forests

MDPI

Article

Identifying Forest Fire Driving Factors and Related Impacts in China Using Random Forest Algorithm

Wenyuan Ma [1], Zhongke Feng [1,*], Zhuxin Cheng [1], Shilin Chen [1] and Fengge Wang [2]

[1] Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 10083, China; imawenyuan@126.com (W.M.); chengzhuxin1993@foxmail.com (Z.C.); chenshilin@bjfu.edu.cn (S.C.)
[2] Forest Fire Prevention and Monitoring Center in Ministry of Emergency Management of China, Beijing 100054, China; slgaj9204@sina.com
* Correspondence: zhongkefeng@bjfu.edu.cn; Tel.: +86-138-1030-5579

Received: 22 March 2020; Accepted: 28 April 2020; Published: 1 May 2020



Abstract: Reasonable forest fire management measures can effectively reduce the losses caused by forest fires and forest fire driving factors and their impacts are important aspects that should be considered in forest fire management. We used the random forest model and MODIS Global Fire Atlas dataset (2010~2016) to analyse the impacts of climate, topographic, vegetation and socioeconomic variables on forest fire occurrence in six geographical regions in China. The results show clear regional differences in the forest fire driving factors and their impacts in China. Climate variables are the forest fire driving factors in all regions of China, vegetation variable is the forest fire driving factor in all other regions except the Northwest region and topographic variables and socioeconomic variables are only the driving factors of forest fires in a few regions (Northwest and Southwest regions). The model predictive capability is good: the AUC values are between 0.830 and 0.975, and the prediction accuracy is between 70.0% and 91.4%. High fire hazard areas are concentrated in the Northeast region, Southwest region and East China region. This research will aid in providing a national-scale understanding of forest fire driving factors and fire hazard distribution in China and help policymakers to design fire management strategies to reduce potential fire hazards.

Keywords: forest fire driving factors; forest fire occurrence; random forest; forest fire management; China

1. Introduction

Forests are ecosystems with rich biodiversity [1–3], and they play an important role in soil and water conservation, climate regulation, the carbon cycle and other aspects [4,5]. Fire, which affects the biodiversity, species composition and ecosystem structure of forest ecosystems, is the dominant disturbance factor in many forest ecosystems [6–9]. Moreover, fire also affects human lives, regional economies and environmental health [10–12]. In short, forest fires threaten the sustainable development of modern forestry and human security [13]. Therefore, as an important component of global environmental change, forest fires have become the focus of forestry and ecological research [14,15]. An important aspect of forest fire management and prevention is studying forest fire driving factors and their impacts, which can help fire prevention departments to accurately assess forest fire hazards and effectively implement forest fire prevention strategies [11,16]. Forest fires are affected complexly by many driving factors, so it is very important to select appropriate forest fire driving factors and prediction models.

Forest fire driving factors have generally been divided into four types, namely, climate, vegetation, topography and socioeconomic [17,18], which vary at different temporal and spatial scales [19]. Regarding impact modes, climate factors control the accumulation and water content of forest

fuels [20,21], which are usually considered as the major determinants of forest fire occurrence [22]. Vegetation is a source of forest fuel and directly affects the ignition capacity [23,24]. Topography can affect the structure and distribution of vegetation, thus affecting the possibility of forest fires as well as the spread speed and direction of forest fires [25]. Socioeconomic factors affect forest fire occurrence via building expansion, traffic network construction and human-related activities, which increase pressure on wildlands, bringing ignition sources close to forests [23,26]. In terms of impact scope, climate affects forest fires on a larger scale while the vegetation, topography and socioeconomic factors affect forest fires on a smaller scale [27]. In terms of impact relationship, there are nonlinear relationships and thresholds between forest fire driving factors and forest fire occurrence [28–30]. Random forest is a machine learning algorithm, which can automatically select important variables and flexibly evaluate the complex interaction between variables. In recent years, random forest has been used in the study of forest fire driving factors and has shown better prediction ability than multiple linear regression [31] and logistic regression [18].

Previous studies have analysed forest fire distribution, forest fire frequency and burnt area at the national scale in China. Tian [32] analysed the spatial and temporal distribution characteristics of wildfires for 2008–2012 in mainland China. Chang [33] explored the environmental factors influencing the spatial variation in the mean number of fires and mean burned forest area from 1987 to 2007 at a provincial level using cluster analysis and redundancy analysis. Zhong [34] analysed the changes in fire frequency and burnt area during 1992–1999 in China. Lu [35] analysed the impacts of annual temperature and precipitation on the burnt area dynamics in China. Ying used ground-based data to analyse the environmental and social factor contributions to the spatial variation of forest fire frequency and burnt area summarized at a county level during 1989–1991 in China [36]. However, previous studies have used models to analyse the driving factors and their impacts of forest fire occurrence in China, mainly at the provincial scale, such as in Fujian province [18], Heilongjiang province [29,37] and Shanxi province [38]. There is still a lack of nationwide research on forest fire driving factors and their influence on recent forest fires. The value of this study lies in conducting the nationwide research which can provide a detailed analysis and practical information of the forest fire hazard and would help governments to formulate more accurate forest fire prevention strategies and allocate resources rationally. In this study, we used the random forest model and forest fire ignitions for 2010~2016 (obtained from MODIS Global Fire Atlas dataset) to evaluate the impact of four types of forest fire driving factors and the regional differences of these factors in China. This study has three objectives: (1) to determine the forest fire driving factors in various geographical regions of China and analyse how they affect forest fire occurrence; (2) to map the likelihood of forest fire occurrence in China and (3) to discuss forest fire prevention strategies in different geographical areas of China.

2. Materials and Methods

2.1. Study Area

The study area covered mainland China (Hong Kong, Macao and Taiwan were not analysed due to a lack of data). The driving factors of forest fires and their effect were analysed in 6 geographical regions: Northeast region (NE), North China region (N), East China region (E), Northwest region (NW), Southwest region (SW) and Mid-south region (MS). Each region is an aggregation of provinces with adjacent locations and similar topography, economy and climate. The details of each region are shown in Figure 1 and Table 1.

Table 1. Basic description of the six geographical regions in China.

Study Area	Province	Main Climate Types	Topography	Dominant Vegetation Types	Socioeconomic Conditions
Northeast region	Heilongjiang, Jilin, and Liaoning	Middle temperate monsoon climate	Dominant terrain is plains and mountains. The elevation in most areas is below 500 m.	Temperate coniferous broadleaved mixed forests and cold temperate coniferous forest. Forest coverage is 41.59% [39].	The total population is 108.75 million. The per capita GDP is ¥49,891 yuan [40].
North China region	Inner Mongolia, Shanxi, Beijing, Tianjin, and Hebei	Middle temperate continental climate and warm temperate monsoon climate	The dominant terrain types are plateaus and hills. The elevation in most areas is below 2000 m.	Warm temperate deciduous broadleaved forest and temperate grassland. Forest coverage is 21.09% [39].	The total population is 174.79 million. The per capita GDP is ¥64,194 yuan [40].
East China region	Shandong, Jiangsu, Anhui, Zhejiang, Shanghai, Jiangxi, and Fujian	Warm temperate monsoon climate and subtropical monsoon climate	The dominant terrain types are plains and mountains. The elevation in most areas is below 1000 m.	Warm temperate deciduous broadleaved forest and subtropical evergreen broadleaved forest. Forest coverage is 40.64% [39].	The total population is 408.98 million. The per capita GDP is ¥78,271 yuan [40] (China Statistical Yearbook, 2018).
Northwest region	Xinjiang, Gansu, Ningxia, Qinghai, and Shaanxi	Warm temperate continental climate	The dominant terrain types, which fluctuate greatly, are deserts and high mountains. The elevation in most areas is between 500 and 5000 m.	Temperate desert and alpine vegetation on the Qinghai-Tibetan plateau. Forest coverage is 8.21% [39].	The total population is 101.86 million. The per capita GDP is ¥45,463 yuan [40].
Southwest region	Tibet, Sichuan, Chongqing, Yunnan, and Guizhou	Subtropical monsoon climate and alpine climate	The terrain is complex and consists of basins, plateaus and mountains. The elevation in most areas is between 500 and 6000 m.	Alpine vegetation on the Qinghai-Tibetan plateau, subtropical evergreen broadleaved forest and tropical rainforest. Forest coverage is 25.75% [39].	The total population is 200.95 million. The per capita GDP is ¥43,609 yuan [40].
Mid-south region	Henan, Hubei, Hunan, Guangxi, Guangdong, and Hainan	Subtropical monsoon climate and tropical monsoon climate	The dominant terrain types are plains and mountains. The elevation in most areas is below 1000 m.	Subtropical evergreen broadleaved forest and tropical rainforest. Forest coverage is 44.63% [39].	The total population is 393.01 million. The per capita GDP is ¥57,664 yuan [40].

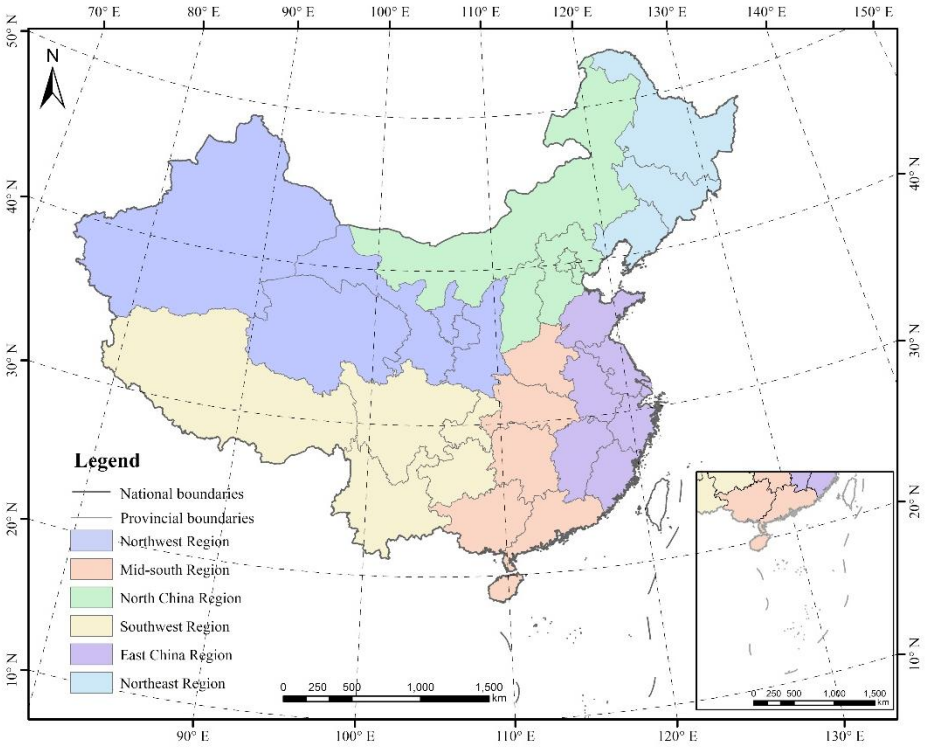

Figure 1. Six geographical regions in China.

2.2. Data Preparation

2.2.1. Dependent Variables

We identified 17,466 forest fires (ignitions) between 2010 and 2016 across mainland China with the Global Fire Atlas dataset (downloaded from the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC), https://daac.ornl.gov) and Chinese land-use type dataset (downloaded from the Resource and Environment Data Cloud Platform, http://www.dsac.cn). The timing and location of the fire ignitions were provided by the Global Fire Atlas, which is a global dataset that records the daily dynamics of individual fires based on the Global Fire Atlas algorithm and estimated burn dates from the Moderate Resolution Imaging Spectroradiometer (MODIS) [41]. A Chinese land-use type dataset provided the forest land range in mainland China for 2015 at a 1000-m spatial resolution. According to this range, we identified Chinese forest fire ignitions for 2010–2016 from the Global Fire Atlas dataset in ArcGIS10.2 software (Environmental Systems Research Institute, RedLands, CA, USA). Figure 2 shows the distribution of the forest fire ignitions for six geographical regions in China.

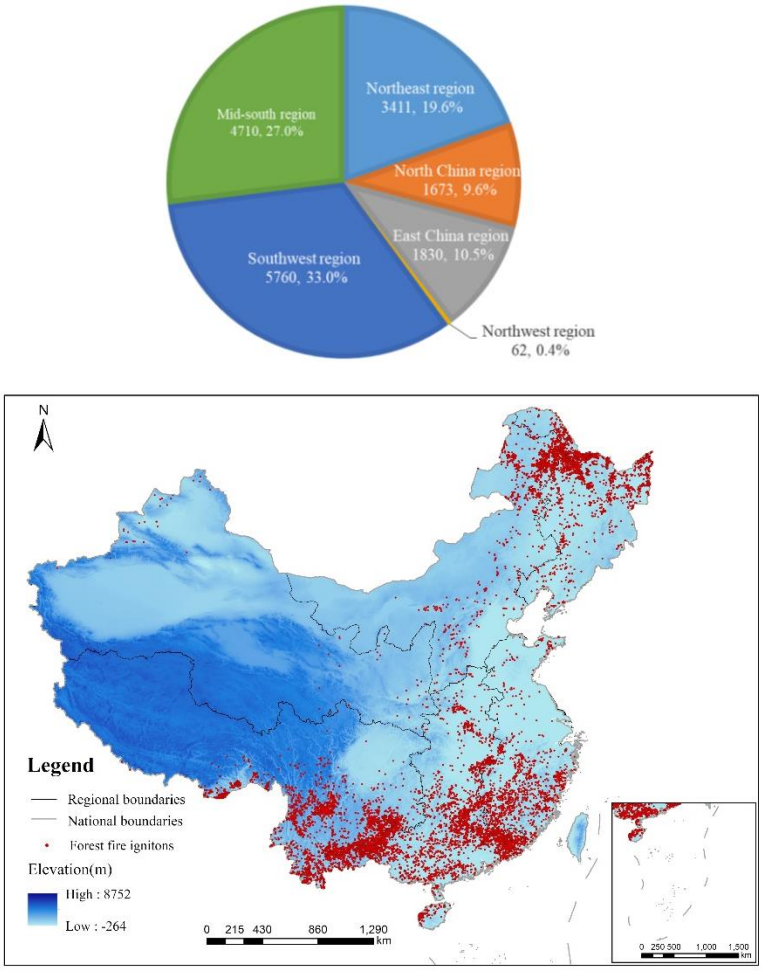

Figure 2. Distribution of the number, proportion and location of forest fire ignitions in six geographical regions from 2010 to 2016.

Modelling forest fire occurrence requires a binary target variable, so a certain percentage of control points (nonfire points) were generated randomly according to three principles: (1) the ratio of forest fire ignition points to control points was 1:1.5 [29], (2) the control points were located within the forest land range in mainland China and (3) the points were random in both time and space. ArcGIS10.2 software was used to randomly generate the control points, and the dates of the control points were randomly selected during 2010–2016 to meet the randomness of time.

2.2.2. Explanatory Variables

A total of 21 variables, grouped into climate, topography, vegetation and socioeconomic categories, were selected as the initial forest fire driving factors (Table 2). All variables were integrated in ArcGIS10.2 software and extracted to the forest fire ignition points and control points.

Table 2. Initial explanatory variables that were collected for inclusion as forest fire drivers.

Variable Type	Variable Name	Code	Source	Resolution, Units	References
Climatic	Annual precipitation in the year before the fire	Pre_year0	National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China (http://www.geodata.cn)	0.25°, 0.1 mm	[18,31,42]
	Annual precipitation in the year of the fire	Pre_year1		0.25°, 0.1 mm	
	Annual soil moisture in the year before the fire	Soil_mois0		0.25°, m^3 m^{-3}	
	Annual soil moisture in the year of the fire	Soil_mois1		0.25°, m^3 m^{-3}	
	Daily average ground surface temperature	GST_avg	Daily Data Set of China's Surface Climate Data (V3.0), National Meteorological Information Centre (http://data.cma.cn)	0.1 °C	[18,31,37,43,44]
	Daily precipitation	Pre_daily		0.1 mm	
	Daily average air pressure	Pres_avg		0.1 hPa	
	Daily average relative humidity	RH_avg		%	
	Daily minimum relative humidity	RH_min		%	
	Sunshine hours	SSD		0.1 h	
	Daily average temperature	Tem_avg		0.1 °C	
	Daily average wind speed	Win_avg		0.1 m/s	
	Daily maximum wind speed	Win_max		0.1 m/s	
Topographic	Elevation	Elevation	Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn)	90 m, m	[28,29,45,46]
	Slope	Slope		90 m, °	
	Aspect	Aspect		90 m	
Vegetation	Fractional vegetation cover	FVC	Resource and Environment Data Cloud Platform (http://www.resdc.cn)(Xu,2018)	1 km	[18]
Socioeconomic	The distance to road and railway	Dis_road	National Catalogue Service for Geographic Information website (http://www.webmap.cn)	1:1,000,000, m	[18,28,29,46,47]
	The distance to settlement	Dis_sett		1:1,000,000, m	
	Population density	POP	National Earth System Science Data Center (http://www.geodata.cn)	1 km, number/km^2	
	Per capita gross national product	GDP		1 km, RMB/km^2	

Climate Variables

Climate variables affect fuel accumulation and moisture which largely determine the time, location and occurrence probability of forest fires [31]. In this study, the initial climate variables include annual variables and daily variables. The annual variables are precipitation and soil moisture. As climate factors in the period before the fire can also affect vegetation accumulation and the fuel moisture content, precipitation and soil moisture in the year before individual forest fire ignition during 2010–2016 were also taken into consideration [29,31]. We downloaded precipitation data with a 1-km spatial and monthly temporal resolution [42] and soil moisture data with a 0.25° spatial and monthly temporal resolution from the National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China (http://www.geodata.cn). Based on these data, we calculated the annual cumulative precipitation and the annual average soil moisture for 2009–2016.

The daily initial climate variables include daily average temperature, daily average ground surface temperature, daily average relative humidity, daily minimum relative humidity, daily precipitation, daily average atmospheric pressure, sunshine hours, daily average wind speed and daily maximum wind speed. The daily humidity, precipitation, wind speed and sunshine hours affect the possibility of forest fire occurrence by reflecting fuel moisture. Daily temperature is the key condition triggering fire ignition. Atmospheric pressure can affect the oxygen content in the air, and the pressure obviously differs due to significant altitude differences and the complex terrain in China; therefore, atmospheric pressure was also considered as an initial climate variable. Daily climate data were obtained from the Daily Data Set of China's Surface Climate Data (V3.0) of the National Meteorological Information Centre (http://data.cma.cn), and we included daily data from 824 national weather stations in China. The daily climate variable values for each fire ignition and control point were provided by the weather station nearest that point.

Topographic Variables

Topography influences the possibility of forest fire occurrence by affecting the vegetation composition and distribution and local microclimate [25]. In this study, the initial topographic variables include elevation, slope and aspect. Data for these variables were extracted from digital elevation model (DEM) data in China with 90-m spatial resolution (obtained from Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences, http://www.gscloud.cn). Aspect was divided into 8 categories according to the criteria in Table 3.

Table 3. Aspect classification criteria.

Aspect	Azimuth (°)
North	337.5~22.5
Northeast	22.5~67.5
East	67.5~112.5
Southeast	112.5~157.5
South	157.5~202.5
Southwest	202.5~247.5

Vegetation Variable

The initial vegetation variable is the fractional vegetation cover (FVC), which is the percentage of the vertical projection of vegetation area to the ground surface within a unit area [48] and can well represent the amount of forest fuel [18,29]. The normalized difference vegetation index (NDVI) is significantly better than other vegetation indices in estimating FVC [49,50], so we calculated FVC based on the annual NDVI dataset for 2010–2016. The calculation formula is as follows:

$$FVC = (NDVI - NDVI_{soil}) / \left(NDVI_{veg} - NDVI_{soil}\right) \quad (1)$$

where $NDVI_{soil}$ is the NDVI of bare soil and $NDVI_{veg}$ is the NDVI of dense vegetation canopy. The annual NDVI dataset for 2010–2016 was from the Resource and Environment Data Cloud Platform (http://www.resdc.cn), and the resolution was 1 km [51].

Socioeconomic Variables

Socioeconomic variables affect the probability of forest fire occurrence by affecting human activities. Human travel and engaging in production activities in or around forests will increase the occurrence probability of forest fires. In this study, the initial socioeconomic variables include the distance to the road and railway, the distance to the settlement, gross national product (GDP) and population density. Collectively, these variables can reflect the accessibility of a forest and the possibility of people engaging in fire-prone behaviours in forests [23,26]. The road, railway and settlement datasets were from the National Basic Geographic Database of 1:1 million, which was published on the National Catalogue Service for Geographic Information website (http://www.webmap.cn). The distance between the forest fire ignitions and control points to the nearest road and railway and settlement areas was calculated using the ArcGIS 10.2 "near analysis tool." The population density dataset and GDP dataset were downloaded by the National Earth System Science Data Center (http://www.geodata.cn), and the resolution was 1 km.

2.3. Model

The random forest model was used to identify the forest fire driving factors and their corresponding impacts on forest fire occurrence in 6 geographical regions of China and the whole study area. Random forest is an ensemble learning technique that is derived from classification or regression trees (CARTs). Random forest has a high prediction accuracy and high tolerance to outliers and "noise," and it has shown good prediction ability in forest fire forecasting [30,52]. The random forest model is composed of a combination of various classification trees, which are individually generated by bootstrap samples. Two-thirds of the data are used to train the random forest model and one-third of the data (the out-of-bag samples, OOB) for model validation [53]. Variable importance can also be measured by OOB, which compares increases in OOB error with that variable randomly permuted and all others unchanged [54,55]. The importance score of a variable is as follows [56]:

$$VI\left(X^j\right) = \frac{1}{ntree}\sum_t\left(err'OOB_t^j - errOOB_t^j\right) \quad (2)$$

where X^j is the jth variable, $ntree$ is the number of trees, $errOOB_t^j$ is the OOB error of each tree t and $err'OOB_t^j$ is the OOB error when X^j is permuted, while all other variables remain unchanged among OOB data. For regression, the OOB error is the mean square error; meanwhile, for classification, the OOB is misclassification probability.

In this study, RF was used for classification, which divided dependent variables into two categories: forest fire occurrence and forest fire nonoccurrence. When using an RF model, the number of trees to run (ntree) and the number of variables to try at each split (mtry) need to be defined. According to previous experience [56,57], the value of mtry was set as $\sqrt{number\ of\ variables}$ and the value of ntree was set to 2000. The varSelRF package in R statistical software was applied to select significant variables from the initial variables. Then, we measured and ranked the variable importance of these variables. The partialPlot function was used to draw partial dependence plots which can describe the relationship between the dependent variables and explanatory variables.

To eliminate bias, in each study region and the whole study area, we selected 80% of the original dataset (training dataset) to build the model, and the remaining 20% of the original dataset (independent validation dataset) was used to assess the performance of the final model. Each training dataset was divided into an inner training dataset (60%) and an inner validation dataset (40%) randomly [52]; this procedure was repeated 5 times, and 5 random subsamples of data in each study region and the whole study area were obtained. Each subsample contained an inner training dataset and an inner validation

dataset, and each subsample generates an intermediate model. The variables that were selected as significant variables in at least 3 of the 5 intermediate model were considered as the forest fire driving factors in a region. In each region, the driving factors and training dataset were used to build a final model, and the independent validation dataset was used to validate the model [30].

2.4. Prediction Accuracy of the Models

The receiver operating characteristic (ROC) curve, a coordinate schema analysis method, was applied to measure the predictive capability of the RF models using the area under the curve (AUC) [28,58,59]. The AUC values ranged from 0.5 to 1, with values closer to 1 indicating a relatively higher accuracy, while an AUC value of >0.8 usually indicates good predictive capability [18,60]. We used the Youden criterion, calculated according to the sensitivity and specificity of ROC (Youden criterion = sensitivity + specificity − 1) [28,61], to determine the cut-off point, which was the threshold for judging whether a fire occurred in the models [62]. If the predicted probability was higher than the cut-off point, it was assumed that a forest fire had occurred and vice versa. The prediction accuracy of the model was calculation based on the cut-off point.

2.5. Mapping Forest Fire Occurrence Likelihood

Based on the fire occurrence probability calculated by the random forest model for fire ignitions and nonfire points, we used ordinary kriging interpolation to map the forest fire occurrence likelihood in mainland China in ArcGIS 10.2 [30].

3. Results

3.1. Identification of Forest Fire Driving Factors and Their Importance Ranks

Table 4 and Figure 3 show the forest fire driving factors and their importance rank in six regions and the whole study area. Table A1 and Figure A1 show the significant variables and their importance rank of each intermediate model.

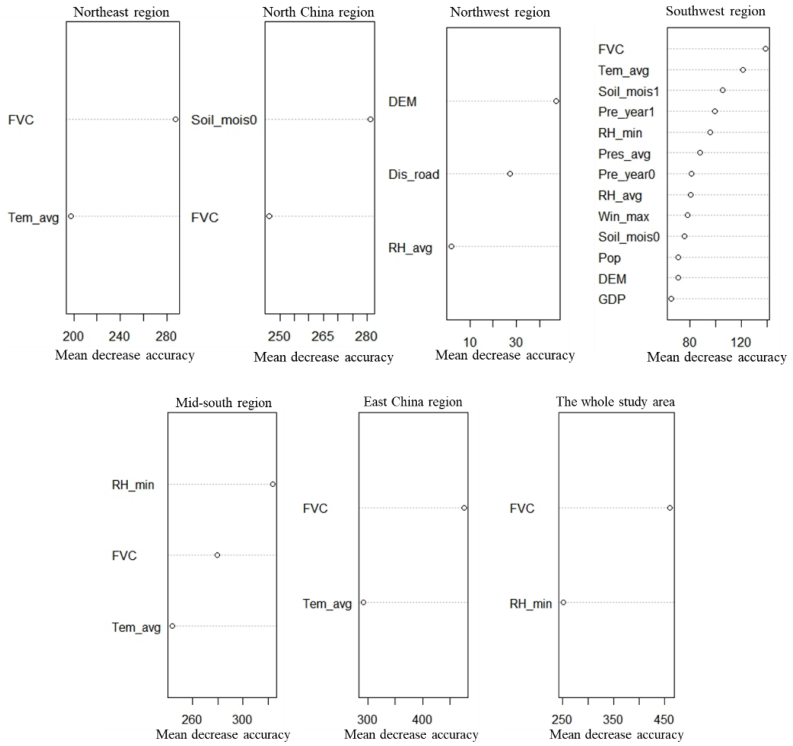

Figure 3. Importance rankings of the forest fire driving factors according to the mean decrease accuracy in six geographical regions and the whole study area. The abbreviated variable names are the same as in Table 4.

Table 4. The results when identifying the forest fire driving factors in six geographical regions and the whole study area.

Variable Type	Variable	NE	N	NW	SW	MS	E	The Whole Study Area
Climatic	Pre_year0				+			
	Pre_year1				+			
	Soil_mois0		+		+			
	Soil_mois1	/	/	/	+	/	/	/
	Tem_avg	+			+	+	+	
	GST_avg	/	/	/	/	/	/	/
	RH_avg			+	+			
	RH_min				+	+		+
	Pre_daily							
	Pres_avg				+			
	SSD							
	Win_avg							
	Win_max				+			
Topographic	DEM			+	+			
	Aspect							
	Slope							
Vegetation	FVC	+	+		+	+	+	+
Socioeconomic	Dis_road			+				
	Dis_sett							
	Pop				+			
	GDP				+			

VIF (variance inflation factor) was used to measure the amount of multicollinearity in the explanatory variables. When VIF > 10, then collinearity in the explanatory variable exists and is excluded in the random forest model. "+" indicates that the variable was identified as being a forest fire driving factor in a given region, and "/" indicates that the variable is excluded due to multicollinearity. NE: Northeast region; N: North China region; NW: Northwest region; SW: Southwest region; MS: Mid-south region; E: East China region; Pre_year0: annual precipitation in the year before the fire; Pre_year1: annual precipitation in the year of the fire; Soil_mois0: annual soil moisture in the year before the fire; Soil_mois1: annual soil moisture in the year of the fire; Tem_avg: daily average temperature; GST_avg: daily average ground surface temperature; RH_avg: daily average relative humidity; RH_min: daily minimum relative humidity; Pre_daily: daily precipitation; Pres_avg: daily average air pressure; SSD: sunshine hours; Win_avg: daily average wind speed; Win_max: daily maximum wind speed; DEM: elevation; FVC: fractional vegetation cover; Dis_road: the distance to road and railway; Dis_sett: the distance to settlement; Pop: population density; GDP: gross national product.

3.2. Influence of the Forest Fire Driving Factors on Forest Fire Occurrence in Different Regions

Partial dependence plots of each forest fire driving factor in each region were drawn to analyse the variables' influence intervals and trends on the probability of forest fire occurrence, where x is the variable value and y is logit of the probability of forest fire occurrence/2 [30]. The markers on the x-axis show the data distribution, where fewer marks indicate less training data and inaccurate model predictions; therefore, only the impact trends within the dense data range are discussed in this study.

Figure 4 shows a nonlinear relationship between each forest fire driving factor and the probability of forest fire occurrence. The vegetation variable shows the same influence trend on the forest fire occurrence probability in each region, and the overall trend is fluctuating. When the fractional vegetation cover is approximately 0.9, the probability of forest fire occurrence shows a peak value and then shows a sharp decline trend, while the probability is lowest when the fractional vegetation cover is approximately 0.98. The impact of climate variables is complex. The daily average temperature shows the same influence trend in the Northeast region and Southwest region: it was positively correlated with the probability of forest fire occurrence initially and negatively correlated after the values exceeded thresholds (12 °C in the Northeast region and 21 °C in the Southwest region). However, it shows another influence trend in the Mid-south region and East China region: the probability of forest fire occurrence is stable at higher values within 20 °C and decreases sharply when the daily average temperature exceeds 20 °C. The average daily relative humidity and the minimum daily relative humidity are generally negative correlated with the probability of forest fire occurrence in the respective regions. The annual soil moisture shows different influence trends in the North China and Southwest regions: in the North China region, it shows a fluctuating trend, while in the Southwest region, the probability of forest fire occurrence increases initially and then decreases as the annual soil moisture increases. For other climate variables, annual precipitation in the year before the fire and the

year of the fire, daily average air pressure and daily maximum wind speed were generally positively correlated with the probability of forest fire occurrence. The elevation shows similar influence trends in the Northwest region and Southwest region and is negatively correlated with the probability of forest fire occurrence. Among socioeconomic variables, the probability of forest fire occurrence decreases with increasing distance from roads and increases initially and then declines with increasing population density and GDP.

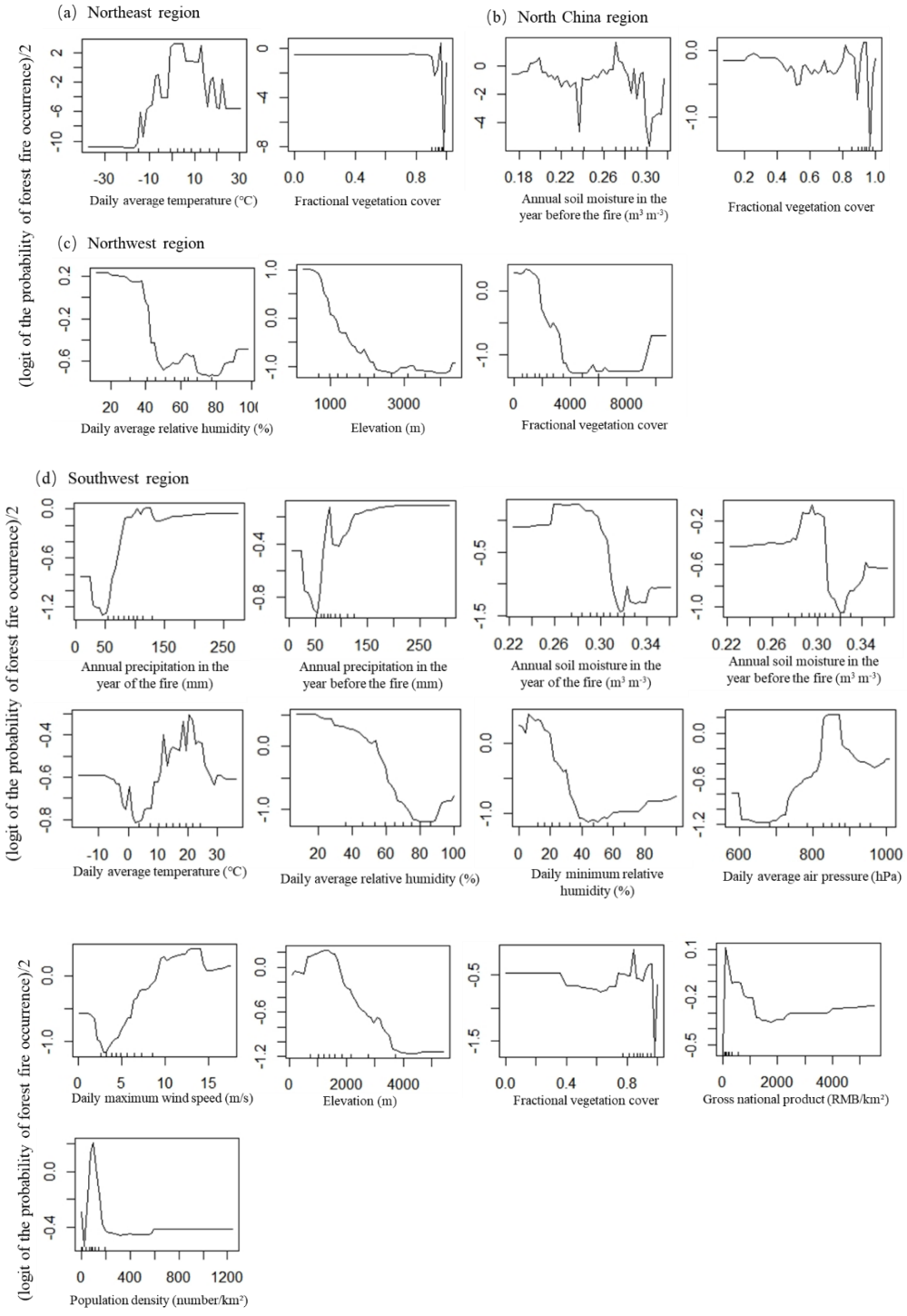

Figure 4. *Cont.*

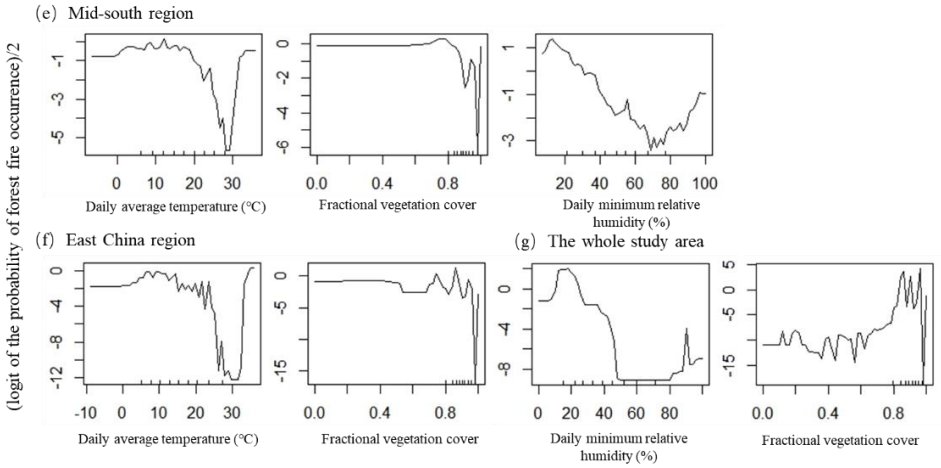

Figure 4. Partial dependence plots of each forest fire driving in six geographical regions and the whole study area.

3.3. Model Prediction Accuracy in Different Regions

The AUC values of each final model and intermediate model are greater than 0.85, and the prediction accuracy is between 70.0% and 91.4% (Table 5), which indicates that the model predictive capability is good. In the final models, the AUC (0.974) and prediction accuracy (91.4% for training and 89.3% for testing) in the East China region were the highest. The AUC (0.871) and prediction accuracy (81.75% for training and 70.52% for testing) in the Northwest region were the lowest, which may be due to the too-few fire ignition in the Northwest region.

Table 5. Comparisons of prediction accuracy of random forest models.

Regions	Model	AUC Value	Cut-Off	Prediction Accuracy (%)	
		(Intermediate Model 1/2/3/4/5)	(Intermediate Model 1/2/3/4/5)	Training (Subtraining 1/2/3/4/5)	Validation (Subvalidation 1/2/3/4/5)
Northeast region	Intermediate model	0.978/0.981/0.971/0.969/0.981	0.466/0.494/0.488/0.466/0.420	92.8/93.0/90.3/89.2/92.5	93.2/92.4/89.4/87.5/93.1
	Final model	0.976	0.493	92	91.5
North China region	Intermediate model	0.974/0.969/0.969/0.970/0.971	0.396/0.426/0.435/0.433/0.390	92.7/92.5/92.7/91.8/92.7	91.6/92.2/91.5/93.5/91.6
	Final model	0.971	0.402	92.8	92.2
East China region	Intermediate model	0.963/0.956/0.963/0.957/0.955	0.470/0.397/0.415/0.457/0.462	90.9/86.8/89.4/89.5/88.8	88.4/86.3/86.7/88.0/89.6
	Final model	0.955	0.467	89	86.2
Northwest region	Intermediate model	0.974/0.981/0.959/0.979/0.960	0.315/0.414/0.462/0.644/0.238	91.0/93.6/91.0/93.6/89.7	86.5/84.6/88.5/90.4/90.4
	Final model	0.964	0.373	91.4	90.3
Southwest region	Intermediate model	0.971/0.966/0.969/0.968/0.968	0.382/0.384/0.379/0.375/0.393	91.5/91.0/91.0/90.6/90.6	90.1/91.2/90.1/90.3/90.2
	Final model	0.966	0.393	90.6	91
Mid-south region	Intermediate model	0.965/0.965/0.953/0.968/0.965	0.397/0.391/0.470/0.418/0.451	89.1/88.9/88.6/89.2/90.4	88.4/88.7/87.4/89.2/90.3
	Final model	0.979	0.481	90.2	87.1
The whole study area	Intermediate model	0.949/0.954/0.951/0.912/0.942	0.419/0.431/0.448/0.372/0.327	86.8/88.2/87.8/83.2/84.4	85.9/88.0/87.6/83.0/83.9
	Final model	0.944	0.415	86	85.8

3.4. Likelihood of Forest Fire Occurrence

Figures 5 and 6 show that the areas with high probability of forest fires are concentrated in the Northeast and Mid-south regions as well as the south of East China region and the northwest of Northwest region. To compare the results of the national model and the regional model, we drew a map of the difference in the likelihood of forest fire occurrence calculated based on the whole study area model and the regional models (Figure 7). The map shows that the probability of the whole model was higher than those of the regional models in most areas of the Southwest region and North China region and in the centre of Northwest region and lower than those in most areas of the Northeast, Northwest, East China and Mid-south regions and in the north of Northwest region.

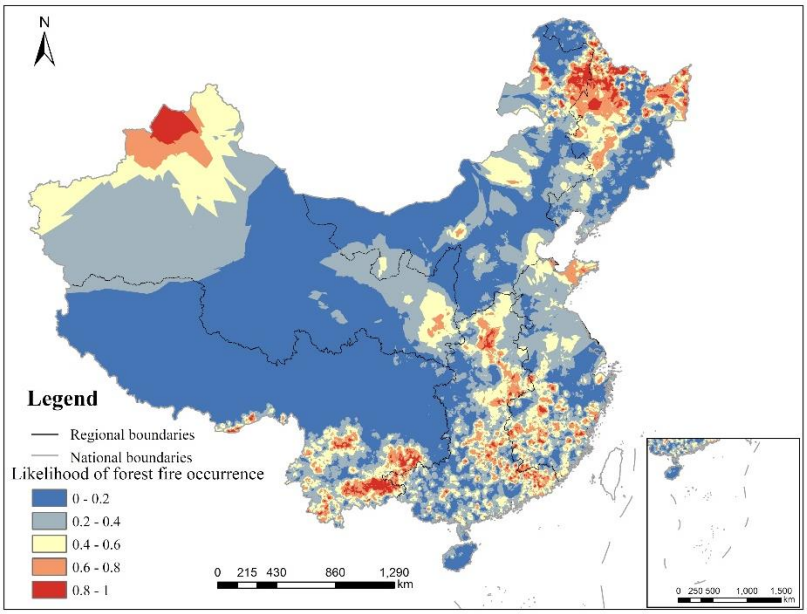

Figure 5. Map of the likelihood of forest fire occurrence in China obtained from the regional models.

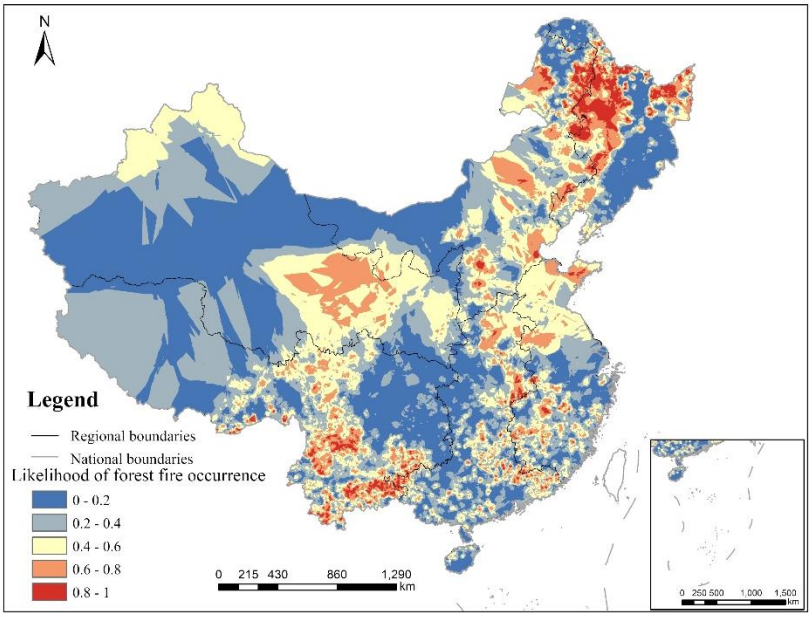

Figure 6. Map of the likelihood of forest fire occurrence in China obtained from the whole study area model.

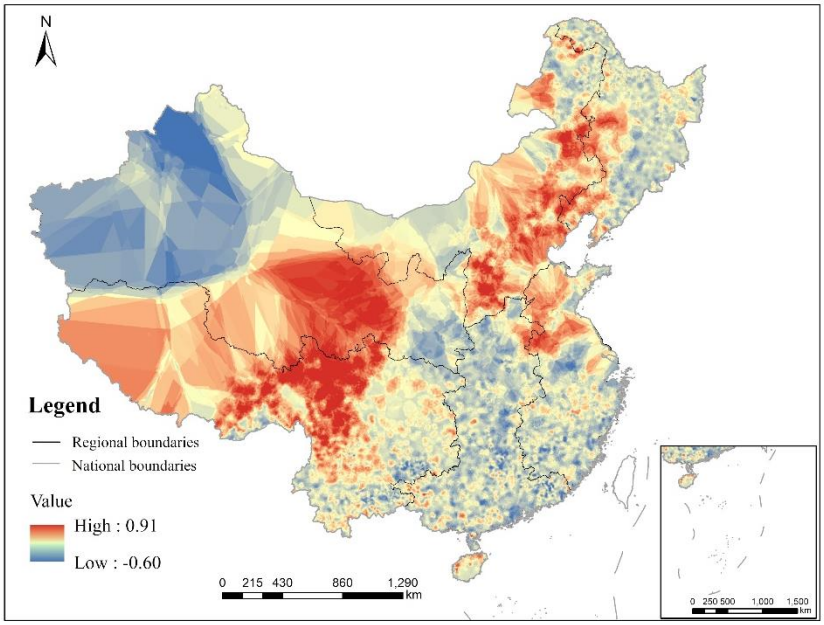

Figure 7. Map of the likelihood of forest fire occurrence obtained from the whole study area model minus that obtained from the regional model.

4. Discussion

4.1. Forest Fire Driving Factors and Their Influence

Previous studies have found regional and scale differences in forest fire factors [18,36,63]. This study also confirmed this point. We found that due to the differing geographical and social conditions in China from region to region, the forest fire driving factors vary in different regions, and the same variables also operate differently depending on the region and the scale of analysis, which illustrates the spatial applicability of forest fire research and the importance of formulating forest fire management systems based on regional characteristics. All final models included a smaller number of variables selected from the initial set. Previous studies have also shown that the simplified model is more stable. Previous studies have also noted that a parsimonious model would be more stable [28,49].

In this study, all final models included climate variables, which are considered the dominant factors affecting forest fires [64–66]. Among climate variables, daily average temperature was the forest fire driving factor in the most regions (Northeast, Southwest, Mid-south, East China and the whole study area). Previous studies [29,30] have shown thresholds and complex nonlinear relationships between temperature and forest fire occurrence probability, and our study confirms this point. The probability of forest fire occurrence initially increases or stabilizes at a higher value with the increase in temperature. When the temperature exceeds a certain threshold (12 °C in the Northeast region, 21 °C in the Southwest region and 20 °C in the Mid-south region and East China region), the probability shows a sharp downward trend. There may be two reasons for this situation. (1) Although high temperatures can increase plant evaporation, thereby reducing the moisture content of forest fire fuels [67], most parts of China experience a monsoon climate (Table 1), and rainfall and heat are synchronous. Therefore, high-temperature weather is often accompanied by high relative humidity levels, which have opposite effects on forest fires, so there were impact thresholds. (2) At high temperatures, forest fire prevention departments are vigilant, implementing strict fire prevention systems and limiting the occurrence of forest fires [68]. Relative humidity is also one of the main forest fire driving factors. In this study, relative humidity showed a similar influence trend in the respective regions, and it was negatively correlated with the occurrence probability of forest fires despite some moderate fluctuations. This is because high relative humidity increases the moisture content of combustible materials and reduces the

possibility of fire [64]. It is noteworthy that the daily minimum relative humidity was also selected as the forest fire driving factor in the whole study area, which indicates that this variable operates at both regional and large scales. Air pressure affects the oxygen content and fuel ignition temperature, and a relatively lower pressure will lead to a lower oxygen content and higher ignition temperature, thus reducing the possibility of forest fire occurrence [69]. However, in the Southwest region, when the daily air pressure is higher than 860 hPa, the probability of forest fire occurrence shows a small decrease, which indicates that there is also an impact threshold of air pressure. The daily maximum wind speed is also one of the driving factors of forest fires in southwest China. The wind will increase evaporation capacity, and the higher the wind speed, the smaller the water content of forest combustibles; hence, the wind speed has a positive correlation with the occurrence probability of forest fires, which is consistent with the research results of Guo [18] in Fujian province of China. The soil moisture in the year of the fire directly affects the water content of forest combustibles [31], so this variable is negatively correlated with the occurrence probability of forest fire. Annual precipitation promotes the accumulation of plant fuels, thus having a positive impact on the occurrence probability of forest fires.

Among the topographic factors, elevation is a forest fire driving factor in the Northwest region and Southwest region, and it is negatively correlated with the occurrence probability of forest fires in both regions. We suspect that this is because the surface of these two areas fluctuates greatly; the elevation in most areas is 500~5000 m in the Northwest region and 500~6000 m in the Southwest region. As elevation increases, human activity decreases, and its impact on weather conditions, vegetation and soil moisture is also not conducive to forest fire occurrence [49,70,71]. Tian's research [32] also showed that forest fires mainly occurred at low elevations in China.

The vegetation variable (fractional vegetation cover) is a forest fire driving factor in all regions except the Northwest region, and its importance ranked first in four regions (Northeast region, Southwest region, East China region and the whole study area). Previous studies have also shown that vegetation cover has an important impact on forest fires [29,67]. Generally, the higher the vegetation coverage, the more fuel is available, so high vegetation coverage leads to a high forest fire rate. However, in this study, fractional vegetation cover showed a complicated influence trend: when the fractional vegetation cover is between 0.8 and 0.97, the occurrence probability of forest fires fluctuates at a higher value, and then it drops rapidly, reaching a minimum value when the fractional vegetation cover is approximately 0.98. We suspect that this is because in forest land with high vegetation coverage, canopy occlusion will lead to some small fires that are not easily detected by MODIS [67].

Compared with other variables, socioeconomic variables are the forest fire factors in few regions (Northwest region and Southwest region), with low degrees of importance (Figure 4). Distance from the road is negatively correlated with the probability of forest fires in the Northwest region because the forests close to the road are vulnerable to traffic accidents and human activities (i.e., smoking and picnics) [26,61]. GDP and population density show similar influence trends, and they have a positive impact on the occurrence probability of forest fires initially and then have a negative impact after exceeding a certain threshold (GDP of 200 RMB/km^2 and population density of 100 number/km^2). This may be because within a certain range, the increase in population density and GDP will increase human activity in forests, thereby promoting forest fire occurrence [51,72,73]. However, in economically prosperous and high-population-density areas, the forest coverage rate is low and there are few forest-related production activities conducted by humans, so the occurrence probability of forest fires decreases [18,29,32].

4.2. Implications for Forest Fire Prevention

There are differences in the forest fire driving factors (Figure 3 and Table 4) and the prediction results (Figures 5–7) between the regional models and the whole study area model. Therefore, it is necessary to study forest fire driving factors based on geographical regions, and regional differences should also be fully considered in forest fire management. Forest fire management departments should formulate forest fire prevention strategies according to the differences in forest fire driving factors and

impact thresholds in different regions. E.g., in the Northwest region, elevation has the greatest impact on forest fire occurrence, and the probability of forest fires is higher in low-elevation areas. Therefore, the Northwest region forest fire management departments should strengthen forest fire monitoring in low-elevation areas, such as setting up more forest fire observation towers and forest fire brigades [30]. In the North China region, soil moisture has the greatest impact, so changes in soil moisture should be taken into account when developing a forest fire prevention strategy. In the Northeast region and East China region, when the daily average temperature reaches the impact threshold, the occurrence probability of forest fires reaches a maximum; hence, forest fire management departments should be more vigilant in the corresponding weather. In the Southwest region, there are 13 forest fire driving factors. These factors should be integrated into the local assessment index systems of forest fire hazard, and the influence of these factors should be considered comprehensively when judging the forest fire hazard. In the Mid-south region, forest fire management departments should pay attention to monitoring the daily minimum relative humidity.

The map of the likelihood of forest fire occurrence is also crucial to forest fire management [74]. Understanding the distribution of forest fire occurrence likelihood can help to determine the location and number of fire observation towers [28], contributing to more effective use of financial and human resources. Figure 5 shows that areas with a high probability of forest fires are concentrated in the Northeast and Mid-south regions as well as the south of East China region and the northwest of Northwest region; thus, more stringent forest fire prevention systems should be implemented in these areas.

4.3. Strengths and Limitations

Previous forest fire research has usually been based on eco-geographical areas or forest types [29,75,76]. However, these zoning methods ignore administrative boundaries, and forest fire management strategies are often formulated by administrative areas. Therefore, we chose a zoning method that takes administrative divisions into account, trying to provide a more practical reference for China's fire prevention department. Our research is based on geographical regions in China, a division method that considers both administrative divisions and natural conditions that has been used in some forestry analysis [77,78]. Each region is an aggregation of provinces with adjacent locations and similar topography, economy and climate. However, this zoning method has its shortcomings. First, if a province has complex topography and different climate and vegetation types (such as Tibet in the Southwest region), it must also be included in one region. We think that this may have led to far higher number of forest fire driving factors in the Southwest region than in the other regions. The second point is about the model. To reveal the nonlinear relationship and influence threshold between forest fire driving factors and forest fire occurrence probability, we used the random forest model, which has shown good prediction ability in previous studies on forest fire [18,30,31]. However, behaving as a "black box," this method cannot calculate regression coefficients or confidence intervals [63,79]. Based on these two points, in future research, we will try to use geographically weighted regression, a spatially explicit technique that would overcome the necessity of building predetermined regions, to analyse forest fire driving factors to address these limitations.

5. Conclusions

We used the random forest model to analyse the forest fire driving factors in different geographical regions in China for 2010 to 2018. The model predictive capability is good, with a prediction accuracy between 70.0% and 91.4%. Furthermore, we mapped the probability of forest fire occurrence in China based on the results of the model. In China, there are obvious regional differences in the types of forest fire driving factors and their impacts. Climate variables (especially temperature and humidity) have major impacts on forest fires occurrence, and the vegetation variable is secondary. Topographic variables and socioeconomic variables are only the forest fire driving factors in the Southwest and Northwest regions. There is a nonlinear relationship and influence threshold between forest fire driving

factors and forest fire occurrence probability. High fire hazard areas are concentrated in the Northeast and Mid-south regions as well as the south of East China region and the northwest of Northwest region. This research will aid in providing a national-scale understanding of forest fire driving factors and fire hazard distribution in China and help policymakers to design fire management strategies and allocate resources reasonably to reduce potential fire hazards.

Author Contributions: Conceptualization, W.M.; Data curation, W.M. and Z.C.; Formal analysis, W.M.; Funding acquisition, Z.F.; Investigation, W.M.; Methodology, W.M.; Resources, F.W.; Supervision, Z.F.; Writing–original draft, W.M.; Writing–review & editing, W.M. and S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities (No. 2015ZCQ-LX-01) and the National Natural Science Foundation of China (No. U1710123).

Acknowledgments: We would like to acknowledge for the data support from National Earth System Science Data Center, National Science & Technology Infrastructure of China. (http://www.geodata.cn)

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A Appendix

Table A1. The results when identifying the significant variables in each intermediate model in six geographical regions and the whole study area.

Variable Type	Variable	Intermediate Models					Selected Frequency
		1	2	3	4	5	
(a) Northeast region							
Climatic	Pre_year0						0
	Pre_year1						0
	Soil_mois0						0
	Soil_mois1	/	/	/	/	/	/
	Tem_avg	+	+	+	+	+	5
	GST_avg	/	/	/	/	/	/
	RH_avg						0
	RH_min						0
	Pre_daily						0
	Pres_avg						0
	SSD						0
	Win_avg						0
	Win_max						0
Topographic	DEM						0
	Aspect						0
	Slope						0
Vegetation	FVC	+	+	+	+	+	5
Socioeconomic	Dis_road						0
	Dis_sett						0
	Pop						0
	GDP						0
(b) North China region							
Climatic	Pre_year0						0
	Pre_year1						0
	Soil_mois0	+		+	+	+	4
	Soil_mois1	/	/	/	/	/	/
	Tem_avg	+	+				2
	GST_avg	/	/	/	/	/	/
	RH_avg						0
	RH_min						0
	Pre_daily						0
	Pres_avg	+					1
	SSD						0
	Win_avg						0
	Win_max						0
Topographic	DEM	+					1
	Aspect						0
	Slope						0
Vegetation	FVC	+	+	+	+	+	5
Socioeconomic	Dis_road						0
	Dis_sett						0
	Pop	+					1
	GDP						0

Table A1. *Cont.*

Variable Type	Variable	Intermediate Models					Selected Frequency
		1	**2**	**3**	**4**	**5**	
(c) Northwest region							
Climatic	Pre_year0						0
	Pre_year1		+				1
	Soil_mois0						0
	Soil_mois1	/	/	/	/	/	/
	Tem_avg						0
	GST_avg	/	/	/	/	/	/
	RH_avg		+	+		+	3
	RH_min						0
	Pre_daily						0
	Pres_avg	+			+		2
	SSD						0
	Win_avg					+	1
	Win_max						0
Topographic	DEM	+	+	+		+	4
	Aspect					+	1
	Slope						0
Vegetation	FVC						0
Socioeconomic	Dis_road			+	+	+	3
	Dis_sett						0
	Pop						0
	GDP	+					1
(d) Southwest region							
Climatic	Pre_year0	+	+	+	+	+	5
	Pre_year1	+	+	+	+	+	5
	Soil_mois0	+	+	+	+	+	5
	Soil_mois1	+	+	+	+	+	5
	Tem_avg	+	+	+	+	+	5
	GST_avg	/	/	/	/	/	/
	RH_avg	+	+	+	+	+	5
	RH_min	+	+	+	+	+	5
	Pre_daily						0
	Pres_avg	+	+	+	+	+	5
	SSD						0
	Win_avg		+	+			2
	Win_max	+			+	+	3
Topographic	DEM	+	+	+	+	+	5
	Aspect						0
	Slope						0
Vegetation	FVC	+	+	+	+	+	5
Socioeconomic	Dis_road						0
	Dis_sett						0
	Pop	+	+	+	+	+	5
	GDP	+	+	+	+	+	5
(e) Mid-south region							
Climatic	Pre_year0			+	+		2
	Pre_year1				+		1
	Soil_mois0				+		1
	Soil_mois1	/	/	/	/	/	/
	Tem_avg	+	+	+	+	+	5
	GST_avg	/	/	/	/	/	/
	RH_avg				+		1
	RH_min	+	+	+	+	+	5
	Pre_daily				+		1
	Pres_avg				+		1
	SSD						0
	Win_avg						0
	Win_max						0
Topographic	DEM						0
	Aspect						0
	Slope						0
Vegetation	FVC	+	+	+	+	+	5
Socioeconomic	Dis_road						0
	Dis_sett						0
	Pop						0
	GDP				+		1

Table A1. *Cont.*

Variable Type	Variable	Intermediate Models					Selected Frequency
		1	2	3	4	5	
(f) East China region							
Climatic	Pre_year0						0
	Pre_year1						0
	Soil_mois0						0
	Soil_mois1	/	/	/	/	/	/
	Tem_avg	+	+	+	+	+	5
	GST_avg	/	/	/	/	/	/
	RH_avg						0
	RH_min						0
	Pre_daily						0
	Pres_avg						0
	SSD						0
	Win_avg						0
	Win_max						0
Topographic	DEM						0
	Aspect						0
	Slope						0
Vegetation	FVC	+	+	+	+	+	5
Socioeconomic	Dis_road						0
	Dis_sett						0
	Pop						0
	GDP						0
(g) The whole study area							
Climatic	Pre_year0						0
	Pre_year1						0
	Soil_mois0						0
	Soil_mois1	/	/	/	/	/	/
	Tem_avg						0
	GST_avg	/	/	/	/	/	/
	RH_avg						0
	RH_min	+	+	+	+	+	5
	Pre_daily						0
	Pres_avg						0
	SSD						0
	Win_avg						0
	Win_max						0
Topographic	DEM						0
	Aspect						0
	Slope						0
Vegetation	FVC	+	+	+	+	+	5
Socioeconomic	Dis_road						0
	Dis_sett						0
	Pop						0
	GDP						0

VIF (variance inflation factor) was used to measure the amount of multicollinearity in the explanatory variables. When VIF > 10, then collinearity in the explanatory variable exists and is excluded in the random forest model. "+" indicates that the variable was identified as being a forest fire driving factor in a given region, and "/" indicates that the variable is excluded due to multicollinearity. Pre_year0: annual precipitation in the year before the fire; Pre_year1: annual precipitation in the year of the fire; Soil_mois0: annual soil moisture in the year before the fire; Soil_mois1: annual soil moisture in the year of the fire; Tem_avg: daily average temperature; GST_avg: daily average ground surface temperature; RH_avg: daily average relative humidity; RH_min: daily minimum relative humidity; Pre_daily: daily precipitation; Pres_avg: daily average air pressure; SSD: sunshine hours; Win_avg: daily average wind speed; Win_max: daily maximum wind speed; DEM: elevation; FVC: fractional vegetation cover; Dis_road: the distance to road and railway; Dis_sett: the distance to settlement; Pop: population density; GDP: gross national product.

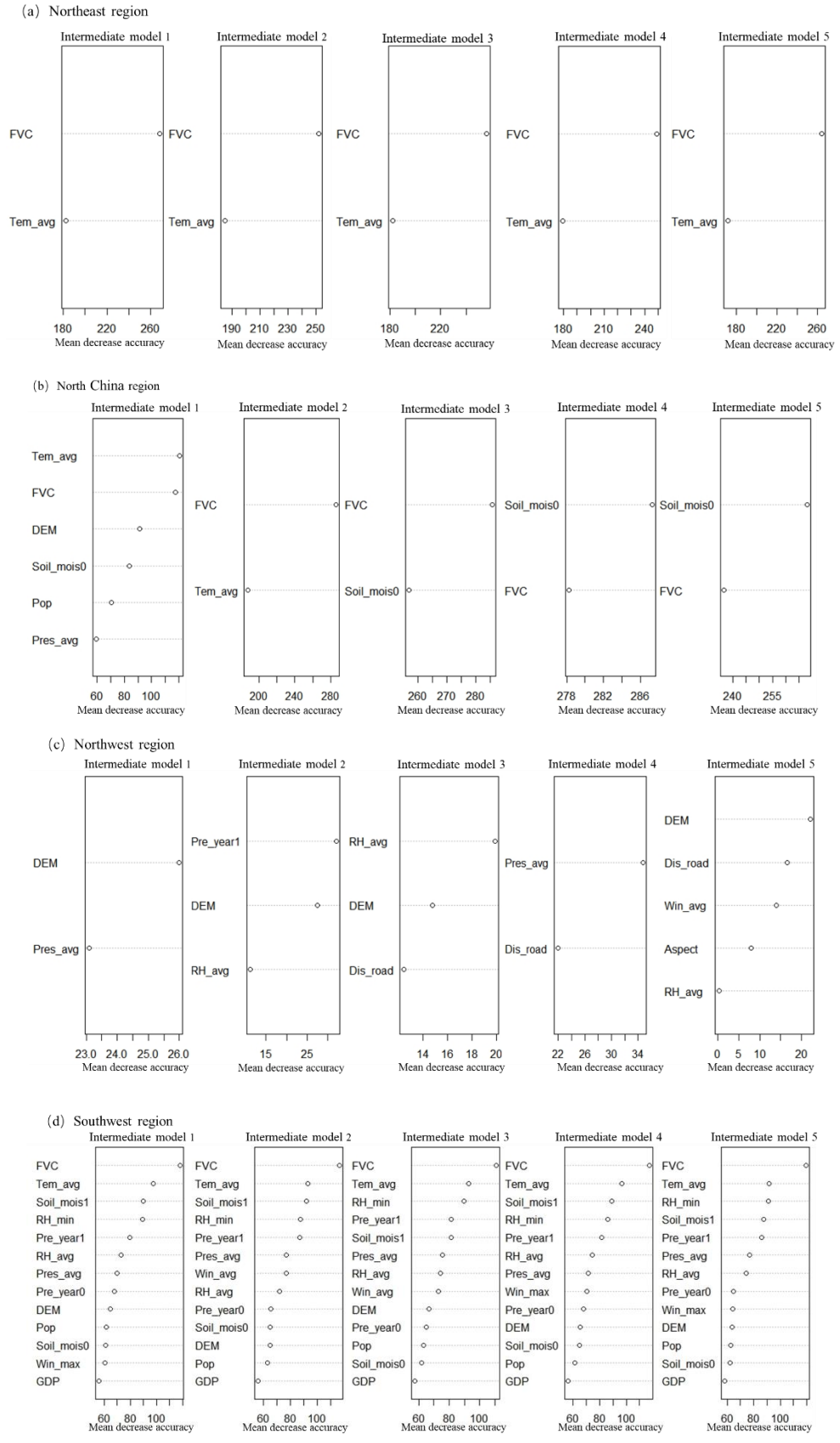

Figure A1. *Cont.*

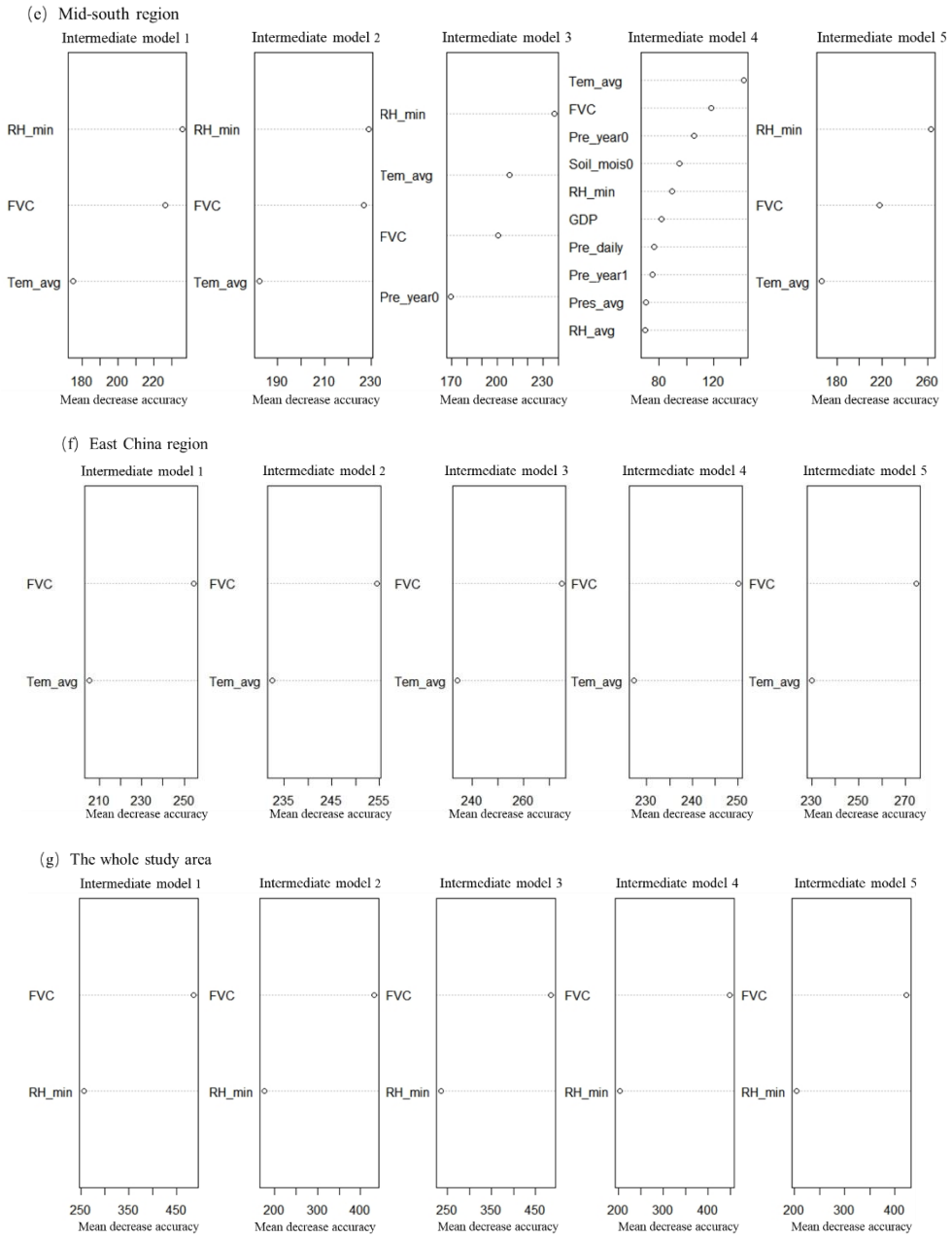

Figure A1. Importance rankings of the significant variables according to the mean decrease accuracy in each intermediate model in six geographical regions and the whole study area. The abbreviated variable names are the same as in Table A1.

References

1. Morales-Hidalgo, D.; Oswalt, S.N.; Somanathan, E. Status and trends in global primary forest, protected areas, and areas designated for conservation of biodiversity from the global forest resources assessment 2015. *For. Ecol. Manag.* **2015**, *352*, 68–77. [CrossRef]
2. Köhl, M.; Lasco, R.; Cifuentes, M.; Jonsson, Ö.; Korhonen, K.T.; Mundhenk, P.; Navar, J.D.J.; Stinson, G. Changes in forest production, biomass and carbon: Results from the 2015 un fao global forest resource assessment. *For. Ecol. Manag.* **2015**, *352*, 21–34. [CrossRef]
3. Keenan, R.J.; Reams, G.A.; Achard, F.; Freitas, J.V.D.; Grainger, A.; Lindquist, E. Dynamics of global forest area: Results from the fao global forest resources assessment 2015. *For. Ecol. Manag.* **2015**, *352*, 9–20. [CrossRef]

4. Bergeron, Y.; Gauthier, S.; Flannigan, M.; Kafka, V. Fire regimes at the transition between mixedwood and coniferous boreal forest in northwestern quebec. *Ecology* **2004**, *85*, 1916–1932. [CrossRef]
5. Piao, S.; Huang, M.; Zhuo, L.; Wang, X.; Ciais, P.; Canadell, J.G.; Kai, W.; Bastos, A.; Friedlingstein, P.; Houghton, R.A. Lower land-use emissions responsible for increased net land carbon sink during the slow warming period. *Nat. Geosci.* **2018**, *11*, 739–743. [CrossRef]
6. Podur, J.; Martell, D.L.; Csillag, F. Spatial patterns of lightning-caused forest fires in Ontario, 1976–1998. *Ecol. Model.* **2003**, *164*, 1–20. [CrossRef]
7. Bond, W.J.; Keeley, J.E. Fire as a global 'herbivore': The ecology and evolution of flammable ecosystems. *Trends Ecol. Evol.* **2005**, *20*, 387–394. [CrossRef]
8. Pastro, L.A.; Dickman, C.R.; Letnic, M. Burning for biodiversity or burning biodiversity? Prescribed burn vs. Wildfire impacts on plants, lizards, and mammals. *Ecol. Appl.* **2011**, *21*, 3238–3253. [CrossRef]
9. Thom, D.; Seidl, R. Natural disturbance impacts on ecosystem services and biodiversity in temperate and boreal forests. *Biol. Rev. Camb. Philos. Soc.* **2016**, *91*, 760–781. [CrossRef]
10. Westerling, A.; Bryant, B. Climate change and wildfire in California. *Clim. Chang.* **2008**, *87*, 231–249. [CrossRef]
11. Hering, A.S.; Bell, C.L.; Genton, M.G. Modeling spatio-temporal wildfire ignition point patterns. *Environ. Ecol. Stat.* **2009**, *16*, 225–250. [CrossRef]
12. Mckenzie, D.; Shankar, U.; Keane, R.E.; Stavros, E.N.; Heilman, W.E.; Fox, D.G.; Riebau, A.C. Smoke consequences of new wildfire regimes driven by climate change. *Earths Future* **2014**, *2*, 35–59. [CrossRef]
13. Shun, L.; Zhiwei, W.; Yu, L.; Hongshi, H. A review of fire controlling factors and their dynamics in boreal forest. *World For. Res.* **2017**, *30*, 41–45. [CrossRef]
14. Dimopoulou, M.; Giannikos, I. Towards an integrated framework for forest fire control. *Eur. J. Oper. Res.* **2004**, *152*, 476–486. [CrossRef]
15. Flannigan, M.D.; Krawchuk, M.A.; Groot, W.J.D.; Wotton, B.M.; Gowman, L.M. Implications of changing climate for global wildland fire. *Int. J. Wildland Fire* **2009**, *18*, 483–507. [CrossRef]
16. Moreno, M.V.; Chuvieco, E. Characterising fire regimes in Spain from fire statistics. *Int. J. Wildland Fire* **2013**, *22*, 296–305. [CrossRef]
17. Ganteaume, A.; Camia, A.; Jappiot, M.; San-Miguel-Ayanz, J.; Long-Fournel, M.; Lampin, C. A review of the main driving factors of forest fire ignition over Europe. *Environ. Manag.* **2013**, *51*, 651–662. [CrossRef]
18. Guo, F.; Wang, G.; Su, Z.; Liang, H.; Liu, A. What drives forest fire in Fujian, China? Evidence from logistic regression and random forests. *Int. J. Wildland Fire* **2016**, *25*, 505–519. [CrossRef]
19. Morgan, P.; Hardy, C.C.; Swetnam, T.W.; Rollins, M.G.; Long, D.G. Mapping fire regimes across time and space: Understanding coarse and fine-scale fire patterns. *Int. J. Wildland Fire* **2001**, *10*, 329–342. [CrossRef]
20. Rollins, M.G.; Morgan, P.; Swetnam, T. Landscape-scale controls over 20th century fire occurrence in two large rocky mountain (USA) wilderness areas. *Landsc. Ecol.* **2002**, *17*, 539–557. [CrossRef]
21. Sharples, J. An overview of mountain meteorological effects relevant to fire behaviour and bushfire risk. *Int. J. Wildland Fire* **2009**, *18*, 737–754. [CrossRef]
22. Minnichl, R.A.; Bahrez, C.J. Wildland fire and chaparral succession along the California-Baja California boundary. *Int. J. Wildland Fire* **1995**, *5*, 13–24. [CrossRef]
23. Pew, K.L.; Larsen, C.P.S. Gis analysis of spatial and temporal patterns of human-caused wildfires in the temperate rain forest of vancouver island, Canada. *For. Ecol. Manag.* **2001**, *140*, 1–18. [CrossRef]
24. Pausas, J.G.; Paula, S. Fuel shapes the fire—climate relationship: Evidence from mediterranean ecosystems. *Glob. Ecol. Biogeogr.* **2012**, *21*, 1074–1082. [CrossRef]
25. Maingi, J.K.; Henry, M.C. Factors influencing wildfire occurrence and distribution in eastern kentucky, USA. *Int. J. Wildland Fire* **2007**, *16*, 23–33. [CrossRef]
26. Cardille, J.A.; Ventura, S.J.; Turner, M.G. Environmental and social factors influencing wildfires in the upper midwest, united states. *Ecol. Appl.* **2001**, *11*, 111–127. [CrossRef]
27. Turco, M.; Llasat, M.C.; von Hardenberg, J.; Provenzale, A. Impact of climate variability on summer fires in a mediterranean environment (northeastern iberian peninsula). *Clim. Chang.* **2013**, *116*, 665–678. [CrossRef]
28. Catry, F.X.; Rego, F.C.; Bação, F.; Moreira, F. Modeling and mapping wildfire ignition risk in portugal. *Int. J. Wildland Fire* **2009**, *18*, 921–931. [CrossRef]
29. Guo, F.; Su, Z.; Wang, G.; Sun, L.; Tigabu, M.; Yang, X.; Hu, H. Understanding fire drivers and relative impacts in different chinese forest ecosystems. *Sci. Total Environ.* **2017**, *411*, 605–606. [CrossRef]

30. Su, Z.; Hu, H.; Wang, G.; Ma, Y.; Yang, X.; Guo, F. Using GIS and random forests to identify fire drivers in a forest city, Yichun, China. *Geomat. Nat. Hazards Risk* **2018**, *9*, 1207–1229. [CrossRef]
31. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M.C. Modeling spatial patterns of fire occurrence in mediterranean europe using multiple regression and random forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. [CrossRef]
32. Tian, X.; Zhao, F.; Shu, L.; Wang, M. Distribution characteristics and the influence factors of forest fires in China. *For. Ecol. Manag.* **2013**, *310*, 460–467. [CrossRef]
33. Chang, Y.; Zhu, Z.; Bu, R.; Li, Y.; Hu, Y. Environmental controls on the characteristics of mean number of forest fires and mean forest area burned (1987–2007) in China. *For. Ecol. Manag.* **2015**, *356*, 13–21. [CrossRef]
34. Zhong, M.; Fan, W.; Liu, T.; Li, P. Statistical analysis on current status of China forest fire safety. *Fire Saf. J.* **2003**, *38*, 257–269. [CrossRef]
35. Aifeng, L.U. Study on the relationship among forest fire, temperature and precipitation and its spatial-temporal variability in China. *Agric. Sci. Technol.* **2011**, *12*, 1396–1400. [CrossRef]
36. Ying, L.; Han, J.; Du, Y.; Shen, Z. Forest fire characteristics in China: Spatial patterns and determinants with thresholds. *For. Ecol. Manag.* **2018**, *424*, 345–354. [CrossRef]
37. Huiling, L. *Based on Spatial and Non-Spatial Model and Influence Factors Analysis of the Space-Time Characteristics of Fujian Forest Fires*; Fujian Agriculture and Forestry University: Fuzhou, China, 2016.
38. Ma, W.; Feng, Z.; Cheng, Z.; Wang, F. Study on driving factors and distribution pattern of forest fires in shanxi province. *J. Cent. South Univ. For. Technol.* **2020**. Available online: https://kns.cnki.net/KCMS/detail/43.1470.S.20200115.1043.001.html (accessed on 5 March 2020). [CrossRef]
39. State Forestry Administ-Ration. *China Forest Resources Inventory Repor*; China Forestry Publishing House: Beijing, China, 2014; pp. 80–81, ISBN 978-7-5038-7424-6.
40. National Bureau of Statistics. China statistical Yearbook. 2018. Available online: http://www.stats.gov.cn/tjsj/ndsj/ (accessed on 10 July 2019).
41. Andela, N.; Morton, D.; Giglio, L.; Paugam, R.; Chen, Y.; Hantson, S.; Werf, G.; Randerson, J. The global fire atlas of individual fire size, duration, speed, and direction. *Earth Syst. Sci. Data Discuss.* **2018**, *11*, 1–28. [CrossRef]
42. Peng, S.; Ding, Y.; Liu, W.; Li, Z. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. *Earth Syst. Sci. Data* **2019**, *11*, 1931–1946. [CrossRef]
43. Zhang, Z.X.; Zhang, H.Y.; Zhou, D.W. Using GIS spatial analysis and logistic regression to predict the probabilities of human-caused grassland fires. *J. Arid Environ.* **2010**, *74*, 386–393. [CrossRef]
44. Yu, M. *The Research of Forest Fire Prediction Model in Fangshan District, Beijing and Sublot Fire Danger Rating Division*; Beijing Forestry University: Beijing, China, 2016.
45. Vilar, L.; Woolford, D.; Martell, D.; Martín, M. A model for predicting human-caused wildfire occurrence in the region of madrid, Spain. *Int. J. Wildland Fire* **2010**, *19*, 325–337. [CrossRef]
46. Prasad, V.K.; Badarinath, K.V.S.; Eaturu, A. Biophysical and anthropogenic controls of forest fires in the deccan plateau, india. *J. Environ. Manag.* **2008**, *86*, 1–13. [CrossRef] [PubMed]
47. Sturtevant, B.; Cleland, D. Human and biophysical factors influencing modern fire disturbance in northern wisconsin. *Int. J. Wildland Fire* **2007**, *16*, 398–413. [CrossRef]
48. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87. [CrossRef]
49. Leprieur, C.; Verstraete, M.M.; Pinty, B. Evaluation of the performance of various vegetation indices to retrieve vegetation cover from avhrr data. *Remote Sens. Rev.* **1994**, *10*, 265–284. [CrossRef]
50. Purevdorj, T.S.; Tateishi, R.; Ishiyama, T.; Honda, Y. Relationships between percent vegetation cover and vegetation indices. *Int. J. Remote Sens.* **1998**, *19*, 3519–3535. [CrossRef]
51. Xu, X. *China Quarterly Vegetation Index (ndvi) Spatial Distribution Data Set*; Data Registration and Publishing System of Resource and Environment Science Data Center of Chinese Academy of Science: Beijing, China, 2018. Available online: http://www.resdc.cn/10.12078/2018060603 (accessed on 10 January 2019). [CrossRef]
52. Rodrigues, M.; Riva, J. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environ. Model. Softw.* **2014**, *57*, 192–201. [CrossRef]
53. Duro, D.C.; Franklin, S.E.; Dubé, M.G. Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *Int. J. Remote Sens.* **2012**, *33*, 4502–4526. [CrossRef]

54. Marston, C.G.; Danson, F.M.; Armitage, R.P.; Giraudoux, P.; Pleydell, D.R.J.; Wang, Q.; Qui, J.; Craig, P.S. A random forest approach for predicting the presence of echinococcus multilocularis intermediate host ochotona spp. Presence in relation to landscape characteristics in western China. *Appl. Geogr.* **2014**, *55*, 176–183. [CrossRef]
55. Abdel-Rahman, E.M.; Ahmed, F.B.; Ismail, R. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using eo-1 hyperion hyperspectral data. *Int. J. Remote Sens.* **2013**, *34*, 712–728. [CrossRef]
56. Liang, H.; Lin, Y.; Yang, G.; Su, Z.; Wang, W.; Guo, F. Application of random forest algorithm on the forest fire prediction in Tahe area based on meteorological factors. *Sci. Silvae Sin.* **2016**, *52*, 89–98. [CrossRef]
57. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.
58. Jiménez-Valverde, A. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Glob. Ecol. Biogeogr.* **2012**, *21*, 498–507. [CrossRef]
59. Chang, Y.; Bu, R.; Chen, H.; Feng, Y.; Li, Y.; Hu, Y.; Wang, Z. Predicting fire occurrence patterns with logistic regression in heilongjiang province, China. *Landsc. Ecol.* **2013**, *28*, 1989–2004. [CrossRef]
60. Vilar, L.; Martín, M.; Martinez-Vega, J. Logistic regression models for human-caused wildfire risk estimation: Analysing the effect of the spatial accuracy in fire occurrence data. *Eur. J. For. Res.* **2011**, *130*, 983–996. [CrossRef]
61. Martínez, J.; Vega-Garcia, C.; Chuvieco, E. Human-caused wildfire risk rating for prevention planning in Spain. *J. Environ. Manag.* **2009**, *90*, 1241–1252. [CrossRef]
62. Vega-Garcia, C.; Woodard, P.M.; Titus, S.J.; Adamowicz, L.; Lee, B.S. A logit model for predicting the daily occurrence of human caused forest-fires. *Int. J. Wildland Fire* **1995**, *5*, 101–111. [CrossRef]
63. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199. [CrossRef]
64. Zumbrunnen, T.; Pezzatti, G.B.; Menéndez, P.; Bugmann, H.; Bürgi, M.; Conedera, M. Weather and human impacts on forest fires: 100 years of fire history in two climatic regions of switzerland. *For. Ecol. Manag.* **2011**, *261*, 2188–2199. [CrossRef]
65. Wotton, M.; Martell, D.; Logan, K. Climate change and people-caused forest fire occurrence in Ontario. *Clim. Chang.* **2003**, *60*, 275–295. [CrossRef]
66. Varela, V.; Vlachogiannis, D.; Sfetsos, A.; Karozis, S.; Politi, N.; Giroud, F. Projection of forest fire danger due to climate change in the french mediterranean region. *Sustainability* **2019**, *11*, 4284. [CrossRef]
67. Chuvieco, E.; Cocero, D.; Riaño, D.; Martin, P.; Martínez-Vega, J.; de la Riva, J.; Pérez, F. Combining ndvi and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. *Remote Sens. Environ.* **2004**, *92*, 322–331. [CrossRef]
68. Hu, T.; Zhou, G. Drivers of lightning- and human-caused fire regimes in the great xing'an mountains. *For. Ecol. Manag.* **2014**, *329*, 49–58. [CrossRef]
69. Song, Z. *Principle and Forecast of Forest Fire*, 1st ed.; China Meteorological Press: Beijing, China, 1991; pp. 56–57, ISBN 7502905820.
70. Sebastian, A.; Salvador, R.; Gonzalo Jimenez, J.; San-Miguel-Ayanz, J. Integration of socio-economic and environmental variables for modelling long-term fire danger in southern europe. *Eur. J. For. Res.* **2008**, *127*, 149–163. [CrossRef]
71. González, J.R.; Palahí, M.; Trasobares, A.; Pukkala, T. A fire probability model for forest stands in Catalonia (north-east Spain). *Ann. For. Sci.* **2006**, *63*, 169–176. [CrossRef]
72. Syphard, A.; Radeloff, V.; Keeley, J.; Hawbaker, T.; Clayton, M.; Stewart, S.; Hammer, R. Human influence on California fire regimes. *Ecol. Appl. Publ. Ecol. Soc. Am.* **2007**, *17*, 1388–1402. [CrossRef] [PubMed]
73. Pereira, M.G.; Malamud, B.D.; Trigo, R.M.; Alves, P.I. The history and characteristics of the 1980–2005 Portuguese rural fire database. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 3343–3358. [CrossRef]
74. Saglam, B.; Bilgili, E.; Dincdurmaz, B.; Kadiogulari, A.I.; Kücük, Ö. Spatio-temporal analysis of forest fire risk and danger using landsat imagery. *Sensors* **2008**, *8*, 3970–3987. [CrossRef]
75. Tian, X.; Shu, L.; Zhao, F.; Wang, M. Dynamic characteristics of forest fires in the main ecological geographic districts of China. *Sci. Silvae Sin.* **2015**, *51*, 71–77. [CrossRef]
76. Wu, Z.; He, H.S.; Keane, R.E.; Zhu, Z.; Wang, Y.; Shan, Y. Current and future patterns of forest fire occurrence in China. *Int. J. Wildland Fire* **2020**, *29*, 104–119. [CrossRef]

77. Lu, J.; Feng, Z.; Zhu, Y. Estimation of forest biomass and carbon storage in China based on forest resources inventory data. *Forests* **2019**, *10*, 650. [CrossRef]
78. Qiu, Z.; Feng, Z.; Song, Y.; Li, M.; Zhang, P. Carbon sequestration potential of forest vegetation in China from 2003 to 2050: Predicting forest vegetation growth based on climate and the environment. *J. Clean. Prod.* **2020**, *252*, 119715. [CrossRef]
79. Cutler, D.R.; Edwards, T.C., Jr.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [CrossRef] [PubMed]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).