       | √                        | √                        | √                        | 5                   |
| Population density | √                       | √                       | √                        | √                        | √                        | 5                   |
| Slope | √                       | √                       | √                        | √                        | √                        | 5                   |
| Altitude | √                       | √                       | √                        | √                        | √                        | 5                   |
| Nearest distance from residential area to fire point | ×                       | ×                       | ×                        | ×                        | ×                        | 0                   |
| Nearest distance from road to fire point | √                       | √                       | √                        | √                        | √                        | 5                   |
| Nearest distance from railway to fire point | √                       | √                       | √                        | √                        | ×                        | 4                   |
| Aspect | √                       | √                       | √                        | √                        | √                        | 5                   |
| Vegetation types | √                       | √                       | √                        | √                        | √                        | 5                   |

Note: √ denotes that the variable is nonstationary in space, × denotes that the variable is stationary in space.

From the fitting results of the GWLR model parameters (Table 5), it can be seen that among all the variables, the daily maximum temperature, 24-hour precipitation, vegetation type, slope direction, and the nearest distance from the railway to the fire point show positive or negative correlation changes in the whole region. The three variables, namely,
sunshine duration, daily maximum pressure and maximum wind speed were positively correlated with the occurrence of forest fire. Seven variables, including average relative humidity, minimum relative humidity, daily maximum surface temperature, population density, slope, altitude, and the nearest distance from highway to fire point, were negatively correlated with the occurrence of forest fire in the study area. In order to better reflect the changes of variable parameter estimation coefficients in the study area, Kriging interpolation is used to visualize the variable coefficients in space (Figure 2). Figure 2 mainly shows the change of the estimated GWLR model coefficient in local space, indicating that the variables have a certain spatial heterogeneity, and the spatial relationship between the independent variables and the occurrence of forest fire is a complex spatial non-stationary relationship.

**Table 5** Parameter fitting results of GWLR model full sample data model.

| Variation coefficient | Model parameter | Minimum value | Quarter quantile | Mean value | Median | Three-quarters quantile | Maximum value |
|-----------------------|-----------------|---------------|------------------|-----------|--------|------------------------|---------------|
| Intercept term        |                 | -0.9552       | -0.6720          | -0.5402   | -0.5019| -0.4009                | -0.0263       |
| Average daily Relative humidity |     | -0.9916       | -0.8617          | -0.7205   | -0.6727| -0.6081                | -0.5362       |
| Minimum daily Relative humidity |       | -1.3720       | -1.2620          | -1.2138   | -1.2055| -1.1604                | -0.5634       |
| 24-hour sunshine hours |                 | 0.2378        | 0.4770           | 0.6093    | 0.5910 | 0.7287                 | 0.8488        |
| Daily maximum pressure |                 | 0.2859        | 0.4662           | 0.5501    | 0.5131 | 0.6375                 | 0.7210        |
| Daily maximum temperature |             | -0.0628       | 0.0057           | 0.0663    | 0.0761 | 0.1239                 | 0.1687        |
| 24-hour precipitation |                 | -0.9641       | -0.6235          | -0.5259   | -0.4860| -0.4104                | 0.0777        |
| Daily maximum wind speed |               | 0.0041        | 0.0660           | 0.1146    | 0.0984 | 0.1661                 | 0.2320        |
| Daily maximum surface temperature | | -0.2337       | -0.1534          | -0.1031   | -0.0726| -0.0572                | -0.0302       |
| Population density |                 | -0.6021       | -0.5315          | -0.4989   | -0.5021| -0.4754                | -0.0669       |
| Vegetation types |                 | -0.0046       | 0.0017           | 0.0150    | 0.0162 | 0.0225                 | 0.0557        |
| Slope |                 | -0.1160       | -0.1061          | -0.0914   | -0.0908| -0.0802                | -0.0530       |
| Aspect |                 | -0.0472       | -0.0131          | 0.0014    | 0.0021 | 0.0169                 | 0.0323        |
| Altitude |               | -0.4268       | -0.3010          | -0.2508   | -0.2185| -0.1919                | -0.1190       |
| Nearest distance from road to fire point | | -0.1417       | -0.1332          | -0.1187   | -0.1217| -0.1090                | -0.0610       |
| Nearest distance from railway to fire point | | -0.0150       | 0.0655           | 0.1286    | 0.1553 | 0.1883                 | 0.2286        |

**Note**: If the maximum and minimum values in the coefficient are positive or negative, indicating that the correlation between the variable and forest fire is consistent throughout the study area, and vice versa.
Fig. 2: Spatial heterogeneity distribution of variable parameter estimation coefficients: (a) constant term; (b) Daily relative humidity; (c) Minimum daily relative humidity; (d) 24-hours sunshine; (e) Daily maximum pressure; (f) Daily maximum temperature; (g) 24-hour precipitation; (h) Daily maximum wind speed; (i) Daily maximum surface temperature; (j) Population density; (k) Vegetation types; (l) Slope; (m) Aspect; (n) Digital Elevation Model; (o) Nearest distance from road to fire point; (p) Nearest distance from railway to fire point.

In addition, in order to better represent the significant spatial heterogeneity estimated GWLR model coefficient in local space, the t-test value is also spatially visualized through Kriging interpolation (Fig. 3). Fig. 3 mainly shows that the significance of the estimated coefficients of model variables also has a strong spatial heterogeneity in the study area. According to Table 5 and Fig. 3, seven variables, including average relative humidity, minimum relative humidity, daily maximum surface temperature, population density, slope, altitude, and the nearest distance from highway to fire point, were negatively correlated with the occurrence of forest fire in the whole study area and the correlation was statistically significant. Among them, four variables, including altitude, 24-hour precipitation, daily maximum surface temperature, and daily minimum relative humidity, were significantly negatively correlated with the occurrence of forest fire in the western part of the study area, Dehong Prefecture, Baoshan City, Diqing Prefecture, Lincang City, Nujiang Prefecture, Pu’er City, Dali Prefecture, Xishuangbanna Prefecture, and Lijiang City. The three variables of sunshine hours, daily maximum pressure and maximum wind speed were positively correlated with the occurrence of forest fire in the whole study area, and they were mainly positively correlated in Wenshan, Qujing, Zhaotong, Honghe and Yuxi. The slope aspect, daily maximum temperature, vegetation type and the nearest distance from railway to fire point are positively and negatively correlated in the study.
Fig. 3 Spatial Distribution of significance of variable coefficient. (a) constant term; (b) Daily relative humidity; (c) Minimum daily relative humidity; (d) 24-hours sunshine; (e) Daily maximum pressure; (f) Daily maximum temperature; (g) 24-hour precipitation; (h) Daily maximum wind speed; (i) Daily maximum surface temperature; (j) Population density; (k) Vegetation types; (l) Slope; (m) Aspect; (n) Digital Elevation Model; (o) Nearest distance from road to fire point; (p) Nearest distance from railway to fire point. Cold (blue) indicates significant negative correlation and warm (red) indicates significant positive correlation.

3.4 Model evaluation

The area under the ROC curve (AUC) is an important indicator for judging the accuracy and fitting effect of the model. At present, it is widely used in China and abroad. Studies have shown that the AUC range is generally between 0.5 and 1.0. When 0.5 < AUC ≤ 0.6, the accuracy is weak, general when 0.6 < AUC ≤ 0.7, moderate when 0.7 < AUC ≤ 0.8, high when 0.8 < AUC ≤ 0.9 and extremely high when AUC > 0.9. While when AUC is 0.5, it indicates that the method has no diagnostic value, that is, it also indicates that the model has low accuracy, poor fitting effect, no practical significance, or is not applicable in the research area. Table 6 shows the fitting effect and prediction accuracy of the logistic regression model and the geographically weighted logistic regression model in the fitting process of five sample groups and full samples. Table 6 shows that the AUC value of geographically weighted logistic regression model is 0.902, and the prediction accuracy is 82.7%. The AUC of logistic regression model was 0.891, and the prediction accuracy was 80.1%. At the same time, combined with Fig. 4, it can be more accurately concluded that the fitting effect of GWLR model is better than that of LR model thereby showing that the geographically weighted logistic regression is better than the logistic regression model in the fitting effect and prediction accuracy. Therefore, the geographically weighted logistic regression is more suitable for the prediction of forest fires in Yunnan Province.
### Table 6 Evaluation of LR Model and GWLR Model

| Training sample group | Model  | AUC   | Best cut-off point | Rate of correctly prediction |
|-----------------------|--------|-------|--------------------|-----------------------------|
| Trainingsample group  |        |       |                    |                             |
| Training              | LR     | 0.893 | 0.650              | 80.3%                       |
| LR                    | 0.890  | 0.648 |                    | 81%                         |
| GWLR                  | 0.907  | 0.674 |                    | 80.5%                       |
| LR                    | 0.891  | 0.646 |                    | 80.2%                       |
| GWLR                  | 0.902  | 0.663 |                    | 80.7%                       |
| LR                    | 0.894  | 0.651 |                    | 80.1%                       |
| GWLR                  | 0.908  | 0.678 |                    | 81%                         |
| LR                    | 0.891  | 0.645 |                    | 80%                         |
| GWLR                  | 0.904  | 0.670 |                    | 81.1%                       |
| Training sample 5 group |        |       |                    |                             |
| LR                    | 0.891  | 0.644 |                    | 80.1%                       |
| GWLR                  | 0.902  | 0.660 |                    | 82.9%                       |

3.4 Forest fire risk probability distribution and fire risk classification in Yunnan Province

According to the comparison results of fitting ability and prediction accuracy of the LR model and the GWLR model, and finally based on the prediction results of geographically weighted logistic regression model, the spatial distribution of forest fire probability in Yunnan Province was interpolated and analyzed by using the empirical Bellekin interpolation tool in ArcGIS10.2, and the spatial distribution map of forest fire probability in Yunnan Province was obtained (Fig. 4a). According to the default threshold 0.5 of the GWLR model and the optimal cut-off value (Cut-off) of the prediction probability of forest fires in Yunnan Province calculated by Haden index, the fire risk classification of Yunnan Province was carried out; low fire risk level area for GWLR model predicted probability value \( P < 0.50 \), medium fire risk level area for \( 0.50 \leq P < 0.660 \), and high fire risk level area for \( P \geq 0.660 \) (Figure 4b). According to Fig. 4, it can be seen that the spatial distribution of forest fire risk probability and fire risk level in Yunnan Province is very obvious. Among
them, the extremely high-risk areas and high-risk areas are mainly distributed in Honghe Prefecture, Wenshan Prefecture and Lijiang Prefecture. Other southern and northwestern in Yunnan Province, Nujiang and Dali prefectures are also scattered. Secondly, middle-risk areas are mainly distributed in the south, northwest and central Yunnan. Areas with low and extremely low risk levels are mainly distributed in Diqing Prefecture, Zhaotong City, Baoshan City and Chuxiong City. From the overall layout of Yunnan Province, the probability of forest fires in Yunnan Province is mainly in the southeast, south and northwest of the disaster, and the probability of occurrence in the northwest and central regions is low, and the fire risk level is low.

Fig.5 Probability distribution and classification of forest fire risk in Yunnan Province based on GWLR models. a is the predicted probability value of the GWLR model, b is the fire risk level map of GWLR model.

4. Discussion

In this paper, the traditional global logistic regression model and the local geographically weighted logistic regression model are used to analyze the forest fire data of Yunnan Province from 2010 to 2020. In the comparative analysis of the fitting ability and prediction accuracy of the two models, it is found that the meteorological factors such as daily average relative humidity, daily maximum pressure, daily maximum temperature, sunshine duration, daily maximum surface temperature, and the factors such as altitude, population density, distance from highway to fire point, and distance from railway to fire point are the main explanatory variables of the two models, which also shows that these driving factors constitute the main contributors to forest fire in Yunnan Province. In addition, through research, it is found that the traditional global model assumes that the explanatory variables of the model are stable in space, while the local geographically weighted logistic regression model considers the spatial non-stationary relationship of the model variables, that is, the explanatory variables of each fire have corresponding parameter coefficients, rather than using the parameter value from the global model applied to all prediction for different regions. Therefore, to a certain extent, the local model should have better prediction accuracy and model fitting rate than the global model. The results of this study show that the local model considering the spatial relationship between forest fire drivers does predict the occurrence of fires more accurately. Compared with the global logistic regression model, the local geographically weighted logistic regression model can explain the spatial relationship of model variables better, and has higher accuracy and fitting ability. Therefore, because similar environmental variables may have different contributions to forest fires in different regions, it is necessary to consider the complexity and heterogeneity of the regional space in future studies on forest fires and fire risk assessment.

By comparison, it is found that the difference in the prediction accuracy and fitting ability between the global model and the local model in Yunnan is not large. Since the geographically weighted extended model geographically weighted logistic regression model is affected by the selection of kernel function and bandwidth in the process of fitting, in our paper, the adaptive double square kernel function and bandwidth used in the geographically weighted logistic regression model according to the previous fitting of geographically weighted logistic regression model are applied. However, it is well known that the kernel function and bandwidth have a significant impact on the fitting and prediction of the model and the spatial distribution of the model parameter estimation coefficient. Therefore, one of the limitations in this paper is choosing the best kernel function and bandwidth of the geographically weighted logistic model.

According to the fire risk zoning map of Yunnan Province, the forest fires in Yunnan Province are mainly distributed in Wenshan Prefecture and Xishuangbanna Prefecture in the south and southeast. Lijiang and Nujiang prefectures in the
north and northwest. The main cause of high-risk grade fires in southern Yunnan Province is whether the southern part of Yunnan Province belongs to the geothermal valley area, and a part of the region belongs to the tropical area, in November to April of the following year, the climate is dry and hot with less precipitation, and Xishuangbanna, Wenshan and other cities are the main forest areas in Yunnan Province, so fires occur frequently. At the same time, according to the research results, it is found that Kunming, Qujing and other population concentration areas and other simple forest coverage and larger areas are not high or high risk areas of forest fire. Therefore, it shows that human activity such as population density and residential areas are not enough to lead to frequent forest fires. It is rather multivariable. Therefore, future researches on contributors to forest fire, it is necessary to take into account the combined effect of multiple data such as natural factors, meteorological factors and human factors, and combine them with a strong fitting ability, so that the model with prediction accuracy can more reasonably reveal the mechanism of forest fire occurrence and carry out scientific forest fire prediction.

By comparing the prediction ability of the two models for the probability of forest fire in Yunnan Province, it can be seen that, to a certain extent, the model considering the spatial heterogeneity between variables is more in line with the real fire situation, and can improve the accuracy of forest fire prediction. However, these models do not reveal the complex mechanism of forest fires, which can only explain the types of contributors to forest fires, or the weight of certain fire driving variables. So whether the future research on forest fire prediction can be focus on the following two aspects: firstly, using certain geographical methods or principles to solve the spatial relationship between model explanatory variables, and combining it with the corresponding mathematical model or machine learning method to form a compound mechanism model to promote its better fitting ability and higher prediction accuracy. Secondly, most of the current studies are trying to find the main driving factors or types of driving factors for forest fires in a certain region, such as meteorological factors and human factors. There are few studies on the contribution of a single driving factor to the occurrence of forest fires, or the most likely driving factor to cause forest fires in a certain threshold range. So in the future forest fire forecast research, whether can study from these two starting points, make it more conducive to reveal the complex mechanism of forest fire process and evolution process.

5. Conclusions

Scientific and reasonable forest fire risk analysis and zoning plays an important role in preventing forest fires. This study is based on the forest fire monitoring data of Yunnan Province from 2010 to 2020, combined with meteorological data such as relative humidity data, sunshine and precipitation. Slope, aspect, elevation and other terrain data. Human data such as population density, residential area, road network and vegetation type data are fitted by logistic regression model and geographically weighted logistic regression model, and the fitting ability and prediction accuracy of the two models are compared and analyzed. The forest fire probability model of Yunnan Province is constructed, and the fire risk level is divided. The results show that:

In the classification and discrimination of forest fires in Yunnan Province, compared with the traditional global logistic regression model, the local geographically weighted logistic regression model has better fitting effect and higher prediction accuracy. It is more suitable for the data structure of forest fires in Yunnan Province, and has more reference value in the fire prevention work in Yunnan Province.

The meteorological factors such as relative humidity, temperature, air pressure, sunshine hours, daily precipitation, wind speed and surface temperature were determined by fitting the LR model and GWLR model. Vegetation type, terrain factor. Human activity factors such as population density, roads and railways will become the incentives for forest fires in Yunnan Province.

It is determined that, in the future, the spatial relationship between model explanatory variables should be fully considered in the process of forest fire prediction, mainly the heterogeneity and complexity of variables in space, which is beneficial to improve the accuracy of forest fire prediction and scientific fire prevention.

Author Contributions Conceptualization, W.L, Q.X., J.Y. and J.L.; methodology, W.L, Q.X. and J.Y.; software, W.L. and J.L.; validation, W.L., Q.X. and J.Y.; formal analysis, W.L. and Q.X.; investigation, W.L. and J.Y.; resources, Q.X.; data curation, W.L. and J.L.; writing—original draft preparation, W.L.; writing—review and editing, Q.X. and J.Y.; visualization, W.L. and J.L.; supervision, Q.X. and J.Y.; funding acquisition, Q.X. and J.Y.; All authors have read and agreed to the published version of the manuscript.

Funding This research was funded by the National Natural Science Foundation of China (grant number 42161065, 41461038), Yunnan Province Science and Technology Fundamental Special Key Project (grant number 202001AS070032), Yunnan Province Young and Middle-aged Academic and Technical Leaders Reserve Talent Project (grant number 202105AC160059), Yunnan Normal University Graduate Research Innovation Fund (grant number YJSJJ21-B80)

Availability of data and materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations
Ethics approval: Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Competing interests: The authors declare no competing interests.

Reference:

Bisquert M, Caselles E, Sánchez JM, Caselles V (2012): Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. International Journal of Wildland Fire 21, 1025-1029

Boubeta M, Lombardía MJ, Marey-Pérez MF, Morales D (2015): Prediction of forest fires occurrences with area-level Poisson mixed models. Journal of Environmental Management 154, 151-158

Cai Q, Zeng A, Su Z, Guo F (2020): Driving factors of forest fire in Zhejiang province based on Logistic regression model. Journal of Northwest A & F University (Natural Science Edition) 48, 102-109

Catry FX, Rego FC, Bação FL, Moreira F (2009): Modeling and mapping wildfire ignition risk in Portugal. International Journal of Wildland Fire 18, 921-931

Chang Y, Zhu Z, Bu R, Chen H, Feng Y, Li Y, Hu Y, Wang Z (2013): Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. Landscape Ecology 28, 1989-2004

Chen D (2019): Prediction of Forest Fire Occurrence in Daxing'an Mountains Based on Logistic Regression Model. Forest Resources Management, 116-122

Chen W, Zhou Y, Zhou E, Xiang Z, Zhou W, Lu J (2021): Wildfire risk assessment of transmission-line corridors based on naïve Bayes network and remote sensing data. Sensors 21, 634

Crosby JS (1954): Probability of fire occurrence can be predicted.

Cunningham AA, Martell DL (1973): A stochastic model for the occurrence of man-caused forest fires. Canadian Journal of Forest Research 3, 282-287

Dayananda P (1977): Stochastic models for forest fires. Ecological Modelling 3, 309-313

Del Hoyo LV, Isabel MPM, Vega FJM (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

Fielding AH, Bell JF (1997): A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation 24, 38-49

Fischer AP, Spies TA, Steelman TA, Moseley C, Johnson BR, Bailey JD, Ager AA, Bourgeron P, Charnley S, Collins BM (2016): Wildfire risk as a socioecological pathology. Frontiers in Ecology and the Environment 14, 276-284

GAO X, LIAO S (2017): Design and implementation of forest fire probability prediction system based on Bayesian network. Computer Engineering and Applications 13, 246-251

Garcia CV, Woodard P, Titus S, Adamowicz W, Lee B (1995): A logit model for predicting the daily occurrence of human caused forest-fires. International Journal of Wildland Fire 5, 101-111

Guo F, Hu H, Jin S, Ma Z, Zhang Y (2010a): Relationship between forest lighting fire occurrence and weather factors in Daxing’an Mountains based on negative binomial model and zero-inflated negative binomial models Chinese Journal of Plant Ecology 34, 571-577

Guo F, Hu H, Ma Z, Zhang Y (2010b): Applicability of different model in simulating the relationships between forest fire occurrence and weather factors in Daxing’an Mountains. Chinese Journal of Applied Ecology 21, 159-164

Guo F, Su Z, Ma X, Song Y, Sun L, Hu H, Yang T (2015a): Climatic and non-climatic factors driving
lightning-induced fire in Tahe, Daxing'an mountain. Acta Ecologica Sinica 35, 6439-6448

Guo F, Selvalakshmi S, Lin F, Wang G, Wang W, Su Z, Liu A (2016a): Geospatial information on geographical and human factors improved anthropogenic fire occurrence modeling in the Chinese boreal forest. Canadian Journal of Forest Research 46, 582-594

Guo F, Zhang L, Jin S, Tigabu M, Su Z, Wang W (2016b): Modeling anthropogenic fire occurrence in the boreal forest of China using logistic regression and random forests. Forests 7, 250

Guo F, Su Z, Tigabu M, Yang X, Lin F, Liang H, Wang G (2017): Spatial modelling of fire drivers in urban-forest ecosystems in China. Forests 8, 180

Guo H, Zhang M, Xu L, Yuan Z, Chen T (2015b): Geographically weighted regression based on estimation of regional forest carbon storage. Journal of Zhejiang A & F University 32, 497-508

Liang H 2016: Based on spatial and non-spatial model and influence factors analysis of the space-time characteristics of Fujian forest fire. master Thesis, Fujian Agriculture and Forestry University

Liang H, Wang W, Guo F, Lin F, Lin Y (2017): Comparing the application of logistic and geographically weighted logistic regression models for Fujian forest fire forecasting. Acta Ecologica Sinica 37, 4128-4141

Ma W, Feng Z, Cheng Z, Wang F (2020): Study on driving factors and distribution pattern of forest fires in Shanxi province. Journal of Central South University of Forestry & Technology 40, 57-69

Marlon JR, Bartlein PJ, Gavin DG, Long CJ, Anderson RS, Briles CE, Brown KJ, Colombaroli D, Hallett DJ, Power MJ (2012): Long-term perspective on wildfires in the western USA. Proceedings of the National Academy of Sciences 109, E535-E543

Martínez-Fernández J, Chuvieco E, Koutsias N (2013): Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. Natural Hazards and Earth System Sciences 13, 311-327

Monjarás-Vega NA, Briones-Herrera CI, Vega-Nieva DJ, Calleros-Flores E, Corral-Rivas JJ, López-Serrano PM, Pompa-García M, Rodríguez-Trejo DA, Carrillo-Parra A, González-Cabón A (2020): Predicting forest fire kernel density at multiple scales with geographically weighted regression in Mexico. Science of the Total Environment 718, 137313

NorTh MP, Stephens SL, Collins BM, Agee JK, ApleT G, Franklin JF, Fule PZ (2015): Reform forest fire management. Science 349, 1280-1281

Oliveira S, Pereira JM, San-Miguel-Ayanz J, Lourenço L (2014): Exploring the spatial patterns of fire density in Southern Europe using Geographically Weighted Regression. Applied Geography 51, 143-157

Pan D, Yu P, Wu Q (2018a): Application of Random Forest Algorithm on the Forest fire Prediction Based on Meteorological Factors in the Hilly Area, Central Hunan Province. Journal of Northwest Forestry University 33, 169-177

Pan D, Yu P, Wu Q (2018b): Application of random forest algorithm on the forest fire prediction based on meteorological factors in the hilly area, central Hunan Province. Journal of Northwest Forestry University 33, 175-183

Parisien M-A, Snetsinger S, Greenberg JA, Nelson CR, Schoennagel T, Dobrowski SZ, Moritz MA (2012): Spatial variability in wildfire probability across the western United States. International Journal of Wildland Fire 21, 313-327

Peng X, Jin Q, Zhan Q, Guo F (2021): Relevant Factor Analysis of Wildfire of Zhejiang Province Using Geographically Weighted Logistic Regression Models. Journal of Northeast Forestry University 49, 57-66
Renard Q, Pélissier R, Ramesh B, Kodandapani N (2012): Environmental susceptibility model for predicting forest fire occurrence in the Western Ghats of India. International Journal of Wildland Fire 21, 368-379

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. Applied Geography 48, 52-63

Shu L, Tian X (1997): Present situation and prospect of forest fire prevention abroad. World Forestry Research, 29-37

Shu L, Tian X, Li H (1998): A review of the world’s forest fire situation. World Forestry Research, 42-48

Wen B, Xie X, Sun M, Du Z, Li S, Huang P, Zhu Y, Xie B (2019): Forest Fire Prediction Based on Weighted Logistic Regression Model. Forestry and Environmental Science 35, 79-83

Westerling AL (2016): Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring. Philosophical Transactions of the Royal Society B: Biological Sciences 371, 20150178

Wotton B, Martell DL (2005): A lightning fire occurrence model for Ontario. Canadian Journal of Forest Research 35, 1389-1401

Wu H, Zhu L, Liu Z, Kong L, Guo X, Zhang F (2018): A Study of Regularity and Prediction Model for Forest Fire in China. World Forestry Research 31, 64-70

Xiao Y, Tian Z, Wei Y (2013): Testing for spatial-temporal nonstationarity based on geographically and temporally weighted regression model. Systems Engineering-Theory & Practice 33, 1537-1542

Xu A, Li Q, Fang L, Wu D (2003): Research and discussion on forest fire forecasting model based on GIS. Journal of Zhejiang A & F University, 61-64

Yang X, Jin X, Zhou Y (2021): Wildfire Risk Assessment and Zoning by Integrating Maxent and GIS in Hunan Province, China. Forests 12, 1299

Yue C, Luo C, Shu L, Shan Z (2020): A review on wildfire studies in the context of global change. Acta Ecologica Sinica 40, 385-401

Zhang G, Wang M, Liu K (2019): Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China. International Journal of Disaster Risk Science 10, 386-403

Zhang H, Han X, Dai S (2013): Fire occurrence probability mapping of northeast China with binary logistic regression model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6, 121-127

Zhang H, Qi P, Guo G (2014): Improvement of fire danger modelling with geographically weighted logistic model. International Journal of Wildland Fire 23, 1130-1146

Zhou X-H, McClish DK, Obuchowski NA (2009): Statistical methods in diagnostic medicine, 569. John Wiley & Sons