Article

A Forest Fire Prediction Method for Lightning Stroke Based on Remote Sensing Data

Zhejia Zhang [1,†], Ye Tian [1,†], Guangyu Wang [2], Change Zheng [1,*] and Fengjun Zhao [3,*]

1   School of Technology, Beijing Forestry University, Beijing 100083, China; zhejiazhang@bjfu.edu.cn (Z.Z.);
    tytoemail@bjfu.edu.cn (Y.T.)
2   Heilongjiang Ecological Engineering Vocational College, Harbin 150025, China; nimi416@163.com
3   Key Laboratory of Forest Protection of National Forestry and Grassland Administration, Ecology and Nature
    Conservation Institute, Chinese Academy of Forestry, Beijing 100091, China
*   Correspondence: zhengchange@bjfu.edu.cn (C.Z.); zhaofj@caf.ac.cn (F.Z.)
†   These authors contributed equally to the work.

Abstract: Forest fires ignited by lightning accounted for 68.28% of all forest fires in the Greater Khin-gan Mountains (GKM) region of northeast China. Forecasting the incidence of lightning-triggered forest fires in the region is imperative for mitigating deforestation, preserving biodiversity, and safeguarding distinctive natural habitats and resources. Lightning monitoring data and vegetation moisture content have emerged as pivotal factors among the various influences on lightning-induced fires. This study employed innovative satellite remote sensing technology to swiftly acquire vegetation moisture content data across extensive forested regions. Firstly, the most suitable method to identify the lightning strikes that resulted in fires and two crucial lightning parameters correlated with fire occurrence are confirmed. Secondly, a logistic regression method is proposed for predicting the likelihood of fires triggered by lightning strikes. Finally, the method underwent verification using five years of fire data from the GKM area, resulting in an AUC value of 0.849 and identifying the primary factors contributing to lightning-induced fires in the region.

Keywords: forest fires; lightning-induced fires; remote sensing data; logistic regression; prediction model

1. Introduction

Forest fires are significant natural disasters characterized by sudden onset [1], high destructiveness [2,3], and challenging mitigation and rescue efforts [4]. Specifically, lightning-induced fires, i.e., forest fires triggered by lightning strikes, pose a more serious threat to affected areas [5], firefighting difficulty, and financial losses. In China, lightning-induced fires occur mainly in high-latitude forest regions, with the GKM forest area being susceptible [6]. Between 1966 and 2009, there were 1592 forest fires reported in the GKM area, of which 613 were attributed to lightning strikes, accounting for 53.21% of the total incidents. It has been observed similarly in North America and Canada, where lightning strikes account for a significant proportion of forest fires. In the boreal forests of North America, approximately 90% of the burned area can be attributed to lightning-induced fires [7]. In Canada, lightning-caused fires make up around 45% of the total number of fires, but they account for approximately 80% of the total burned area [8]. The significance of investigating lightning-induced forest fires is evident, as they diverge from those ignited by human activities. The potential development of fires post-lightning is influenced by natural factors, enabling predictions based on lightning and environmental data.

In the three factors of a fire (ignition sources, combustibles, and combustion conditions), lightning strikes play a crucial role as an ignition source. In the forecasting of lightning-induced fires, a vital step involves determining whether an ignition source is

present in large numbers of lightning strikes. The criteria for assessment include the characteristics of the lightning and the environmental conditions during the lightning occurrence. The former influences the generation of fire sources, while the latter impacts the persistence of fire sources. The primary approach for discerning the characteristics of lightning leading to a fire involves analyzing the lightning parameters associated with previous fire incidents and examining the conditions under which a lightning strike transforms into a source of ignition. Schumacher et al. surveyed lightning fires in Brazil, focusing on the impact of dry thunderstorms and the Lightning Climate Change (LCC) indices on the ignition of combustible materials [9,10]. However, their study did not include a comparative analysis of buffer distance and holding time, crucial factors in accurately identifying ignited lightning strikes. These parameters play a significant role in determining the accuracy of lightning strike identification. However, it is noteworthy that in 2019, out of the 92,517 cloud-to-ground flash incidents in the GKM region, a mere 21 (0.023%) forest fires were caused by lightning. Various factors, such as weather conditions and topography, contribute to the conditions that support fires, while combustibles, such as vegetation type and moisture content, also play a role. This situation underscores the challenge and importance of identifying the specific lightning strikes responsible for initiating fires, as only a minor fraction of the numerous lightning bolts serve as ignition sources.

Concerning the nature of combustibles, existing research primarily relies on the Fire Weather Index (FWI) [11], a well-established and widely employed indicator for predicting the risk of fires. The FWI data can be adjusted to account for regional variations, thereby ensuring the accuracy of fire risk predictions in different areas. Nadeem et al. successfully employed the FWI index to achieve an AUC greater than 0.9 in forecasting the danger of lightning-induced fires [11]. However, one should recognize that the FWI relies on successive iterative calculations, and its spatial accuracy is constrained by the limited variability of the weather indicators within a given region. This limitation undermines efforts to enhance the precision of fire predictions. Alternatively, remote sensing data offers numerous advantages that can significantly contribute to fire risk assessment. These advantages include the swift availability of information, comprehensive coverage [12,13], and the capacity for continuous monitoring [14]. Moreover, remote sensing data can provide real-time information about the moisture content of surface combustible materials [8,15–17] by utilizing various spectral bands [18]. As a result, remote sensing data have the potential to replace FWI data in fire risk prediction [19].

When considering the weather conditions that lead to fires, precipitation, temperature, humidity, and wind all play varying roles in the occurrence of fires. Furthermore, lightning-induced forest fire areas tend to concentrate in specific locations due to the relatively fixed path of the lightning-strike thunderstorm system. The investigation revealed that most lightning-induced fires occurred on mountain ridges and high slopes, indicating the influence of topography on the ignition source of lightning fires. Additionally, topography also affects the combustion conditions of the fire (e.g., weather) and its combustible nature (e.g., moisture content). Therefore, this study included topographic data in the subsequent analysis.

Presently, the predominant methodology in research involves the utilization of a grid-based approach. For instance, studies conducted by Nami et al. in Iran and Nicholas et al. in Australia have employed this method, subdividing study regions into 20 × 20 km grids [20,21]. This approach allows for categorizing fire risk within a 1–8 day time scale, enabling focused monitoring of regions with high fire risk [21–25]. However, the investigation into the origins of lightning-induced fires, specifically lightning strikes, lacks a well-defined direction. In their study of lightning fires in the Portuguese region, Couto et al. conducted simulations of vegetation and lightning in a controlled laboratory setting to gain theoretical insights into the mechanisms underlying lightning-induced fires [26]. However, one must note that the laboratory environment does not represent real fire situations. In predicting future lightning-induced fires, researchers have relied on historical

data on lightning frequency. For example, Ruth et al. used historical data for lightning fire prediction in the Australian region [9,27,28].

However, due to the increasing occurrence of abnormal weather patterns in recent years, historical data on lightning strike frequencies may be subject to delays and reduced predictive reliability. Therefore, it becomes necessary to address these issues by combining realistic lightning fire data to identify the patterns of lightning fire occurrences and utilizing current lightning strikes to determine fire incidents rather than relying solely on historical lightning strike frequencies. Moris et al. [29] demonstrated a noteworthy delay between lightning strikes and subsequent incidents of lightning-induced fires. Considering this delay, one can predict the likelihood of each lightning strike causing a fire.

In conclusion, this study contributes to the understanding of lightning-induced fires by examining various methods of lightning detection and analyzing data on these fires, including the timestamps of lightning strikes. This study emphasizes the importance of identifying ignition lightning strikes in fire prediction and underscores the significance of remote sensing data in predicting such fires by integrating it with weather and topographical information. Moreover, a logistic regression model was used to assess the likelihood of lightning-induced fires occurring with each lightning strike. This model enables early detection of ignition-causing fires upon lightning occurrence, which aids in implementing proactive fire suppression measures before further escalation. Ultimately, this study identifies key factors that contribute to lightning-induced fires.

2. Materials and Methods

2.1. Study Area

The study area is situated in northern China, specifically in the GKM region (latitude 50.7°–53.7° N, longitude 121.1°–127.1° E). The approximate size of the study area is 1.36×10^7 hectares, as shown in Figure 1. The regional climate is categorized as a cool temperate zone, with an average annual temperature ranging from −2 to −4 °C. Additionally, annual precipitation levels in the area range between 350 and 500 mm. The elevation within the study area varies from 300 to 1400 m. The primary forest types found in this region consist of *Larix gmelinii* forests, which include grass *Larix gmelinii* forests, Rhododendron *Larix gmelinii* forests, and *Thino-pyrum pumila Larix* forests. Other forest types include *Pinus sylvestris* forests, coniferous and broad-leaved mixed forests (white *Larix gmelinii* forests), and broad-leaved forests, which encompass white forest Mongolia forests, *Populus davidiana* forests, and others. Shrub vegetation predominantly comprises coniferous shrubs (*Pinus pumila* shrubs) and broad-leaved shrubs such as *Betula platyphylla* and hazelnut. The dominant grassland type in the GKM area is Goutang meadow (comprising *Carex*, *Cinnamomum camphora*, etc.) [30]. This region is prone to frequent summer thunderstorms, making it a high-incidence area for forest lightning fires in China. It holds the highest rank in the nation in terms of both the number of forest lightning fires and the extent of forest damage.

2.2. Fire and Lightning Monitoring Data

This study was conducted from 2005 to 2009, utilizing fire record data and lightning monitoring data from the relevant government agencies. The fire record in the GKM region contained the location, discovery time, and cause of each fire. During the five years, the region experienced 145 forest fires, with 99 of them attributed to lightning strikes, accounting for 68.28% of all incidents. This study solely focused on data related to lightning-induced fires.

Lightning-induced fires occurred between 10 May and 27 September of each year. We extended the time frame to April to October each year to ensure comprehensive coverage of lightning fire prediction research and minimize potential omissions. Moreover, determining the exact timing of lightning-induced forest fires poses significant challenges. The most recorded forest fires were detected from 9:00 to 20:00, with more than 50% happening

between 13:00 and 17:00 (Figure 2). Incidents during the morning and night were infrequent. Thus, this study is organized based on daily units.
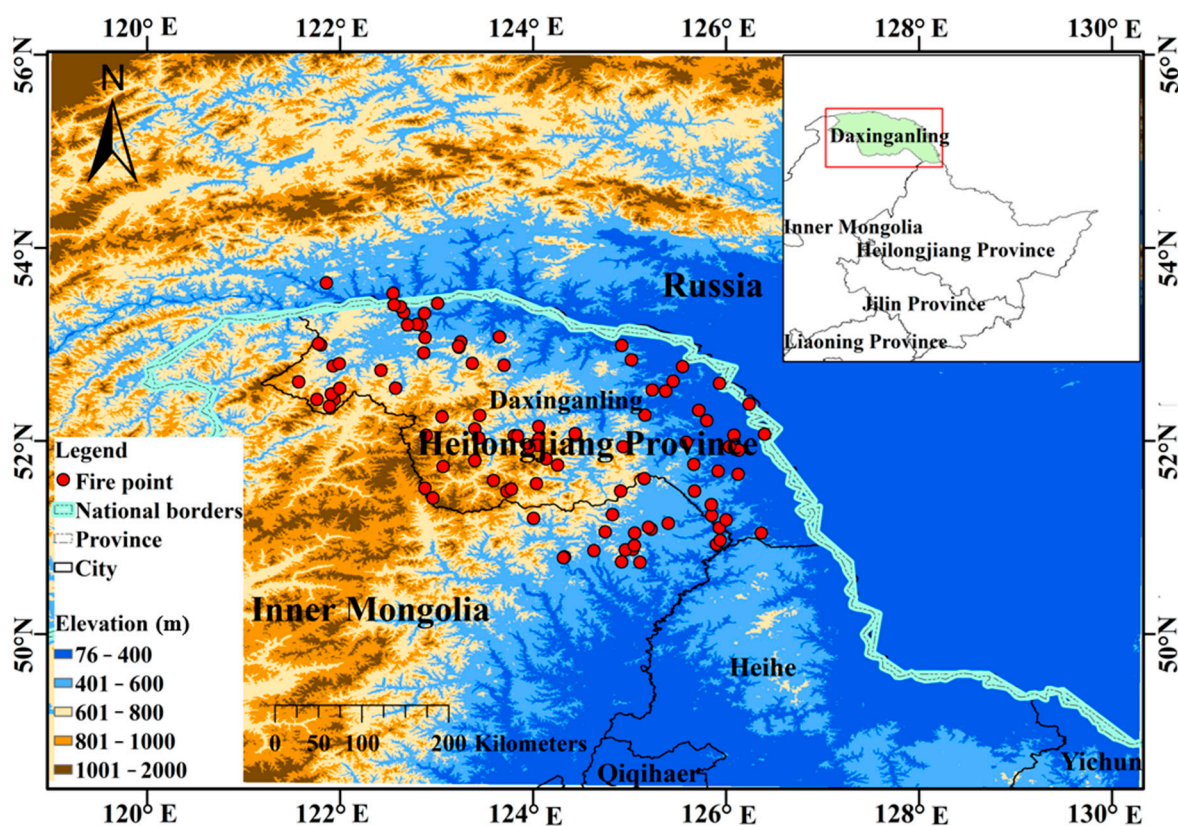

Figure 1. The study area around the GKM is delineated, indicating the locations of lightning-induced fires between 2005 and 2009.
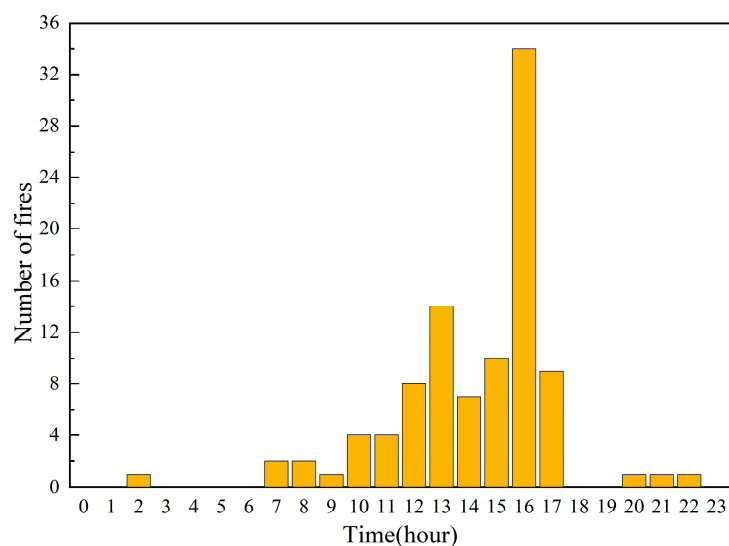
Figure 2. Temporal distribution of lightning forest fires in the Greater Khing'an Mountains from 2005 to 2009.

The lightning monitoring data included records of lightning strike times, locations, and four key parameters: intensity, slope, charge, and energy. This study aimed to identify the initiating lightning strokes and assess the potential significance of lightning stroke parameters in lightning-induced fire events. The analysis focused on ascertaining the

onset time of lightning-induced fires triggered by lightning flashes, facilitating a more comprehensive examination of the meteorological conditions associated with these fires. So, there is a need to identify ignition lightning strikes.

2.3. Identifying Igniting Lightning Strokes

For identifying ignition lightning strikes, it is necessary to analyze the errors in the lightning strike and forest fire data in combination with the spatial and temporal lag between lightning strikes and subsequent fires. Two critical data are available for this purpose: the time lag and the spatial distance between the fire and the lightning strike. They are calculated based on when and where the fire was detected and when and where the lightning strike occurred. The popular approach to correlating lightning strikes with fires involves establishing a temporal and spatial buffer centered on the location of the fire. This buffer helps identify potential lightning strike candidates that caused the fire. When configuring the buffer, two crucial parameters must be considered: the maximum temporal window (Tmax) and the maximum buffer radius (Smax) associated with lightning-induced fires. In this study, Tmax is determined based on commonly used values in related research, with two options: 7 days or 14 days. Smax can be set at either 5 km or 10 km. The next step involves establishing a definitive criterion to identify the most probable lightning cause from the pool of candidate lightning events. The central focus of this discussion revolves around finding a balance between the temporal lag and spatial distance between fires and lightning strikes. Broadly, this discussion can be categorized into two types: one that prioritizes the temporal lag issue and another that emphasizes the spatial distance problem. In this paper, we adopted two methods to address the time lag and the spatial distance problem, respectively. Furthermore, this paper introduces a method for calculating the spatiotemporal index that does not focus only on time lag and spatial distance.

Two identification methods prioritize the temporal window. One identifies the candidate lightning strokes with the minimum temporal window (MinT) concerning lightning fires. The other corresponds to the minimum time per kilometer and aims to find the smallest buffer radius among candidate lightning strokes and lightning fires with a temporal window of less than 1 day. If no candidate lightning stroke has a temporal window of less than 1 day, the search expands to less than 2 days and progressively increases up to Tmax (PerSMinT).

Additionally, two identification methods prioritize the buffer radius. The first method identifies the candidate lightning stroke with the minimum buffer radius (MinS) relative to the lightning fire. The second method looks for the smallest temporary window when the distance between the candidate lightning stroke and the lightning fire is less than 1 km. If no candidate lightning stroke has a buffer radius of less than 1 km, the search expands incrementally (starting from less than 2 km) until reaching Smax (PerSMinT). The last method is based on the maximum spatiotemporal index "a" (MaxA) [31–35], which does not focus exclusively on the temporal window or buffer radius. Calculation details are as follows:

$$A = \left(1 - \frac{T}{Tmax}\right) \times \left(1 - \frac{S}{Smax}\right) \tag{1}$$

This comprehensive approach results in 12 matching outcomes (Table 1). To determine the most effective method for identifying igniting strokes causing lightning fires and the optimal combination of Tmax and Smax, this study analyzes factors such as the number of lightning strokes in different outcomes, the similarity between results, and the temporal window and buffer radius of results. Additionally, the electrical characteristics of lightning fires are investigated.

Within the buffer zone for lightning-induced fires, if multiple candidate lightning strokes exist (all within the buffer zone), these strokes are ranked using five identification methods (Table 1). This process aims to identify the lightning stroke most likely to have ignited the fire.

Table 1. Methods were applied to select candidate lightning.

Selection Criterion		Max A		Min Time		Min Dist Per Day		Min Dist		Min Time Per km	
Maximum distance		5 km	10 km	5 km	10 km	5 km	10 km	5 km	10 km	5 km	10 km
Maximum time	7 d	√	√	×	×	×	×	×	√	×	√
	14 d	√	√	√	√	√	√	×	√	×	√

Max A = maximum index A; Min time = minimum holdover time; Min dist = minimum distance; Min dist per day = daily minimum distance; Min time per km = minimum time per kilometer; d = day. √ = method applied; × = method not applied.

2.4. Supplementary Data Sources for Acquiring Predictive Driving Factors

This study utilized publicly available datasets from the Google Earth Engine (GEE), including meteorological, terrain, and remote sensing data, to investigate the prediction of lightning-induced fires. These datasets have been extensively validated and widely utilized in various scientific studies, including those related to fire risk assessment (Table 2). Subsequently, we computed statistical parameters of variables within a 20 km × 20 km grid (e.g., mean and variance) to assess the strength of influencing factors in predicting lightning fires.

Table 2. Dataset information.

Dataset Name	Website URL	Spatial Resolution	Temporal Resolution	Covariate Name	Units
ERA5 DAILY	https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_DAILY#description (accessed on 1 January 2024)	27,830 m	Daily	mean_2m_air_temperature minimum_2m_air_temperature maximum_2m_air_temperature dewpoint_2m_temperature total_precipitation surface_pressure mean_sea_level_pressure u_component_of_wind_10m v_component_of_wind_10m	K K K K m Pa Pa m/s m/s
ERA5-Land	https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_LAND_DAILY_AGGR#description (accessed on 1 January 2024)	11,132 m	Daily	snow_depth	m
SRTM Digital Elevation Data Version 4	https://developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4#description (accessed on 1 January 2024)	90 m	-	elevation	m
MCD12Q1 V6	https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MCD12Q1#description (accessed on 1 January 2024)	500 m	Yearly	Land Cover Type 1	-
MODO9GA version 6.1	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD09GA#description (accessed on 1 January 2024)	500 m	Daily	sur_refl_b01 sur_refl_b02 sur_refl_b03 sur_refl_b04 sur_refl_b05 sur_refl_b06 sur_refl_b07	- - - - - - -

Table 2. *Cont.*

| Dataset Name	Website URL	Spatial Resolution	Temporal Resolution	Covariate Name	Units
MOD09A1 V6.1	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD09A1#description (accessed on 1 January 2024)	500 m	8-day	sur_refl_b01 sur_refl_b02 sur_refl_b03 sur_refl_b04 sur_refl_b05 sur_refl_b06 sur_refl_b07	- - - - - - -

The meteorological data were obtained from the ERA5 daily dataset provided by the European Center for Medium-Range Weather Forecasts (ECMWF) and served two purposes: to enhance forecast inputs and to calculate the Forest Fire Weather Index (FWI). The FWI, which consists of the Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), and Initial Spread Index (ISI) [36], is crucial for methodological comparisons. These indices were derived from empirical formulas using precipitation, surface temperature, relative humidity, and wind speed. Relative humidity was computed based on dew point temperatures from the ERA5 dataset, employing the Magnus-Tetens approximation method [37]. In addition, snow depth was used to assess snow melt time and calculate the FWI.

Terrain data from the Space Shuttle Radar Terrain Mission (SRTM) facilitated the computation of average elevation and the distribution of slope aspects (East, West, South, and North). These topographical characteristics provided valuable insights into the study area.

Remote sensing data were instrumental in identifying predictive factors for lightning fires. The "Type1" category from the MCD12Q1.006 MODIS Land Cover Type Annual Global 500 m dataset was used for vegetation distribution analysis, focusing on deciduous coniferous forest and grassland. The MOD09GA (Terra Surface Reflection Daily Global 1 km and 500 m) and MOD09A1 (Terra Surface Reflection 8-Day Global 500 m) were used for vegetation water content. These datasets provided essential information on vegetation moisture levels, contributing to the analysis of fire susceptibility and lightning fire prediction.

2.5. Preprocessing of Remote Sensing Data

In previous studies [11], the utilization of remote sensing data has been infrequent, primarily relying on the NDVI index to depict the vegetation status within the research region. This paper emphasizes the daily moisture content of vegetation, employing a broader range of remote sensing indices associated with vegetation moisture content to enhance the prediction of lightning-fire incidents. Consequently, in the satellite selection process, MODIS data, offering daily temporal resolution, was chosen. Nevertheless, remote sensing data frequently encounter the challenge of cloud cover [38,39]. To mitigate this issue, this study employed two datasets, MOD09GA and MOD09A1, in a complementary manner. Given that changes in vegetation moisture content occur gradually for regions in the daily dataset MOD09GA that are obscured by clouds, cloud removal was performed in this study by combining the eight-day composite data from MOD09A1 preceding the relevant day, representing the best-pixel composite image within an eight-day window. In the mosaicking process, the average digital number (DN) values of pixels without clouds in both MOD09A1 and MOD09GA were scaled to match, minimizing potential discrepancies due to daily weather conditions and other variables (Figure 3).
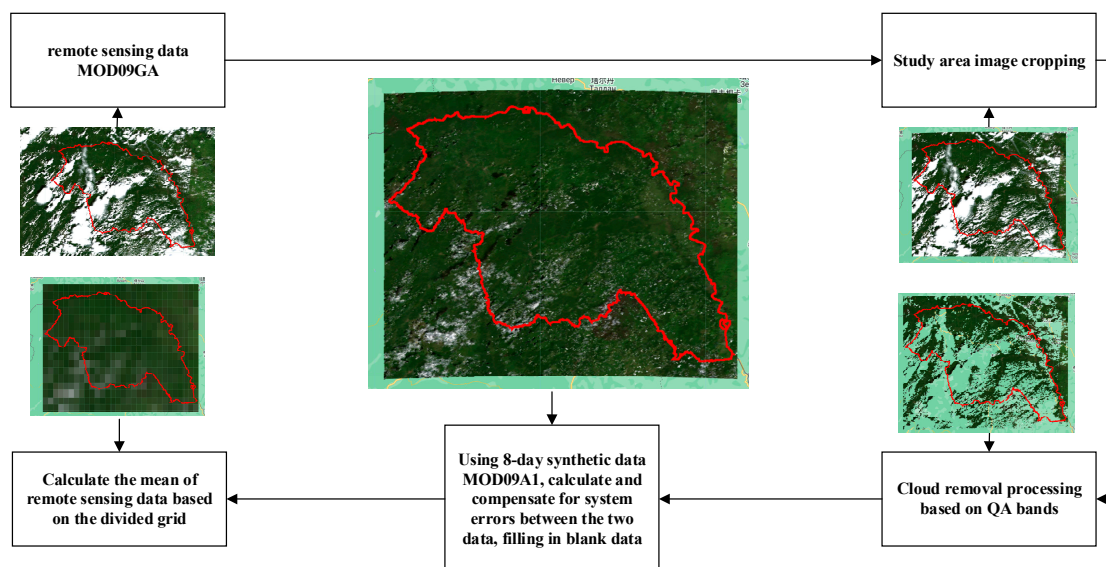
Figure 3. Extraction of Vegetation Information Using MOD09GA and MOD09A1 Datasets.

In the context of extracting vegetation moisture content through remote sensing, two methods were considered: single-band inversion and multi-band remote sensing index inversion [7]. Consequently, among the factors influencing lightning fires, the original band data of the image (sur_refl_b01-sur_refl_b07) and specific remote sensing indices commonly used in prior studies to assess combustible moisture content were selected. The original band data and data from remote sensing indices were averaged within the grid to produce the final driving factor data.

2.6. Lightning Fire Prediction Mode and Variable Selection

This study employs logistic regression models to predict the likelihood of lightning strikes leading to fires by incorporating meteorological, lightning, vegetation, and topographic data. Table 3 lists the specific variables used in the model. Meteorological indicators such as temperature, wind speed, wind direction, and humidity, among others, are well-established and widely recognized for their significance in assessing fire risk [40,41]. Previous research has demonstrated the efficacy of meteorological monitoring data in identifying areas with the highest risk of lightning-induced wildfires [29]. Furthermore, considering lightning data as the ignition source for lightning fires is imperative to assess lightning fire risk [8]. Section 3.1 of this paper describes the lighting parameter selection process in detail.

In the context of vegetation, moisture content significantly influences the incidence of fires [42]. Alterations in the moisture content of vegetation impact its spectral reflectance across various bands. Generally, as vegetation's moisture content rises, its reflectance increases in the near-infrared band and diminishes in the visible spectrum. This phenomenon occurs because moisture within vegetation absorbs light in the near-infrared band while simultaneously reflecting light in the visible spectrum. Consequently, numerous remote sensing indices have been developed to gauge vegetation moisture content [8,43]. Given that the spectral properties of the vegetation vary across different bands, the reflections on vegetation moisture content differ accordingly. Therefore, this study initially computed multiple remote sensing indices, aiming to assess their efficacy in predicting lightning strikes and fires. It includes the Bare Soil Index (BSI) [44,45], Enhanced Vegetation Index (EVI) [46], Greenness Index (GRATIO) [47], and Normalized Difference Moisture Index (NDMI) [48]. These indices have been used in studies related to fires, vegetation, and other related research areas. Vegetation types also have a significant impact on the occurrence of lightning fires [49].

Table 3. All logistic regression model variables in this article.

Variables	Describe	R Model	F Model	R + F Model
Intension	Parameters of individual lightning	√	√	√
Slope	Parameters of individual lightning	√	√	√
Bare soil index (BSI)	Remote sensing index (Mean value of data within the grid)	√		√
Enhanced vegetation index (EVI)	Remote sensing index (Mean value of data within the grid)	√		√
Greenness index (GRATIO)	Remote sensing index (Mean value of data within the grid)	√		√
Normalized Difference Moisture Index (NDMI)	Remote sensing index (Mean value of data within the grid)	√		√
Sur_refl_b02	MODIS Band 2 (Mean value of data within the grid)	√		√
Sur_refl_b04	MODIS Band 4 (Mean value of data within the grid)	√		√
maximum_2m_air_temperature	Maximum air temperature at 2 m height (daily maximum) (Mean value of data within the grid)	√	√	√
total_precipitation	Total precipitation (daily sums) (Mean value of data within the grid)	√	√	√
u_component_of_wind_10m	10m u-component of wind (daily average) (Mean value of data within the grid)	√	√	√
v_component_of_wind_10m	10m v-component of wind (daily average) (Mean value of data within the grid)	√	√	√
aspectArea0	Percentage of area with an east slope within the grid	√	√	√
aspectArea1	Percentage of area with a west slope within the grid	√	√	√
aspectArea2	Percentage of area with a south slope within the grid	√	√	√
aspectArea3	Percentage of area with a north slope within the grid	√	√	√
Elevation variance	The variance of elevation within the grid reflects the degree of ground folding	√	√	√
elevation	Elevation (Mean value of data within the grid)	√	√	√
Vegetation1	Deciduous Needleleaf Forests: dominated by deciduous needleleaf (larch) trees (canopy > 2 m). Tree cover > 60% (Percentage of area within the grid)	√	√	√
Vegetation2	Grasslands: dominated by herbaceous annuals (<2 m) (Percentage of area within the grid)	√	√	√
FFMC	FWI index (Mean value of data within the grid)		√	√
DMC	FWI index (Mean value of data within the grid)		√	√
DC	FWI index (Mean value of data within the grid)		√	√
ISI	FWI index (Mean value of data within the grid)		√	√

R model = Remote sensing data logistic regression; F mode = FWI data logistic regression; R + F model = Remote sensing data and FWI data logistic regression.

In terrain data analysis, elevation within grids is utilized to calculate variance, reflecting the topographic relief, which has been validated in predicting lightning fire risk [11]. Additionally, data such as slope and aspect are commonly used in fire risk prediction as well [50].

In this study, the involvement of several indices reflecting vegetation water content may result in linear correlations between them. These correlations may adversely affect the logistic regression model. Therefore, the correlations between the variables must be evaluated. The variance inflation factor (VIF) method [51] was employed to test for multicollinearity between variables. Subsequently, variables with VIF values equal to or greater than 10 were removed from the analysis to mitigate the multicollinearity problem.

2.7. Data Sampling and Collinearity Diagnosis

Although the dataset used in this study includes over 450,000 records of lightning detection and location data from 2005 to 2009, lightning-caused forest fires were limited to only 99 instances. The data imbalance exists in many directions of fire research, such as fire identification [52,53]. The severe imbalance in the data challenges the construction of predictive models. Early in the experiment, the accuracy of the models derived from the severely unbalanced training set is significantly influenced by the negative samples in the training set, resulting in the model erroneously classifying all samples as counterexamples. In order to solve this problem, Haeng Yeol Oh et al. tried to compare multiple methods for oversampling positive samples in fire risk prediction, among which borderline-SMOTE obtained an accuracy of 79.54% on its data, proving the effectiveness of the method [54]. Nadeem et al. used a simple random sampling method to downsample negative samples [11].

However, the data imbalance in this study is even more severe than in the above articles. To tackle this issue, we implemented a combination of oversampling and undersampling on the training set. Specifically, we employed the Synthetic Minority Oversampling Technique (borderline-SMOTE) [55] algorithm to oversample data representing lightning strikes (positive samples) associated with forest fires. Simultaneously, lightning strike data representing no forest fires (negative samples) were undersampled by random deletion. This approach resulted in a well-balanced training set. The borderline-SMOTE algorithm is specifically designed to improve the performance of classifiers when dealing with unbalanced datasets. It increases the number of minority class samples by identifying boundary samples and generating synthetic data, thereby enhancing the classifier's ability to recognize these instances. This method demonstrated favorable performance in our study.

In the test set, positive sample data were not oversampled, but negative samples underwent undersampling to maintain some level of imbalance and reflect real-world conditions. Ultimately, we established a test set with a positive-to-negative sample ratio of 1:10. When evaluating the model's performance, standard binary classification evaluation methods were applied, incorporating the use of a confusion matrix and ROC curve analysis. Additionally, we scrutinized the significance of testing results to assess the importance of the model's driving factors. Lastly, we explored fitting a logistic regression model excluding remote sensing data but incorporating the Forest Fire Weather Index (FWI) for comparison with the model incorporating remote sensing information.

3. Results

#### 3.1. Analysis of Igniting Strokes Identification Results

This section evaluates the 12 diverse methods used to identify ignited lightning strikes, as outlined in Section 2.3, Determining Ignited Lightning Strikes. There were 99 lightning-ignited fires recorded over five years, from 2005 through 2009. The number of successful matches is displayed in Table 4. In the best case in Table 2 (Tmax = 14 days, Smax = 10 km), only 65 fires (65.66%) were associated with identified igniting strokes. In cases of lightning-induced fires lacking outcomes, determining whether the absence of results arises from errors in lightning or fire records or from lightning strokes surpassing the buffer area of the respective fire becomes unfeasible. These results are excluded from subsequent model development to ensure data accuracy.

Table 4. The number of successful matches is depicted.

| Selection Criterion		Max A		Min Time		Min Dist Per Day		Min Dist		Min Time Per km	
Maximum distance		5 km	10 km	5 km	10 km	5 km	10 km	5 km	10 km	5 km	10 km
Maximum time	7 d	33	51	×	×	×	×	×	51	×	51
	14 d	43	65	43	65	43	65	×	65	×	65

Max A = maximum index A; Min time = minimum holdover time; Min dist = minimum distance; Min dist per day = daily minimum distance; Min time per km = minimum time per kilometer; d = day. × = method not applied.

In scenarios where various approaches result in an equal count of lightning occurrences, variations are evident among the selected lightning events within the buffer zone. Particularly, a higher degree of similarity is observed among methods m7–m12 (Figure 4). Similarly, a distinct resemblance is noted among occurrences of the Max A method. However, a comparatively lower level of similarity is observed between these two groups of methods. This study indicates a decreasing trend in occurrences of lightning-induced fires with an increase in sample size, temporary window duration, and buffer zone radius.
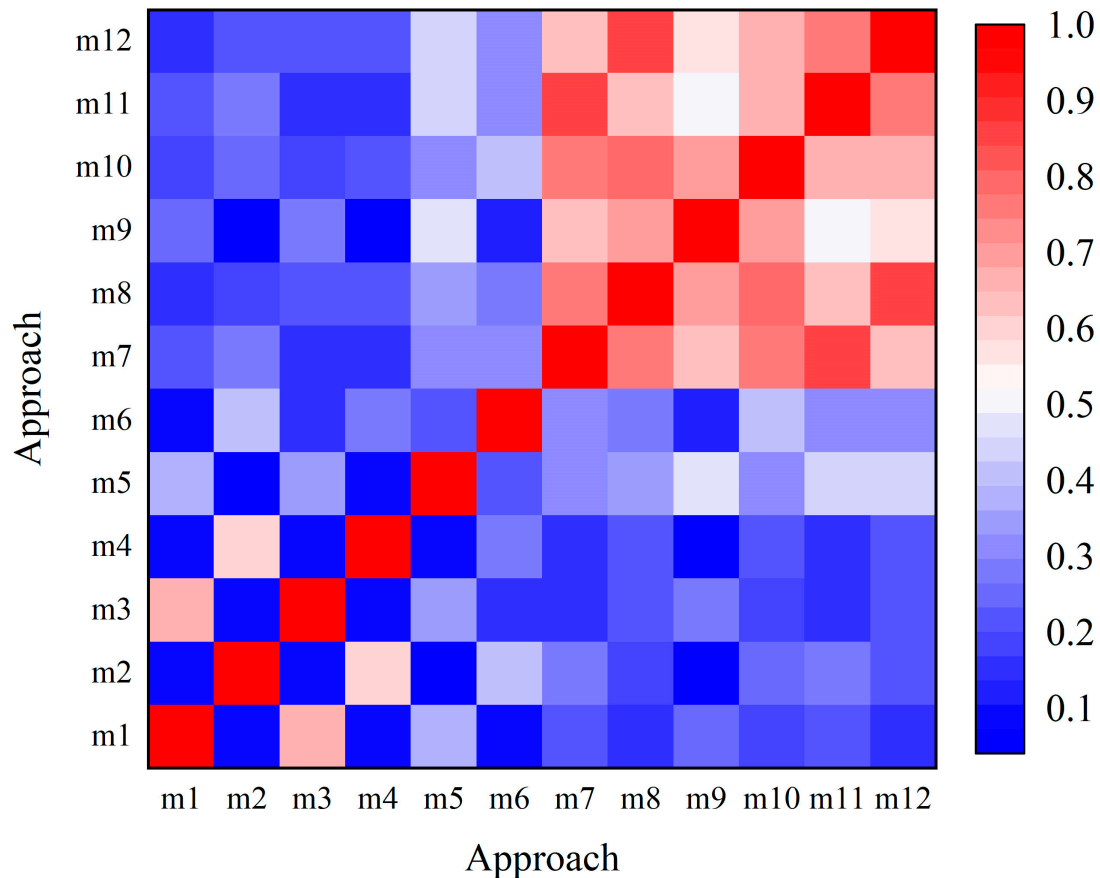

Figure 4. Similarity of results from different approaches. m1–m12. Methods m1 to m12 are sequentially Max A_7d5km; Max A_7d10km; Max A_14d5km; Max A_14d10km; Min time_5km; Min time_10km; Min dist_7d; Min dist_14d; Min dist per day_5km; Min dist per day_10km; Min time per km_7d; Min time per km_14d.

In light of this observation, we analyzed distinct outcomes (Figure 5) to assess the relationship between buffer radius and lightning occurrences. However, in this study, a decreasing trend is evident in the overall distribution of lag time, with the exception of a notable decrease in data volume within a lag time of 1–2 days. Conversely, most distributions in lag distance exhibit an ascending trend. This divergence may be attributed to inaccuracies and significant errors in lightning locating devices. Consequently, it is foreseeable that prioritizing distance methods, including Min dist_7d, Min dist_14d, Min dist per day_5km, Min dist per day_10km, Min time per km_7d, and Min time per km_14d, may yield inconclusive results. However, the comparison of methods hinges primarily on the accuracy of the lightning fire prediction model formulated through their utilization. Various methods are employed to meticulously gather sets of lightning events that initiate lightning strikes. This is conducted with the aim of conducting a rigorous statistical analysis of lightning characteristics susceptible to causing fires.
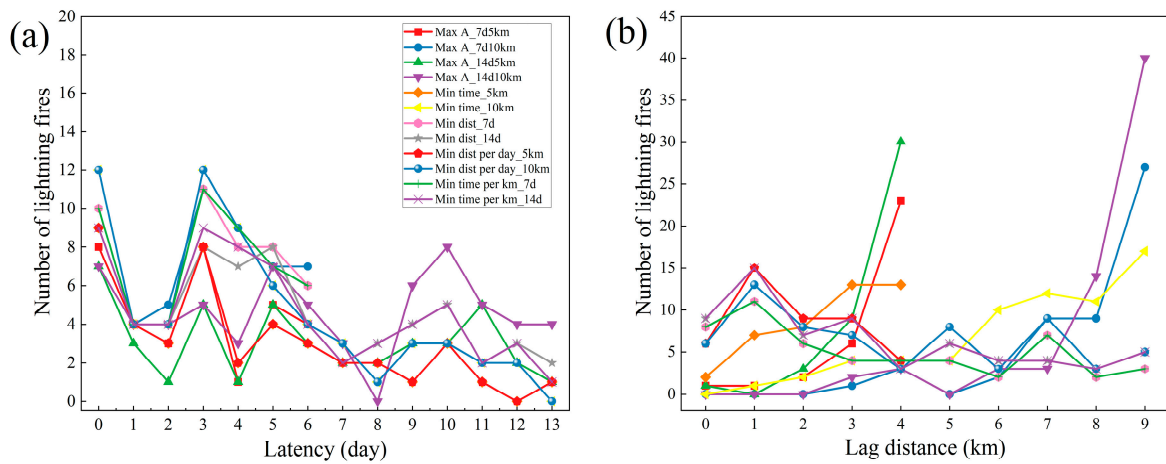
Figure 5. Distribution of latency (**a**) and lag distance (**b**) for each method.

Subsequently, we conducted an analysis of the matched lightning indicators (Figure 6). It is noteworthy that, in comparison to all lightning strikes, the lightning strikes causing fires tend to be more concentrated within the intensity and slope range of −10 to 0. This concentration might indicate a relationship with the occurrence of fires. Regarding energy and charge, no significant differences were observed. Therefore, subsequent lightning parameters were narrowed down to exclusively encompass lightning intensity and slope.
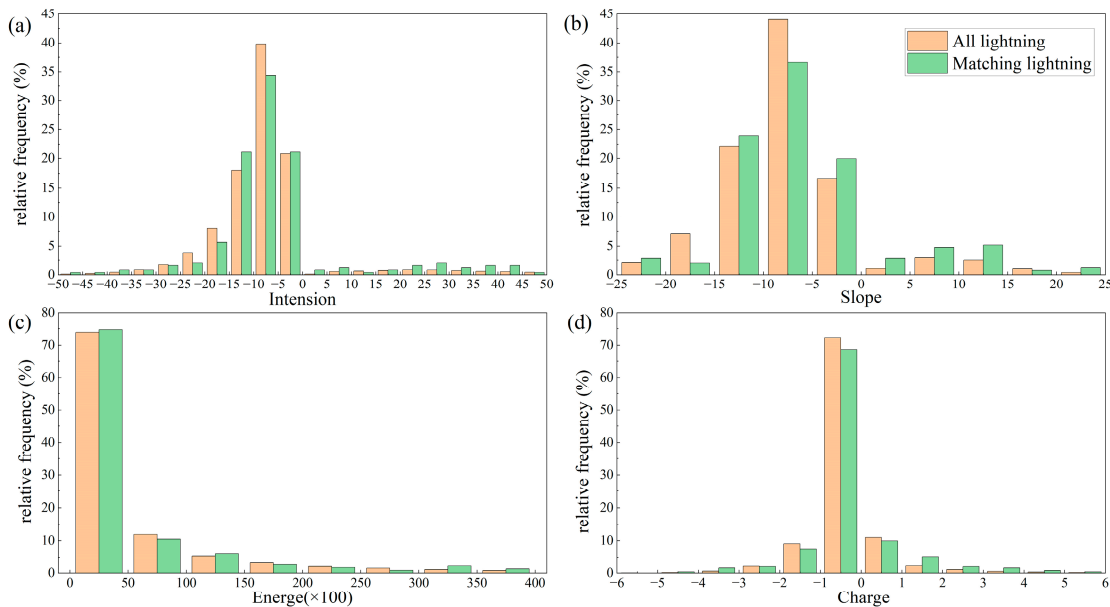
Figure 6. Distribution of all lightning and matching lightning parameters (ignoring some data intervals with a proportion less than 1%, which may be caused by errors) (**a**) Intension, (**b**) Slope, (**c**) Energy, and (**d**) Charge.

3.2. Logistic Regression Prediction Model Results of Remote Sensing Data

To substantiate the validity of the methods in the GKM region, this study adopted a consistent approach to developing logistic regression models for 12 sets of ignited stroke outcomes. Following the alleviation of data collinearity effects, the model retained a total of 20 variables, as outlined in Table 3.

In model training, a significant amount of information becomes lost during the model fitting process owing to the undersampling of negative samples. Therefore, in this study, we employed multiple fittings during the model establishment process, and the final

model results were derived from the average fitting outcomes. This method has been used in the study of Nadeem et al. with good results [11]. Since the ratio of the number of negative samples contained in a training set to the total number of negative samples is about 500, the number of fits for each model is configured at 500 to effectively utilize the substantial majority of negative samples. The training predominantly employs scikit-learn 1.2.0 (sklearn) [56], with max_iter set to 1000 and other parameters left at their default values.

The average AUC results for 500 model iterations are depicted in Table 5. Firstly, it is evident in the logistic regression (R) model using remote sensing data that different numbers of results significantly impact the model's accuracy. Among these, when Tmax = 14 days and Smax = 10 km, the results are notably better. This indicates that these results provide a richer sample set, leading to the best-fitting effect of the model. Secondly, for the four groups of data with successful matches of fewer than 51, their positive sample data are reduced by 51.2% to 97.0%. While these data may show better identification effects within the sample after undergoing the same sampling process, their performance on the validation set is inferior to the former. However, overall, the differences in their accuracy are not significant.

Table 5. All the AUC results of logistic regression models in this article.

Selection Criterion	Successful Matches	R Model	F Model	R + F Model
maxA_14d10km		0.849	0.807	0.891
minT_10km		0.803	0.750	0.856
minS_14d	65	0.773	0.712	0.834
perDminS_10km		0.802	0.749	0.855
perSminD_14d		0.766	0.712	0.82
maxA_7d10km		0.798	0.743	0.852
minS_7d	51	0.724	0.663	0.785
perSminD_7d		0.74	0.679	0.804
maxA_14d5km		0.785	0.719	0.85
minT_5km	43	0.754	0.681	0.826
perDminS_5km		0.806	0.744	0.868
maxA_7d5km	33	0.785	0.721	0.872

R model = Remote sensing data logistic regression; F mode = FWI data logistic regression; R + F model = Remote sensing data and FWI data logistic regression.

Among groups with the same number of successful matches, maxA consistently achieves the best or second-best accuracy. Particularly, the maxA_14d10km method yields the optimal results. In Nadeem et al.'s study, this method attained the highest accuracy in identifying ignition points [11]. In this study, its outstanding performance may similarly be attributed to its precise identification of lightning strikes.

To assess the effectiveness of remote sensing data in predicting lightning-induced fires, this study conducted a comparative analysis with the widely utilized Fire Weather Index (FWI) in the field. Following the mitigation of data collinearity, as illustrated in Table 3, this study proceeded to evaluate successful matches and compare the accuracy of different models based on identification methods. The conclusions align with the results from the R model. However, when excluding remote sensing data from the model input and incorporating FWI (Fire Weather Index) data, there is a slight overall decrease in model accuracy. This could be attributed to the fact that the FWI index is calculated based on weather data. While it correlates well with surface moisture content, the information it contains can also be considered part of the weather data. This reduction in input factors may consequently lead to the observed decrease in accuracy.

Finally, the combination of remote sensing data with FWI data was performed, alleviating the impact of data collinearity. In this combined model, 24 variables persisted, as shown in Table 5. The overall AUC values for the model were higher than using any single data source, reaching the highest model accuracy. The optimal AUC value, 0.891, was attained using the method maxA_14d10km. The reason could be that FWI data compensate

for the non-linear effects of weather on fire occurrences, further enhancing the accuracy of the model based on the R model.

4. Discussion

4.1. Response of Fire Prediction to Different Lightning Strike Identification Methods

Our results underscore the critical importance of method selection in the modeling process when predicting whether a lightning strike can further escalate into a wildfire. This includes choices related to the lightning strikes that trigger fires and the driving factors considered in wildfire prediction. The error distribution of unknown lightning strike data significantly influences the outcome, with various spatiotemporal matching windows and methods impacting the final results. In Section 3.1, we analyze the lag time and spatial distribution trends of results from various methods, ultimately identifying a preferred matching method. The findings can be summarized as follows: Distance-based prioritized matching outperforms spatiotemporal index methods, which in turn outperform prioritized time matching. In Section 3.2, we rank matching methods based on different model prediction accuracies. The results can be summarized as follows: spatiotemporal index methods outperform prioritized time matching, which in turn outperforms distance-based prioritized matching. The main contradiction arises from the significant differences in the performance of distance-based prioritized sorting methods. One reason for this may be the trend analysis for the former, possibly because of the high number of potential lightning strikes during thunderstorms within a short time frame. This abundance makes it easier for even misjudged lightning strikes to satisfy the time distribution, leading to a trend that is easily met in an exponential distribution but challenging in a spatial distribution. This results in a more accurate evaluation of the distance-based prioritized method. The second reason is associated with the model's low spatial resolution, diminishing the significance of distance and leading to a less effective assessment of the distance-based prioritized method. Moreover, prevalent errors in the lightning detection and positioning system's distance, combined with limited data on lightning-induced fires, exacerbate this issue. Nevertheless, taking all factors into account, the spatiotemporal index method emerges as the recommended approach for lightning-fire matching in this study. It excels by concurrently addressing the temporal and spatial lag of lightning-induced fires, ensuring enhanced stability and consistently favorable results across all experiments.

4.2. Response of Fire Prediction to Explanatory Variables

In terms of selecting driving factors for wildfire prediction, the Fire Weather Index (FWI) model shows comparable performance to the remote sensing data model. Different lightning selection methods result in models displaying diverse AUC values. This validates the feasibility of the remote sensing data model. Additionally, it suggests that, in future research, enhancing the temporal and spatial resolution of remote sensing data can improve the prediction of lightning-induced fires, providing greater development potential. The combination of remote sensing data with FWI produces the highest predictive accuracy for lightning-induced fires.

The significance analysis of all variables is presented in Table 6. Notably, there is a breakpoint observed in the absolute values of t-stats around 30. Eleven variables, namely BSI, sur_refl_b02, maximum_2m_air_temperature, v_component_of_wind_10m, elefang, elejun, vegetation3, vegetation10, FFMC, DC, and ISI, are deemed important as their absolute t-stats exceed 30.

Within meteorological factors, high temperatures may lead to drier vegetation, consequently elevating the likelihood of lightning fires. Among the factors influencing lightning fires, the maximum daily air temperature exhibits a notably positive impact, demonstrating statistical significance. Conversely, higher wind speeds are associated with a decreased likelihood of lightning fires. The v-component of wind speed at 10 m height exhibits a negative impact, indicating that higher north winds are more likely to cause lightning fires. Northern winds may introduce particular meteorological conditions, such as decreased pre-

cipitation and humidity, contributing to vegetation desiccation and a heightened likelihood of lightning fires.

Table 6. Logistic regression results.

Name	Coef	Std Error	Tval	Pval	Lower Alpha	Upper Alpha
constant	−0.002	0.012	−0.198	0.843	−0.025	0.021
Intension	−0.454	0.021	−21.860	<0.001	−0.495	−0.413
Slope	−0.155	0.021	−7.339	<0.001	−0.196	−0.114
BSI	−1.477	0.026	−57.089	<0.001	−1.528	−1.426
EVI	0.419	0.017	24.817	<0.001	0.386	0.452
GRATIO	−0.462	0.019	−23.875	<0.001	−0.500	−0.424
NDMI	0.254	0.029	8.862	<0.001	0.198	0.310
sur_refl_b02	0.870	0.026	32.980	<0.001	0.819	0.922
sur_refl_b04	−0.271	0.024	−11.118	<0.001	−0.319	−0.223
maximum_2m_air_temperature	1.027	0.016	62.284	<0.001	1.059	0.995
total_precipitation	0.126	0.018	7.114	<0.001	0.091	0.161
u_component_of_wind_10m	0.059	0.012	4.900	<0.001	0.035	0.082
v_component_of_wind_10m	−1.083	0.012	−88.934	<0.001	−1.107	−1.059
aspectArea0	−0.493	0.020	−24.123	<0.001	−0.533	−0.453
aspectArea1	−0.707	0.027	−26.407	<0.001	−0.759	−0.654
aspectArea2	−0.197	0.028	−7.052	<0.001	−0.251	−0.142
aspectArea3	−0.599	0.022	−26.682	<0.001	−0.643	−0.555
elefang	−1.080	0.021	−51.294	<0.001	−1.121	−1.039
elejun	1.389	0.021	65.956	<0.001	1.347	1.430
vegetation3	0.946	0.012	78.286	<0.001	0.922	0.969
vegetation10	0.592	0.011	53.674	<0.001	0.570	0.614
FFMC	1.154	0.025	46.293	<0.001	1.105	1.202
DMC	−0.384	0.017	−22.163	<0.001	−0.418	−0.350
DC	1.474	0.015	98.647	<0.001	1.445	1.503
ISI	0.622	0.018	34.403	<0.001	0.586	0.657

Concerning remote sensing information, the Bare Soil Index (BSI) indicates the extent of exposed soil on the surface. A low BSI likely indicates higher vegetation density and coverage, negatively impacting lightning fire occurrences. The near-infrared band of the MODIS satellite, sensitive to chlorophyll content and vegetation structure, notably influences lightning fires, based on the study results.

Within the Fire Weather Index (FWI), the Fine Fuel Moisture Code (FFMC) reflects the humidity of fine combustible materials, with higher values indicating drier conditions. The Drought Code (DC) represents deep soil moisture conditions, with higher values indicating drier soil. The Initial Spread Index (ISI) reflects the initial spread speed of fires and is influenced by factors such as fine combustible material humidity, wind speed, and fire intensity. All three FWI variables have a significant positive impact on lightning fires in the study.

In terms of terrain information, higher elevations often experience more extreme temperatures and humidity, leading to unstable weather conditions and an increased probability of lightning activity. Lower elevation variances suggest flatter and more stable terrain, fostering organized local airflow and increasing the likelihood of lightning activity. Conversely, higher elevation variances may introduce more turbulence, potentially decreasing the likelihood of lightning occurrences.

Concerning vegetation types, deciduous needleleaf trees, which lose leaves in the dry season, are more susceptible to dry and combustible conditions. Herbaceous plants, with short lifecycles and leaf loss in the dry season, are also prone to dry and combustible conditions. This increases the likelihood of vegetation becoming fuel for lightning fires. Both Deciduous Needleleaf Forests and Grasslands show a significant positive impact on lightning fires.

In conclusion, remote sensing information is highlighted as a crucial factor in predicting lightning fires, alongside meteorological factors, terrain information, vegetation types, and FWI indices, all of which have nearly equal importance. The identified significant factors align with past research results and theoretical models.

4.3. Model Results

We propose a logistic regression model incorporating remote sensing data to predict whether lightning-induced fires occur before they become detectable. The model achieved an AUC of 0.891. In comparison with other studies, direct comparisons are challenging due to differences in research focus. However, previous studies have reported AUC values of 0.859 [21] and 0.84 [57], indicating favorable performance. Additionally, our study improves upon previous research in two aspects. Firstly, by identifying the lightning parameters (intensity and slope) that trigger lightning strikes, the accuracy of the model was improved by 3.6% compared to the model without these two parameters (AUC = 0.860). Secondly, we augmented the model with remote sensing data to conveniently obtain vegetation moisture content information, resulting in a 4.9% improvement over models using only FWI. This enables the early identification of lightning-induced fires when they occur, thereby facilitating proactive measures to extinguish fires before they escalate.

However, our study has several limitations. Firstly, the availability of remote sensing data is affected by cloud cover, leading to decreased data timeliness. Future research could explore the integration of multi-source remote sensing data acquisition to mitigate this issue. Secondly, we did not fully leverage the high spatial resolution of remote sensing data and lightning location information. Utilizing these datasets effectively, along with incorporating extensive historical lightning-induced fire data, holds promise for achieving higher accuracy in forecasting lightning-induced fires.

5. Conclusions

As the frequency of extreme climate conditions rises, predicting future lightning fires becomes increasingly challenging. Due to the lag in the occurrence of lightning fires in relation to lightning events, forecasting efforts can be directed toward individual lightning strikes. This requires matching historical data with lightning events. In the establishment of lightning fire prediction models, it is important to identify the methods of lightning strikes that have caused historical fires. Our results indicate that spatiotemporal indices offer a comprehensive approach. In the fitting process of lightning fire prediction models, this study emphasizes the value of remote sensing data in supplementing ground-based vegetation information, thereby enhancing prediction accuracy. Remote sensing data provide superior resolution compared to traditional lightning fire prediction approaches. Leveraging the spatial resolution of remote sensing data has the potential to enhance the precision of future lightning fire predictions to a new level.

Author Contributions: Methodology, Z.Z. and Y.T.; resources, Z.Z. and C.Z.; software, Z.Z. and G.W.; writing, Z.Z.; format calibration, Y.T., F.Z. and C.Z. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Due to project needs, the data cannot be disclosed for the time being.

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

References

1. Wang, D.; Guan, D.; Zhu, S.; Kinnon, M.M.; Geng, G.; Zhang, Q.; Zheng, H.; Lei, T.; Shao, S.; Gong, P. Economic footprint of California wildfires in 2018. *Nat. Sustain.* **2021**, *4*, 252–260. [CrossRef]
2. Randerson, J.T.; Liu, H.; Flanner, M.G.; Chambers, S.D.; Jin, Y.; Hess, P.G.; Pfister, G.; Mack, M.C.; Treseder, K.K.; Welp, L.R.; et al. The impact of boreal forest fire on climate warming. *Science* **2006**, *314*, 1130–1132. [CrossRef]

3. Tzanis, C.; Tsivola, E.; Efstathiou, M.; Varotsos, C. Forest Fires Pollution Impact on the Solar Uv Irradiance at the Ground. *Fresenius Environ. Bull.* **2009**, *18*, 2151–2158.

4. Pereira, G.H.D.; Fusioka, A.M.; Nassu, B.T.; Minetto, R. Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study. *Isprs J. Photogramm. Remote Sens.* **2021**, *178*, 171–186. [CrossRef]

5. Ager, A.A.; Preisler, H.K.; Arca, B.; Spano, D.; Salis, M. Wildfire risk estimation in the Mediterranean area. *Environmetrics* **2015**, *25*, 384–396. [CrossRef]

6. Guo, F.T.; Wang, G.Y.; Innes, J.L.; Ma, Z.H.; Liu, A.Q.; Lin, Y.R. Comparison of six generalized linear models for occurrence of lightning-induced fires in northern Daxing'an Mountains, China. *J. For. Res.* **2016**, *27*, 379–388. [CrossRef]

7. Hessilt, T.D.; Abatzoglou, J.T.; Chen, Y.; Randerson, J.T.; Scholten, R.C.; van der Werf, G.; Veraverbeke, S. Future increases in lightning ignition efficiency and wildfire occurrence expected from drier fuels in boreal forest ecosystems of western North America. *Environ. Res. Lett.* **2022**, *17*, 054008. [CrossRef]

8. Abdollahi, M.; Dewan, A.; Hassan, Q.K. Applicability of Remote Sensing-Based Vegetation Water Content in Modeling Lightning-Caused Forest Fire Occurrences. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 143. [CrossRef]

9. Schumacher, V.; Setzer, A.; Saba, M.M.; Naccarato, K.P.; Mattos, E.; Justino, F. Characteristics of lightning-caused wildfires in central Brazil in relation to cloud-ground and dry lightning. *Agric. For. Meteorol.* **2022**, *312*, 108723. [CrossRef]

10. Cecil, D.J.; Buechler, D.E.; Blakeslee, R.J. Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. *Atmos. Res.* **2014**, *135*, 404–414. [CrossRef]

11. Nadeem, K.; Taylor, S.W.; Woolford, D.G.; Dean, C.B. Mesoscale spatiotemporal predictive models of daily human- and lightning-caused wildland fire occurrence in British Columbia. *Int. J. Wildland Fire* **2020**, *29*, 11–27. [CrossRef]

12. Zang, N.; Cao, Y.; Wang, Y.B.; Huang, B.; Zhang, L.Q.; Mathiopoulos, P.T. Land-Use Mapping for High-Spatial Resolution Remote Sensing Image Via Deep Learning: A Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5372–5391. [CrossRef]

13. Li, R.; Fu, Y.; Bergeron, Y.; Valeria, O.; Chavardès, R.D.; Hu, J.; Wang, Y.; Duan, J.; Li, D.; Cheng, Y. Assessing forest fire properties in Northeastern Asia and Southern China with satellite microwave Emissivity Difference Vegetation Index (EDVI). *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 54–65. [CrossRef]

14. Maffei, C.; Lindenbergh, R.; Menenti, M. Combining multi-spectral and thermal remote sensing to predict forest fire characteristics. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 400–412. [CrossRef]

15. Massetti, A.; Rüdiger, C.; Yebra, M.; Hilton, J. The Vegetation Structure Perpendicular Index (VSPI): A forest condition index for wildfire predictions. *Remote Sens. Environ.* **2019**, *224*, 167–181. [CrossRef]

16. Gibson, R.; Danaher, T.; Hehir, W.; Collins, L. A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. *Remote Sens. Environ.* **2020**, *240*, 111702. [CrossRef]

17. Maffei, C.; Menenti, M. Predicting forest fires burned area and rate of spread from pre-fire multispectral satellite measurements. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 263–278. [CrossRef]

18. Zhang, B.; Chang, L.; Stein, A. Spatio-temporal linking of multiple SAR satellite data from medium and high resolution Radarsat-2 images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *176*, 222–236. [CrossRef]

19. Chowdhury, E.H.; Hassan, Q.K. Operational perspective of remote sensing-based forest fire danger forecasting systems. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 224–236. [CrossRef]

20. Nami, M.H.; Jaafari, A.; Fallah, M.; Nabiuni, S. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. *Int. J. Environ. Sci. Technol.* **2017**, *15*, 373–384. [CrossRef]

21. Read, N.; Duff, T.J.; Taylor, P.G. A lightning-caused wildfire ignition forecasting model for operational use. *Agric. For. Meteorol.* **2018**, *253*, 233–246. [CrossRef]

22. de Bem, P.P.; de Carvalho Júnior, O.A.; Matricardi, E.A.; Guimarães, R.F.; Gomes, R.A. Predicting wildfire vulnerability using logistic regression and artificial neural networks: A case study in Brazil. *Int. J. Wildland Fire* **2018**, *28*, 35–45. [CrossRef]

23. Bergado, J.R.; Persello, C.; Reinke, K.; Stein, A. Predicting wildfire burns from big geodata using deep learning. *Saf. Sci.* **2021**, *140*, 105276. [CrossRef]

24. Analysis, R. Evaluating Lightning-Caused Fire Occurrence Using Spatial Generalized Additive Models: A Case Study in Central Spain. *Risk Anal.* **2020**, *40*, 1418–1437.

25. Yolanda, S.S.; Antonio, M.-G.A.; Fernando, S.F.; Marina, M.P. Mapping Wildfire Ignition Probability Using Sentinel 2 and LiDAR (Jerte Valley, Cáceres, Spain). *Sensors* **2018**, *18*, 826. [CrossRef]

26. Couto, F.T.; Iakunin, M.; Salgado, R.; Pinto, P.; Viegas, T.; Pinty, J.P. Lightning modelling for the research of forest fire ignition in Portugal. *Atmos. Res.* **2020**, *242*, 104993. [CrossRef]

27. Price, C.; Rind, D. The effect of global warming on lightning frequencies. In Proceedings of the Conference on Severe Local Storms, Provincial Park, AB, Canada, 22–26 October 1990.

28. Coughlan, R.; Di Giuseppe, F.; Vitolo, C.; Barnard, C.; Lopez, P.; Drusch, M. Using machine learning to predict fire-ignition occurrences from lightning forecasts. *Meteorol. Appl.* **2021**, *28*, e1973. [CrossRef]

29. Moris, J.V.; Conedera, M.; Nisi, L.; Bernardi, M.; Cesti, G.; Pezzatti, G.B. Lightning-caused fires in the Alps: Identifying the igniting strokes. *Agric. For. Meteorol.* **2020**, *290*, 107990. [CrossRef]

30. Yang, J.; Huang, X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data* **2021**, *13*, 3907–3925. [CrossRef]

31. Larjavaara, M.; Kuuluvainen, T.; Rita, H.; Venlinen, A. Spatial distribution of lightning-ignited forest fires in Finland. In Proceedings of the International Conference on Forest Fire Research, Coimbra, Portugal, 18–23 November 2002.

32. Menezes, L.S.; de Oliveira, A.M.; Santos, F.L.M.; Russo, A.; de Souza, R.A.F.; Roque, F.O.; Libonati, R. Lightning patterns in the Pantanal: Untangling natural and anthropogenic-induced wildfires. *Sci. Total Environ.* **2022**, *820*, 153021. [CrossRef]

33. Pérez-Invernón, F.J.; Huntrieser, H.; Soler, S.; Gordillo-Vázquez, F.J.; Koutsias, N. *Lightning-Ignited Wildfires and Long-Continuing-Current Lightning in the Mediterranean Basin: Preferential Meteorological Conditions*; Copernicus GmbH: Göttingen, Germany, 2021.

34. Vasiliev, M.S.; Boroyev, R.N. Remote method determining the formation of forest fires from thunderstorms (on the example of Yakutia). In Proceedings of the International Symposium on Atmospheric and Ocean Optics, Atmospheric Physics, Moscow, Russia, 6–10 July 2020.

35. Kharyutkina, E.; Pustovalov, K.; Moraru, E.; Nechepurenko, O. Analysis of Spatio-Temporal Variability of Lightning Activity and Wildfires in Western Siberia during 2016–2021. *Atmosphere* **2022**, *13*, 669. [CrossRef]

36. Vitolo, C.; Di Giuseppe, F.; Barnard, C.; Coughlan, R.; San-Miguel-Ayanz, J.; Liberta, G.; Krzeminski, B. ERA5-based global meteorological wildfire danger maps. *Sci. Data* **2020**, *7*, 216. [CrossRef]

37. Pu, J.H.; Shen, C.; Zhang, C.X.; Liu, X.J. A semi-experimental method for evaluating frosting performance of air source heat pumps. *Renew. Energy* **2021**, *173*, 913–925. [CrossRef]

38. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. Aggregating cloud-free Sentinel-2 images with Google Earth Engine. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *4*, 145–152. [CrossRef]

39. Meraner, A.; Ebel, P.; Zhu, X.X.; Schmitt, M. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS J. Photogramm. Remote. Sens.* **2020**, *166*, 333–346. [CrossRef]

40. An, H.; Gan, J.B.; Cho, S.J. Assessing Climate Change Impacts on Wildfire Risk in the United States. *Forests* **2015**, *6*, 3197–3211. [CrossRef]

41. Shu, Y.; Shi, C.M.; Yi, B.L.; Zhao, P.W.; Guan, L.J.; Zhou, M. Influence of Climatic Factors on Lightning Fires in the Primeval Forest Region of the Northern Daxing'an Mountains, China. *Sustainability* **2022**, *14*, 5462. [CrossRef]

42. Didan, K. Vegetation Phenology and Enhanced Vegetation Index Products from Multiple Long Term Satellite Data Records. In *Vegetation Indices from Remote Sensing Are by Far the Most Widely Used Remote Sensing Tools for Studying Vegetation and Large-Scale Ecosystem Proc*; Federal Reporter; American Geophysical Union: Washington, DC, USA, 2008.

43. Nguyen, C.T.; Chidthaisong, A.; Diem, P.K.; Huo, L.Z. A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. *Land* **2021**, *10*, 231. [CrossRef]

44. Hashimoto, A.; Segah, H.; Yulianti, N.; Naruse, N.; Takahashi, Y. A new indicator of forest fire risk for Indonesia based on peat soil reflectance spectra measurements. *Int. J. Remote Sens.* **2021**, *42*, 1917–1927. [CrossRef]

45. Malak, D.A.; Pausas, J.G.; Pardo-Pascual, J.E.; Ruiz, L.A. Fire Recurrence and the Dynamics of the Enhanced Vegetation Index in a Mediterranean Ecosystem. *Int. J. Appl. Geospat. Res.* **2015**, *6*, 18–35. [CrossRef]

46. Zarco-Tejada, P.J.; Berjón, A.; López-Lozano, R.; Miller, J.R.; Martín, P.; Cachorro, V.; González, M.R.; de Frutos, A. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens. Environ.* **2005**, *99*, 271–287. [CrossRef]

47. Caccamo, G.; Chisholm, L.A.; Bradstock, R.A.; Puotinen, M.L.; Pippen, B.G. Monitoring live fuel moisture content of heathland, shrubland and sclerophyll forest in south-eastern Australia using MODIS data. *Int. J. Wildland Fire* **2012**, *21*, 257–269. [CrossRef]

48. Masoudvaziri, N.; Ganguly, P.; Mukherjee, S.; Sun, K. Impact of geophysical and anthropogenic factors on wildfire size: A spatiotemporal data-driven risk assessment approach using statistical learning. *Stoch. Environ. Res. Risk Assess.* **2021**, *36*, 1103–1129. [CrossRef]

49. Huang, Y.X.; Wu, S.L.; Kaplan, J.O. Sensitivity of global wildfire occurrences to various factors in the context of global change. *Atmos. Environ.* **2015**, *121*, 86–92. [CrossRef]

50. Liu, W.L.; Wang, S.X.; Zhou, Y.; Wang, L.T.; Zhu, J.F.; Wang, F.T. Lightning-caused forest fire risk rating assessment based on case-based reasoning: A case study in DaXingAn Mountains of China. *Nat. Hazards* **2016**, *81*, 347–363. [CrossRef]

51. Jou, Y.-J.; Huang, C.-C.L.; Cho, H.-J. A VIF-based optimization model to alleviate collinearity problems in multiple linear regression. *Comput. Stat.* **2014**, *29*, 1515–1541. [CrossRef]

52. Song, X.R.; Gao, S.; Liu, X.; Chen, C.B. An outdoor fire recognition algorithm for small unbalanced samples. *Alex. Eng. J.* **2021**, *60*, 2801–2809. [CrossRef]

53. Zhou, Z.; Zheng, C.E.; Liu, X.D.; Tian, Y.; Chen, X.Y.; Chen, X.X.; Dong, Z.X. A Dynamic Effective Class Balanced Approach for Remote Sensing Imagery Semantic Segmentation of Imbalanced Data. *Remote Sens.* **2023**, *15*, 1768. [CrossRef]

54. Oh, H.Y.; Jeong, M.H. Grid-based Urban Fire Prediction Using Extreme Gradient Boosting (XGBoost). *Sens. Mater.* **2022**, *34*, 4879–4890. [CrossRef]

55. Han, H.; Wang, W.Y.; Mao, B.H. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In *Advances in Intelligent Computing, Pt 1, Proceedings of the International Conference on Intelligent Computing, ICIC 2005, Hefei, China, 23–26 August 2005*; Huang, D.S., Zhang, X.P., Huang, G.B., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3644, pp. 878–887.

56. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
57. Penman, T.D.; Bradstock, R.A.; Price, O. Modelling the determinants of ignition in the Sydney Basin, Australia: Implications for future management. *Int. J. Wildland Fire* **2013**, *22*, 469–478. [CrossRef]