forests

MDPI

Article

Fuel Load Models for Different Tree Vegetation Types in Sichuan Province Based on Machine Learning

Hongrong Wang [1], Haoquan Chen [2], Hanmin Sheng [1,*], Kai Chen [1], Chen Dong [3] and Zhiqiang Min [3]

[1] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; rongcloud@163.com (H.W.); kaichen@uestc.edu.cn (K.C.)
[2] College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China; 15075185961@163.com
[3] State Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Zhejiang A&F University, Hangzhou 311300, China; dongchen@zafu.edu.cn (C.D.); minzq1026@163.com (Z.M.)
* Correspondence: hmsheng@uestc.edu.cn

Abstract: (1) Objective: To improve forest fire prevention, this study provides a reference for forest fire risk assessment in Sichuan Province. (2) Methods: This research focuses on various forest vegetation types in Sichuan Province. Given data from 6848 sample plots, five machine learning models—random forest, extreme gradient boosting (XGBoost), k-nearest neighbors, support vector machine, and stacking ensemble (Stacking)—were employed. Bayesian optimization was utilized for hyperparameter tuning, resulting in machine learning models for predicting forest fuel loads (FLs) across five different vegetation types. (3) Results: The FL model incorporates not only vegetation characteristics but also site conditions and climate data. Feature importance analysis indicated that structural factors (e.g., canopy closure, diameter at breast height, and tree height) dominated in cold broadleaf, subtropical broadleaf, and subtropical mixed forests, while climate factors (e.g., mean annual temperature and temperature seasonality) were more influential in cold coniferous and subtropical coniferous forests. Machine learning-based FL models outperform the multiple stepwise regression model in both fitting ability and prediction accuracy. The XGBoost model performed best for cold coniferous, cold broadleaf, subtropical broadleaf, and subtropical mixed forests, with coefficient of determination (R^2) values of 0.79, 0.85, 0.81, and 0.83, respectively. The Stacking model excelled in subtropical coniferous forests, achieving an R^2 value of 0.82. (4) Conclusions: This study establishes a theoretical foundation for predicting forest fuel capacity in Sichuan Province. It is recommended that the XGBoost model be applied to predict fuel loads (FLs) in cold coniferous forests, cold broadleaf forests, subtropical broadleaf forests, and subtropical mixed forests, while the Stacking model is suggested for predicting FLs in subtropical coniferous forests. Furthermore, this research offers theoretical support for forest fuel management, forest fire risk assessment, and forest fire prevention and control in Sichuan Province.

Keywords: vegetation type; forest fuel load; meteorological factors; stand factors; site factors; machine learning; feature importance analysis

check for updates

Received: 10 December 2024
Revised: 21 December 2024
Accepted: 27 December 2024
Published: 29 December 2024

Citation: Wang, H.; Chen, H.; Sheng, H.; Chen, K.; Dong, C.; Min, Z. Fuel Load Models for Different Tree Vegetation Types in Sichuan Province Based on Machine Learning. *Forests* **2025**, *16*, 42. https://doi.org/10.3390/f16010042

Copyright: © 2024 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Forest fires are a natural disturbance within ecosystems, playing a crucial role in their formation, development, and succession [1]. However, with global climate change and urban expansion [2], the frequency, area, and intensity of forest fires have shown a significant upward trend [3], posing serious threats to forest ecosystems and human society.

The primary causes of forest fires include lightning, human activities, climate change, and drought conditions [4]. For example, on 30 March 2019, a forest fire in Muli County, Sichuan Province, China, burned approximately 20 ha and resulted in 31 fatalities. On 30 March 2020, another forest fire in Xichang County, Sichuan Province, affected around 3000 ha and caused 19 fatalities [5]. In July 2019, Australia experienced a forest fire that lasted half a year [6]. In August of the same year, the Amazon rainforest, known as the "lungs of the Earth", burned for more than 2 weeks, destroying over 800,000 ha of forest in South America [7]. Additionally, unprecedented high-intensity fires occurred in Eastern Siberia and Alaska in 2019 and 2020, and 2023 marked the most severe year for forest fires in Canada's history [8]. Forest fires release massive amounts of energy and greenhouse gases in a short period, severely disrupting the natural environment. This phenomenon poses significant threats to human life, property, and the ecological environment [9].

Countries frequently affected by forest fires, such as China, the United States, Canada, Spain, and Australia, have been conducting research on forest fire risk assessment and early warning systems. As early as 1996, Pyne et al. [10] introduced the well-known fire triangle model, identifying climate, topography, and fuel as the three primary factors in fire risk assessment. Among these factors, climate and topography data are relatively easy to obtain, making accurate fuel load (FL) estimation essential for effective forest fire risk assessment.

Fuel load (FL) refers to the dry mass of combustible materials per unit area, encompassing all live and dead organic matter [11]. It is a critical parameter in fire behavior modeling, influencing fire intensity [12,13], rate of spread [14], flame length [15,16], fuel consumption [17], and fire severity [18]. FL also serves as an indispensable input variable for fire behavior models [19,20] and helps fire management personnel design proactive measures, such as prescribed burning and increasing fuel spacing, to reduce the risk of fire outbreaks. Importantly, FL specifically includes the combustible portions of organic matter, such as fine branches, leaves, and litter, which are likely to ignite during wildfires. In contrast, biomass represents the total mass of living organisms in a given area, encompassing both combustible (e.g., leaves, fine branches) and non-combustible components (e.g., trunks, coarse roots). Confusing these concepts can lead to inaccuracies in fire behavior predictions, underscoring the need for clear distinctions in fire-related studies.

FL is critical for early warning systems for forest fires [21,22]. Obtaining accurate FL data traditionally requires establishing sample plots and collecting data through extensive field and laboratory collaboration, which demands substantial human and material resources [23]. For expediting forest FL estimation, sample plot data can be used to build and apply forest FL models on a large scale. Scholars domestically and internationally have made progress in this area [24,25]. Initially, research focused on developing univariate regression models between FL and individual stand growth factors, such as diameter at breast height (DBH) and tree height. For instance, Sheng et al. [26] developed a regression model based on measured sample data from a forest fire risk survey in Shanghai, establishing the relationship between FL in the tree layer and stand volume. As single-variable models proved insufficient, multivariate regression models were introduced, incorporating not only stand growth factors but also environmental and site factors. Wang et al. [27] developed a multivariate linear stepwise regression model to explore the main factors influencing surface FL in Yunnan pine forests in Xichang. Similarly, Xiao et al. [28] used a joint estimation method to create a model for FL in Chinese red pine forests, and Chojnacky et al. [29] used generalized allometric biomass equations to develop a total aboveground biomass model for North American tree species. Nguyen et al. [30] developed a refined FL model using stepwise regression, while Sydorenko et al. [31] constructed a model with multiple regression to examine the relationship between surface FL and fixed-effect variables. Research to date has primarily focused on stand growth and environmental factors

influencing FL. As climate change gains attention globally, integrating climate factors with stand and site variables into FL models is expected to become a key research direction in the future.

Most current FL models rely on traditional linear or simple nonlinear modeling approaches, which require certain statistical assumptions, such as data independence, normal distribution, and homoscedasticity. However, the relationships between FL and factors such as stand characteristics and site conditions are often complex and nonlinear, making it difficult for many datasets to meet these assumptions and limiting the accuracy of these models. To address this issue, some researchers have proposed using machine learning algorithms to develop FL models. Unlike traditional statistical models, machine learning methods do not require assumptions about data distribution and can effectively handle high-dimensional data with complex nonlinear interactions, allowing them to extract valuable insights and improve predictive accuracy. For instance, Li et al. [32] utilized a backpropagation (BP) neural network to develop a dynamic model integrating total forest fuel load (FL) and stand factors. However, BP neural networks have limitations, including long training times and a trade-off between predictive and generalization abilities, making them prone to overfitting [33]. In recent years, various improved and classic machine learning algorithms have emerged, but comparative studies using multiple algorithms to construct and evaluate FL models remain rare. Additionally, existing research has primarily focused on small regions with homogeneous tree species, making these models applicable only to specific tree species or limited areas. Studies on large-scale FL models, such as those covering provincial regions and diverse forest types, are still scarce.

Sichuan Province possesses abundant forest resources but is also highly susceptible to forest fires [34,35]. This study utilizes data from 6848 standard tree plots collected during the Forest Fire Risk Survey of Sichuan Province, classifying forests into five vegetation types based on vegetation regions and classifications. The primary objective is to develop a high-precision fuel load model by systematically analyzing the relationships between fuel load and stand factors, site conditions, and climate variables. Various machine learning models—such as Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)—are applied, and their performances are compared with traditional regression models. To further enhance prediction accuracy and applicability, a Stacking ensemble learning model is implemented, leveraging the strengths of multiple individual models. Unlike existing fuel load models, this study integrates stand, site, and meteorological factors for the first time while exploring the application of different machine learning algorithms to fuel load modeling. The most accurate estimation model for each forest type is identified, and key factors influencing fuel load are quantified. These findings provide a theoretical foundation and technical reference for forest fuel management, fire risk assessment, and fire prevention in Sichuan Province.

2. Materials and Methods

2.1. Overview of the Study Area

Sichuan Province, located in Southwestern China, spans geographic coordinates from 97°21′ E to 108°33′ E and 26°03′ N to 34°19′ N. This region is expansive and predominantly characterized by mountainous, hilly, and basin terrains. The climate is primarily subtropical monsoon [36], with variations including humid subtropical and plateau monsoon climates. The average annual temperature ranges from 4.5 °C to 19.2 °C, while annual precipitation varies between 400 and 1600 mm. These climatic conditions, combined with diverse soil types—including purple soil, yellow soil, red soil, mountain yellow–brown soil, and alpine desert soil—have created a highly diverse ecological environment. Based on the topography and climate, Sichuan's vegetation is classified into two major regions: the

subtropical evergreen broadleaf forest zone and the alpine vegetation zone of the Tibetan Plateau. This study uses data from 6848 standard tree plots collected during the Forest Fire Risk Survey of Sichuan Province. Each plot is square-shaped, with a side length of 25.82 m and an area of 0.067 ha. Surveyed factors include topographic variables (altitude, slope, aspect, and slope position) and vegetation growth variables (average tree height in meters, average DBH in centimeters, canopy density, stand density in trees per hectare, and age group), totaling nine indicators. The plot distribution covers the province's main forested areas, as illustrated in Figure 1.

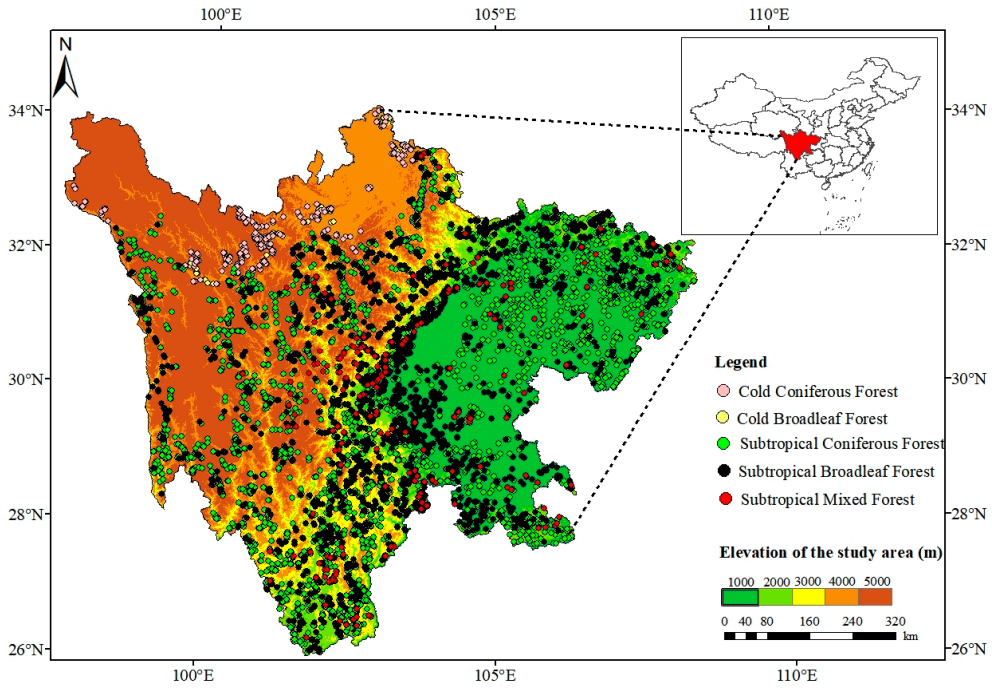


Figure 1. Distribution of Sample Plots in the Study Area.

2.2. Tree Type Classification

Different vegetation regions and types represent diverse ecological environments and climatic conditions. Classifying forest stands based on these factors enables more accurate forest fire risk assessments and more effective forest fire prevention management. Drawing on the Implementation Guidelines for the First Forest and Grassland Fire Risk Survey of Sichuan Province: Forest Fuel Sample Plots, Grassland Fuel Sample Plots, and Large Plot Surveys [37] and incorporating Sichuan's ecological environment, topography, and climate, this study categorizes forest stands into five types: cold coniferous forests, cold broadleaf forests, subtropical coniferous forests, subtropical broadleaf forests, and subtropical mixed forests. Research further indicates that vegetation types are strongly associated with climatic and ecological characteristics, with significant differences in fuel loads, burning behaviors, and fire occurrence mechanisms across forest types. This classification approach highlights variations in fire risk among vegetation types [38,39], allowing for more precise identification of high-risk areas and offering a scientific foundation for forest fire prevention and management. Table 1 presents the number of sample plots and major tree species for each vegetation type.

Table 1. Information on Sample Plots for Different Tree Vegetation Types.

Vegetation Zone	Vegetation Type	Number of Plots	Main Dominant Tree Species
Alpine Vegetation Zone of the Qinghai–Tibet Plateau	cold coniferous forest	308	*Picea asperata*, *Cupressus funebris*, *Abies fabri*
	cold broadleaf forest	59	*Betula* spp.
Subtropical Evergreen Broadleaf Forest Zone	subtropical coniferous forest	4043	*Picea asperata Mast.*, *A. fabri*, *C. funebris*
	subtropical broadleaf forest	2231	*Quercus* spp., *Betula* spp., species of miscellaneous hardwoods and softwoods
	subtropical mixed forest	207	*Quercus* spp., *Pinus yunnanensis*, species of miscellaneous hardwoods and softwoods

2.3. Plot Setup and FL Measurement

This study focuses on the aboveground fuel load within the sample plots. In forest stands, the total aboveground fuel load is the sum of the fuel loads from the tree layer, shrub layer, herbaceous layer, litter layer, and humus layer. It is expressed as follows:

Aboveground fuel load = Tree layer fuel load + Shrub layer fuel load + Herbaceous layer fuel load + Litter layer fuel load

The litter layer is further categorized into three classes: Litter 1 (small twigs, leaves, and weeds with a diameter less than 0.6 cm), Litter 2 (twigs with a diameter $\geq$0.6 cm but <2.5 cm), and Litter 3 (branches with a diameter $\geq$2.5 cm but <7.62 cm). In this study, the tree layer fuel load was calculated based on individual tree volume measurements recorded during field surveys [40], using biomass expansion equations for different tree species as specified in the Implementation Guidelines for the First Forest and Grassland Fire Risk Survey of Sichuan Province: Forest Fuel Sample Plots, Grassland Fuel Sample Plots, and Large Plot Surveys [37]. The standing stock volume of each tree within the standard plots was estimated to calculate the tree layer fuel load. Meanwhile, the fuel loads of the shrub layer, herbaceous layer, litter layer, and humus layer were determined by measuring the dry-to-fresh weight ratio of samples collected from the plots.

Within the subplot setup of the 6848 selected standard plots, the five-point sampling method [41,42] was applied. Four subplots for measuring fuel loads of the shrub layer, herbaceous layer, litter, and humus were placed 3 m from two adjacent edges at each corner of the standard plot. Additionally, one subplot for the same measurements was established at the center of each plot. The layout of the plots is shown in Figure 2.

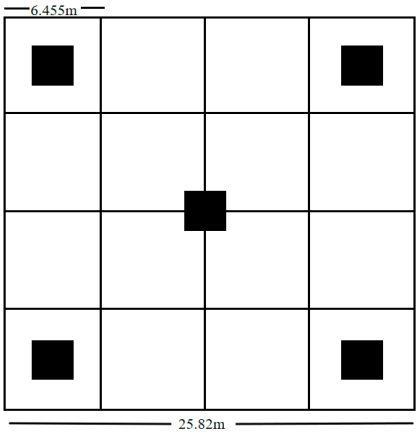


Figure 2. Sample Plot Layout.

Sampling was performed using the harvesting method [43], where five subplots within each standard plot were thoroughly mixed. Fresh samples of 500 g, 300 g, 200 g, and 200 g were collected for shrubs, herbs, litter, and humus, respectively, and their weights were recorded. The litter was further divided into three categories: Litter 1; Litter 2; and Litter 3. The fresh samples were transported to the laboratory, oven-dried, and reweighed. The fresh-to-dry weight ratio was then determined using the fresh and dry weights of the subplots [44], as described in Equation (1).

$$M = \frac{m_B - m_C}{m_C} \times 100\% \quad (1)$$

In this formula, M represents the absolute moisture content of the fuel; m_B is the fresh mass of the fuel (measured in the field during sampling), and m_C is the dry mass of the fuel (measured after drying). The total dry weight for each subplot and plot can be calculated on the basis of the dry-to-fresh weight ratio. Adding the tree layer FL and performing unit conversions yield the FL per unit area for each plot (t/hm^2). The FL data and basic information for different vegetation types are summarized in Table 2. The classification standards for age class, slope, aspect, and slope position are derived from the Implementation Guidelines for the First Forest and Grassland Fire Risk Survey of Sichuan Province: Forest Fuel Sample Plots, Grassland Fuel Sample Plots, and Large Plot Surveys [37].

Table 2. Combustible load and basic information of forest stands in sample plots.

Plot Factors	Cold Coniferous Forest	Cold Broadleaf Forest	Subtropical Coniferous Forest	Subtropical Broadleaf Forest	Subtropical Mixed Forest
FL per Unit Area/($t \cdot hm^{-2}$)	118.27 ± 105.71	129.09 ± 78.59	81.89 ± 41.81	96.16 ± 6.22	115.44 ± 66.6
Average Tree Height/m	12.49 ± 7.69	8.84 ± 2.82	12.56 ± 5.72	12.34 ± 4.73	12.96 ± 4.76
Average DBH/m	0.23 ± 0.14	0.13 ± 0.05	0.19 ± 0.12	0.15 ± 0.07	0.18 ± 0.08
Canopy Closure	0.51 ± 0.19	0.57 ± 0.19	0.6 ± 0.19	0.61 ± 0.16	0.6 ± 0.16
Stand Density (plants·hm^{-2})	731.35 ± 485.06	1370.93 ± 866.71	1432.58 ± 963.5	1457.15 ± 841.54	1494.52 ± 793.46
Elevation	3588.26 ± 319.77	3298.63 ± 278.81	2250.54 ± 1210.58	1672 ± 1199.29	1699.47 ± 961.71
Age Class	Young Forest, Middle-Aged Forest, Nearly Mature Forest, Mature Forest, Overmature Forest				
Slope	Flat Slope, Gentle Slope, Moderate Slope, Steep Slope, Sharp Slope, Hazardous Slope				
Aspect	North Slope, Northeast Slope, East Slope, Southeast Slope, South Slope, Southwest Slope, West Slope, Northwest Slope, No Aspect				
Slope Position	Ridge, Upper Slope Position, Middle Slope Position, Lower Slope Position, Valley, Flatland				

2.4. Meteorological Factor Extraction

Vegetation growth is influenced by climate, which also indirectly contributes to forest fires [45]. To improve the accuracy of fuel load predictions, this study incorporates climatic factors into the analysis. The climate data used in this study were obtained from the

WorldClim database (http://www.worldclim.org/, accessed on 23 September 2024) [46], which provides global climate layers at a 30′ spatial resolution. These layers were generated through spatial interpolation based on observational data from meteorological stations worldwide between 1950 and 2000. Using ArcGIS 10.8.1 software, estimated climate data for each sampling point were extracted. Finally, 19 climatic variables that significantly influence vegetation growth in the sample plots were selected, as shown in Table 3.

Table 3. Symbols for Meteorological Factors.

№	Climate Factor	Abbreviation
1	Annual Mean Temperature	MAT
2	Mean Diurnal Range	MDTR
3	Temperature Uniformity	TU
4	Temperature Seasonality	TS
5	Maximum Temperature of the Warmest Month	MWMT
6	Minimum Temperature of the Coldest Month	MCMT
7	Annual Temperature Range	ATR
8	Mean Temperature of the Wettest Season	MTWS
9	Mean Temperature of the Driest Season	MTDS
10	Mean Temperature of the Warmest Season	MTHS
11	Mean Temperature of the Coldest Season	MTCS
12	Annual Precipitation	AP
13	Precipitation of the Wettest Month	PWM
14	Precipitation of the Driest Month	PDM
15	Precipitation Seasonality	PS
16	Precipitation of the Wettest Season	PWSS
17	Precipitation of the Driest Season	PDSS
18	Precipitation of the Warmest Season	PHSS
19	Precipitation of the Coldest Season	PCSS

Note: Climatic data were obtained from the WorldClim database (http://www.worldclim.org/, accessed on 23 September 2024) and processed for use in this study.

2.5. Quantification of Discrete Data

In the data processing phase, discrete variables, including plot slope, aspect, slope position, and age class, were quantified. Following the quantification method proposed by Dahouda et al. [47], this study established the rules for converting these discrete variables into numerical values, as detailed in Table 4. This approach systematically transforms categorical data into numerical form, facilitating comprehensive analysis within the modeling framework.

Table 4. Data Transformation Rules for Slope, Aspect, Slope Position, and Age Class.

Slope		Aspect		Slope Position		Age Class	
Category	Code	Category	Code	Category	Code	Category	Code
Flat Slope	1	North Slope	1	Ridge	1	Young Forest	1
Gentle Slope	2	Northeast Slope	2	Upper Slope Position	2	Middle-Aged Forest	2
Moderate Slope	3	East Slope	3	Middle Slope Position	3	Nearly Mature Forest	3
Steep Slope	4	Southeast Slope	4	Lower Slope Position	4	Mature/Overmature Forest	4
Sharp Slope	5	South Slope	5	Valley	5		
Hazardous Slope	6	Southwest Slope	6	Flatland	6		
		West Slope	7				
		Northwest Slope	8				
		No Aspect	9				

To enhance the predictive accuracy of the model and mitigate overfitting, this study applied data cleaning procedures to forest stand, topographic, and climatic data. First, the three-standard deviation method was used to identify and remove outliers, ensuring the validity of the dataset. Next, the relationships between variables were analyzed using the Spearman correlation coefficient, while the variance inflation factor (VIF) was applied to filter independent variables. Variables with a Spearman correlation coefficient below 0.8 [48] and a VIF value under 10 [49] were considered free from significant multicollinearity. Afterward, the selected features were standardized to eliminate the effects of dimensional differences on model performance. Finally, the dataset was randomly split into a training set (70%) and a test set (30%) to facilitate model training and evaluate predictive accuracy.

2.6. Machine Learning Models

In this study, the total fuel load per unit area, calculated in Section 2.3, was used as the dependent variable, while the selected sample plot investigation factors and climatic factors served as independent variables. Four algorithms—Random Forest Regression (RF), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors Regression (KNN), and Support Vector Machine Regression (SVM)—were employed as base models to construct fuel load prediction models. On this basis, the Stacking ensemble learning method was utilized to combine these four base models into a strong regressor for fuel load prediction, which was then compared with traditional regression models for performance evaluation. Hyperparameter optimization, a critical step to improve model performance and reduce computational cost, was conducted using Bayesian optimization and five-fold cross-validation. All models were implemented using Python (version 3.9). An introduction to the various models is provided below.

2.6.1. Models and Theoretical Foundation

(1) Random forest (RF). RF is a powerful ensemble learning algorithm that enhances predictive accuracy by constructing multiple decision trees and combining their predictions. By randomly selecting features and samples during training, it reduces the risk of overfitting and exhibits strong resistance to noise, making it well suited for handling complex, high-dimensional data. The RF model consists of multiple independently trained decision trees. For prediction, the outputs of all decision trees are combined to achieve robust learning outcomes [50]. The prediction formula for RF is as follows:

$$\hat{y} = \frac{1}{k}\sum_{k=1} f_k(x) \quad (2)$$

where $\hat{y}$ represents the predicted value; k is the total number of decision trees, and $f_k(x)$ denotes the prediction result of k decision tree;

(2) Extreme gradient boosting (XGBoost). XGBoost is a decision tree-based algorithm that uses gradient boosting to optimize overall prediction accuracy by iteratively correcting errors from previous models. Known for its efficiency and strong nonlinear fitting capabilities, XGBoost is particularly suitable for analyzing the complex effects of climate change and stand factors on FL [51]. The model structure is represented by the following formula:

$$\hat{y} = \sum_{k-1}^{k} f_k(x_i) \quad (3)$$

where $\hat{y}$ is the predicted value; k is the number of trees; f_k is the prediction function of k tree, and x_i is i input sample.

$$Obj(\theta) = L(\theta) + \Omega(\theta) \quad (4)$$

Here, $L(\theta)$ represents the training loss function, which evaluates the model's fit to the training data; $\Omega(\theta)$ serves as the regularization term, aiming to control model complexity and prevent overfitting. The objective function of XGBoost is optimized by minimizing the training loss and the regularization term simultaneously, thereby balancing accuracy and generalization ability;

(3) Support vector machine (SVM). SVM is a supervised learning algorithm that aims to find the optimal hyperplane by maximizing the margin between classes in the feature space, making it particularly suitable for high-dimensional and small-sample data. By introducing kernel functions, SVM can effectively handle nonlinear data, enabling it

to capture the complex relationships between climate factors and vegetation types in this study. For regression tasks [52], the core formula of SVM is expressed through support vector regression as follows:

$$\hat{y} = \omega^T \phi(x) + b \tag{5}$$

In this model, $\hat{y}$ represents the predicted value; ω denotes the weight vector; $\phi(x)$ is the feature representation of input x mapped to a high-dimensional space via a kernel function, and b is the bias term. The loss function used is the ε-insensitive loss function, formulated as:

$$L(y, \hat{y}) = \max(0, |y - \hat{y}| - \varepsilon) \tag{6}$$

The loss function only calculates loss when the prediction error exceeds the threshold ε. By optimizing the weight ω and bias b, SVM can effectively perform regression prediction and capture complex variable relationships.

The k-nearest neighbors (KNN) algorithm is an instance-based learning method assuming that similar data points are close to one another in feature space [53]. It makes predictions by measuring the distance between a new data point and k nearest neighbors in the training set. The prediction process includes the following steps: (1) calculating the distance between the new data point and all data points in the training set (using, for example, Euclidean or Manhattan distances); (2) identifying k nearest neighbors; and (3) predicting the target value for the new data point by averaging or performing a weighted average of the target values of these neighbors. The specific distance metric formulas are as follows:

Euclidean distance:

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{ij})^2} \tag{7}$$

Manhattan distance:

$$d(x, x_i) = \sum_{j=1}^{n} |x_j - x_{ij}| \tag{8}$$

Here, $d(x,x_i)$ represents the distance between the new data point x and the data point x_i in the training set. n denotes the number of features, and x_j and x_{ij} are the values of data points x and x_i for the j-th feature, respectively;

(4) Stacking ensemble (Stacking). Stacking is an ensemble learning method that improves overall predictive performance by combining the predictions of multiple base models. In this study, RF, XGBoost, SVM, and KNN are used as base models, with RF serving as the metalearner to integrate the predictions from each model, thereby enhancing the accuracy of FL predictions (Figure 3).

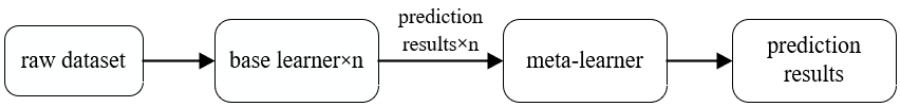


Figure 3. Stacking Learning Prediction Architecture Diagram.

2.6.2. Bayesian Optimization

Bayesian optimization is an optimization algorithm based on Bayes' theorem and Gaussian processes [54]. Its goal is to iteratively refine the estimation of the objective function through a prior distribution model to obtain the optimal solution. The core idea involves selecting the next parameter value to evaluate, on the basis of known points in the parameter space, to maximize information gain through an acquisition function. In each iteration, the posterior distribution of the objective function is updated using existing

experimental results and prior knowledge of the function, continuing until the optimal solution is found or a preset number of iterations is reached. This method improves model performance effectively with limited sample data, making it especially suitable for hyperparameter tuning [55]. The optimization process is given by the following equation:

$$x_{next} = \underset{x}{\operatorname{argmax}} \alpha(x) \tag{9}$$

Here, $\alpha(x)$ represents the acquisition function, which is used to evaluate the potential "value" or "importance" of the current point x. Bayesian optimization selects the next sampling point x_{next} by maximizing the acquisition function $\alpha(x)$.

This study employed four algorithms—RF regression, XGBoost, KNN regression, and SVM—as base models. Through the Stacking learning method, these four base models were combined into a strong regressor for FL prediction, which was then compared with traditional regression models. Hyperparameter optimization is a key step in improving model performance and reducing computational cost. In this study, Bayesian optimization and fivefold cross-validation were used to optimize the hyperparameters of each model. All models were implemented in Python.

2.7. Model Accuracy Validation

To evaluate the predictive performance of RF, XGBoost, KNN regression, SVM, Stacking algorithm, and stepwise regression, this study used the coefficient of determination (R^2) and mean absolute error (MAE) as measures of model fit. Additionally, mean relative error (MRE) and root-mean-square error (RMSE) were adopted to test the model's predictive accuracy. The R^2 value indicates the degree of fit between predicted and observed values, with values closer to 1 reflecting better model fit. Lower values of MAE, MRE, and RMSE indicate improved model performance.

$$R^2 = 1 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{n}(y_i - \overline{y})^2 \tag{10}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \times 100\% \tag{12}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / n} \tag{13}$$

In these formulas, y_i represents the actual observed value; $\hat{y}_i$ denotes the predicted value from the model; n is the number of sample units, and p is the number of parameters.

3. Results

3.1. Selection Results of Different Vegetation Features

After performing correlation analysis on the features using Spearman's correlation coefficient and further screening the independent variables through the Variance Inflation Factor (VIF), the feature selection results for different vegetation types were obtained, as shown in Table 5.

Table 5. Feature Selection Results for Different Vegetation Types.

Vegetation Type	Forest Stand Factors	Topographic Factors	Climatic Factors
Cold Coniferous Forest	Canopy Closure, Stand Density/(plants·hm^{-2}), Average Tree Height/m	Slope, Aspect, Slope Position	MAT, TS, ATR, PWM, PDM, PS, PDSS
Cold Broadleaf Forest	Canopy Closure, Stand Density/(plants·hm^{-2}), Average Tree Height/m, Age Class	Slope, Aspect, Slope, Position, Altitude	MAT, TS, ATR, PWM, PDM, PS
Subtropical Coniferous Forest	Canopy Closure, Stand Density/(plants·hm^{-2}), Average Tree Height/m, Average DBH/m, Age Class	Slope, Aspect, Slope Position	MAT, MDTR, ATR, PS
Subtropical Broadleaf Forest	Canopy Closure, Stand Density/(plants·hm^{-2}), Average Tree Height/m, Average DBH/m, Age Class	Slope, Aspect, Slope Position	MAT, MDTR, ATR, PS
Subtropical Mixed Forest	Canopy Closure, Stand Density/(plants·hm^{-2}), Average Tree Height/m, Average DBH/m, Age Class	Slope, Aspect, Slope Position	PS, MDTR, ATR, AP, PS

Table 5 shows that the selected features differ among vegetation types. However, three forest stand factors (canopy density, stand density, and average tree height), three topographic factors (slope, aspect, and slope position), and three climatic factors (annual mean temperature, annual temperature range, and precipitation seasonality) are shared as common features across all five vegetation types. This indicates that these factors exhibit significant universality and play a crucial role in characterizing different vegetation types.

3.2. Machine Learning Modeling

This study employed five machine learning models to predict the total fuel load per unit area across five distinct vegetation types. The models used were Random Forest Regression (RF), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors Regression (KNN), Support Vector Machine (SVM), and the Stacking ensemble method. The performance and evaluation results of these models are presented in Table 6.

Table 6. Accuracy Evaluation of Different Models.

Vegetation Types	Model Name	Fitting Metrics		Validation Metrics	
		R^2	*MRE*	*MAE*	*RMSE*
Cold Coniferous Forest	RF	0.61	0.66	64.13	76.65
	XGBoost	0.79	0.57	57.16	73.87
	KNN	0.37	0.99	64.34	82.95
	SVM	0.60	0.28	58.54	79.30
	Stacking	0.70	0.68	63.62	84.38
Cold Broadleaf Forest	RF	0.61	0.66	64.20	76.67
	XGBoost	0.85	0.46	58.15	67.64
	KNN	0.64	0.65	50.75	67.73
	SVM	0.66	0.40	51.79	61.96
	Stacking	0.74	0.56	67.70	82.91
Subtropical Coniferous Forest	RF	0.78	0.24	35.47	50.84
	XGBoost	0.73	0.29	35.78	51.28
	KNN	0.34	0.43	37.89	53.90
	SVM	0.36	0.38	36.94	55.54
	Stacking	0.82	0.21	35.88	51.42
Subtropical Broadleaf Forest	RF	0.73	0.32	38.53	56.30
	XGBoost	0.81	0.30	38.31	55.43
	KNN	0.39	0.52	44.12	60.61
	SVM	0.35	0.49	38.62	56.41
	Stacking	0.72	0.25	40.08	58.44
Subtropical Mixed Forest	RF	0.73	0.26	39.85	52.45
	XGBoost	0.83	0.20	43.39	56.17
	KNN	0.38	0.41	43.98	56.52
	SVM	0.76	31.11	43.17	55.72
	Stacking	0.67	0.46	41.06	55.59

In cold coniferous forests, the XGBoost model performed best, with an R^2 value of 0.79, *MRE* of 0.57, *MAE* of 57.16, and *RMSE* of 73.87. Similarly, in cold broadleaf forests, XGBoost showed excellent performance, achieving an R^2 value of 0.85, *MRE* of 0.46, *MAE* of 58.15, and *RMSE* of 67.64. These results indicated that XGBoost effectively captured the underlying structure of FL data across various vegetation types, providing accurate predictions. In subtropical coniferous forests, the Stacking model demonstrated superior performance, with an R^2 value of 0.82, *MRE* of 0.21, *MAE* of 35.47, and *RMSE* of 50.84, highlighting the advantages of ensemble methods in handling complex data. For subtropical broadleaf and mixed forests, XGBoost continued to perform well, with R^2 values of 0.81 and 0.82, *MRE* values of 0.30 and 0.19, *MAE* values of 38.31 and 42.76, and *RMSE* values of 55.43 and 54.90, respectively. These findings confirmed XGBoost's robustness and predictive capability across various vegetation types.

In sum, the performance of different models varied significantly across vegetation types. The XGBoost model consistently demonstrated superior predictive accuracy across all five forest types, making it particularly well-suited for FL prediction in these environments. Meanwhile, the Stacking model exhibited strong comprehensive performance in subtropical coniferous forests, underscoring the importance of considering vegetation-specific characteristics when selecting an appropriate model.

3.3. Model Comparison

To further compare the performance of different models in predicting FL, this study employed a multiple linear stepwise regression method to construct FL models. Linear regression equations were developed for each of the five vegetation types, with the evaluation metrics for these equations presented in Table 7.

Table 7. FL Models for Different Vegetation Types Based on Stepwise Linear Regression.

Vegetation Types	Prediction Model	Fitting Metrics		Validation Metrics	
		R^2	*MRE*	*MAE*	*RMSE*
Cold Coniferous Forest	$Y = -37.26 \times PS + 24.07 \times X8 - 22.07 \times ATR + 22.94 \times PWSS - 20.18 \times PDSS - 14.18 \times X4 + 138.06$	0.31	1.13	67.77	84.10
Cold Coniferous Forest	$Y = -97.52 \times TU + 45.89 \times PWSS + 22.73 \times X6 + 126.61$	0.61	0.67	54.78	69.94
Cold Broadleaf Forest	$Y = 3.90 \times X7 - 5.25 \times X4 + 6.40 \times X5 + 79.01 \times ATR - 160.03 \times MDTR - 39.27 \times TS + 44.94 \times PS + 89.21 \times TU + 4.03 \times X2 + 8.41 \times X8 - 45.05 \times MTWS + 17.03 \times PDM - 60.32 \times PHSS + 50.76 \times AP + 3.40 \times X6$	0.23	0.50	41.28	55.75
Subtropical Coniferous Forest	$Y = 21.53 \times X8 + 9.20 \times X6 - 55.10 \times MAT + 8.45 \times X5 + 42.94 \times MCMT + 7.39 \times X9 + 4.67 \times X2$	0.28	0.61	44.20	60.88
Subtropical Broadleaf Forest	$Y = -23.17 \times MTDS + 18.10 \times X7 + 123.67$	0.20	0.52	41.77	54.02

Note: X2—Slope; X4—Slope Position; X5—Canopy Density; X6—Stand Density; X7—Average Tree Height; X8—Average DBH; X9—Age Group.

The comparative analysis of multiple stepwise regression and machine learning models across different vegetation types clearly demonstrated significant differences in predictive performance. Overall, machine learning models outperformed multiple stepwise regression in fitting capability and predictive accuracy. In cold coniferous forests, for instance, the XGBoost model achieved a high R^2 value of 0.79, which was significantly better than the value of 0.31 obtained by stepwise regression. Additionally, XGBoost showed superior performance in testing metrics, with *MAE* and *RMSE* values of 57.16 and 73.87, respectively, compared with the values of 67.77 and 84.10 for stepwise regression, respectively. In cold broadleaf and subtropical coniferous forests, XGBoost and Stacking models exhibited better fitting and testing metrics, confirming their advantages in handling complex nonlinear relationships. For subtropical coniferous and broadleaf forests, machine

learning models not only effectively capture complex interactions among multiple features but also significantly improve prediction stability.

In summary, the XGBoost and Stacking models demonstrated superior predictive ability across various vegetation types, especially when addressing complex nonlinear relationships. By contrast, the multiple stepwise regression model performed modestly across several vegetation types, with lower fitting and testing metrics than machine learning models, highlighting the limitations of traditional regression methods in environmental data analysis. Therefore, for practical applications in FL prediction, machine learning models are recommended to enhance predictive accuracy and stability.

3.4. Analysis of Standardized Residuals and Prediction vs. Actual Values for the Optimal Model

Based on the comparison of the models, it can be concluded that machine learning models demonstrate superior performance in fuel load prediction, particularly the XGBoost and Stacking ensemble models. To further validate the performance of the models, this study analyzes the residuals and the relationship between predicted and actual values for the optimal fuel load models under different vegetation types. Residual plots (Figure 4) and scatter plots with regression lines for predicted versus actual values are provided to visually illustrate the distribution of prediction errors.

In Figure 4, the horizontal axis represents the predicted fuel load values, while the vertical axis shows the standardized residuals for each model. According to the residual plot in Figure 4, most residual values fall within the range of ± 2, with only a small portion exceeding this range. This indicates that the prediction results of the models are accurate in most cases, with minimal errors and strong stability. These findings further demonstrate the superiority of machine learning models in handling the complex relationships between vegetation characteristics and fuel load. Machine learning models effectively adapt to the prediction needs of different vegetation types while significantly reducing prediction bias.

Figure 5 illustrates scatter plots and regression lines depicting the relationship between predicted and actual values generated by the optimal machine learning models. The results demonstrate satisfactory predictive performance across various vegetation types, with coefficients of determination (R^2) ranging from 0.84 to 0.93. Among these, the cold broadleaf forest achieved the best fit (R^2 = 0.93), characterized by tightly clustered scatter points and a regression line closely following the ideal fit (blue dashed line). This indicates that the XGBoost model can precisely predict fuel loads for this vegetation type. In contrast, the subtropical broadleaf forest exhibited a slightly lower R^2 value (R^2 = 0.84). Nonetheless, the predictions effectively captured the trend in fuel load variation, showcasing the model's adaptability to complex vegetation types. For the cold coniferous forest and subtropical coniferous forest, the model performed well in predicting low to moderate fuel load values but tended to underestimate in high-value regions. Prediction biases across vegetation types may stem from uneven sample distributions or the complexity of feature variables. For instance, in the cold coniferous forest and subtropical mixed forest, the model displayed greater biases when predicting low and high-value samples. This could reflect a lack of representativeness in feature variables or potential limitations in the model's complexity. Overall, the models effectively captured the trends in fuel load variation. However, opportunities for improvement remain. Future research could enhance model performance by incorporating more discriminative environmental variables or refining algorithm structures to improve generalization capability and predictive accuracy.

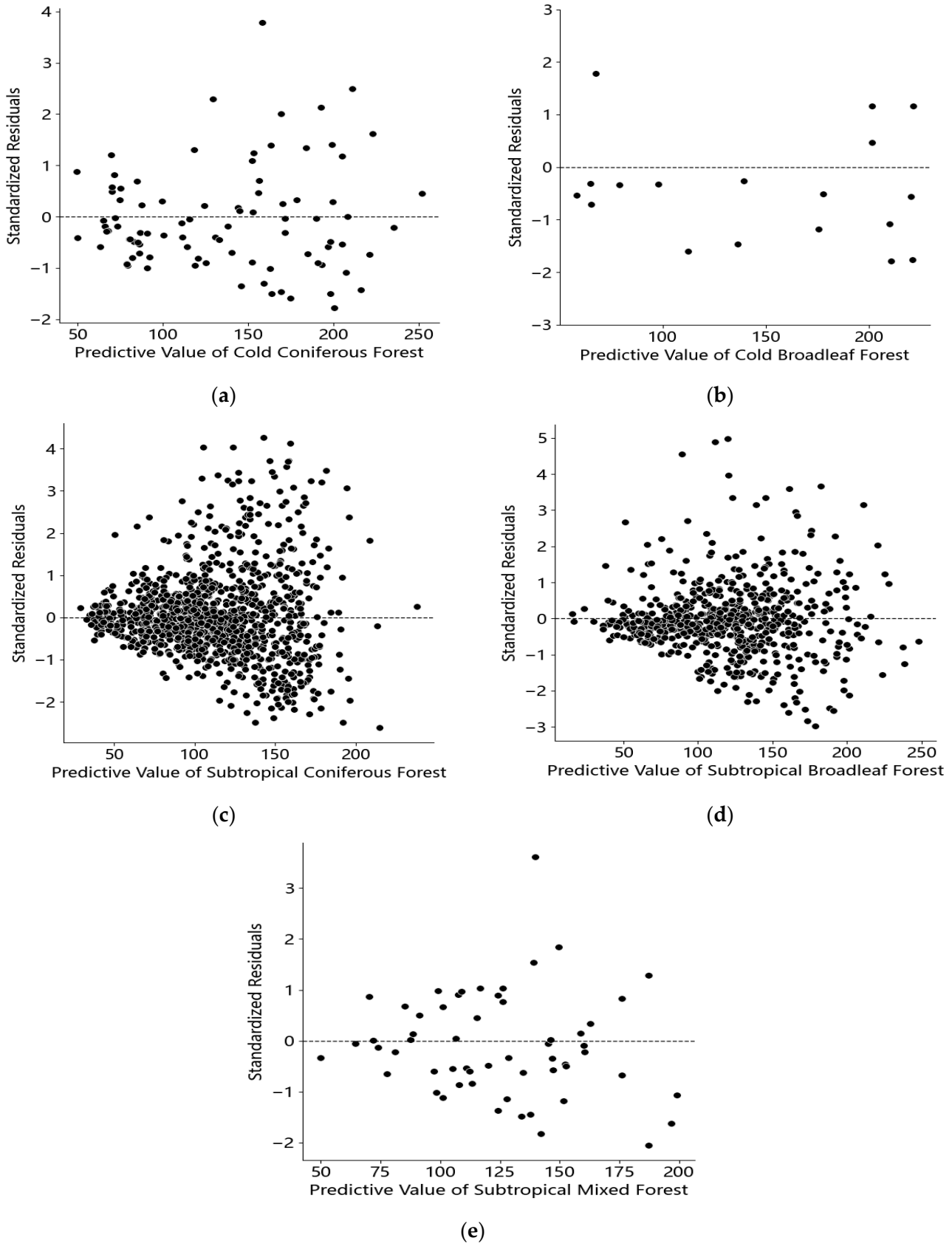


Figure 4. Distribution of Standardized Residuals for Predicted Values in the Optimal Model: (**a**) Cold Coniferous Forest. (**b**) Cold Coniferous Forest. (**c**)Cold Coniferous Forest. (**d**) Subtropical Coniferous Forest. (**e**) Subtropical Broadleaf Forest.

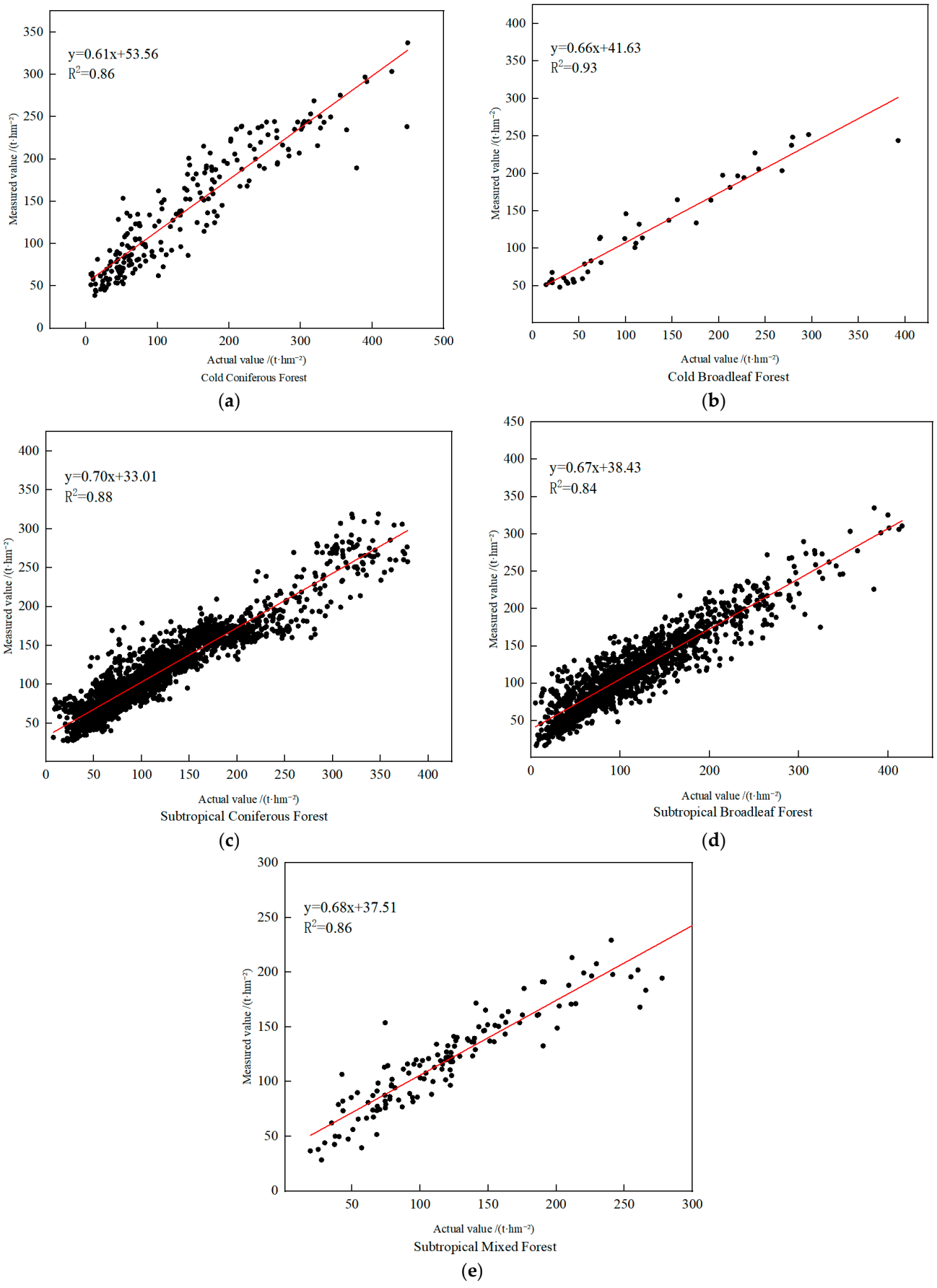


Figure 5. Scatter Plots and Regression Curves of Predicted vs. Actual Aboveground Fuel Loads for Different Vegetation Types: (**a**) Cold Coniferous Forest. (**b**) Cold Coniferous Forest. (**c**) Cold Coniferous Forest. (**d**) Subtropical Coniferous Forest. (**e**) Subtropical Broadleaf Forest.

3.5. Variable Importance Evaluation of the Optimal Models for Different Vegetation Types

In cold coniferous forests, the annual temperature range has the greatest impact on FL, followed by precipitation seasonality, canopy closure, and temperature seasonality. Accordingly, canopy coverage and climate variability play key roles in this forest type. For cold broadleaf forests, canopy closure is the primary influencing factor, indicating that a higher canopy density leads to greater fuel accumulation, consistent with the findings of Alvarez et al. [56]. Elevation and stand density also play significant roles, highlighting the importance of terrain and vegetation density in fuel accumulation. Additionally, climate factors such as temperature seasonality, annual mean temperature, and precipitation during the wettest month have a regulatory effect on FL in this forest type. In subtropical coniferous forests, the annual mean temperature is the main influencing factor, reflecting the dominant role of climate in this ecosystem. Mean diurnal range and average DBH are also important, suggesting that climate fluctuations and tree growth characteristics jointly affect FL. For subtropical broadleaf forests, stand density is the critical factor, followed by average DBH, annual mean temperature, and average tree height, indicating that tree size, structure, and moisture conditions significantly influence fuel accumulation. In subtropical mixed forests, stand density has the largest impact, underscoring the importance of canopy coverage and tree growth characteristics. Moreover, average tree height, precipitation seasonality, and average DBH significantly influence FL, suggesting an integrated effect of climate and topography in this forest type.

Overall, biological structural factors, such as canopy closure, DBH, and tree height, play a particularly important role in cold broadleaf, subtropical broadleaf, and subtropical mixed forests, with tree size and structure significantly impacting FL. By contrast, climate factors, such as annual mean temperature and temperature seasonality, dominate in cold coniferous and subtropical coniferous forests. These findings indicate distinct and consistent driving mechanisms of biological and climate factors across different forest types.

The forest FL across Sichuan Province was estimated using the optimal models for the five vegetation types described above. This initial spatial distribution of forest FL estimates is presented in Figure 6.

As shown in Figure 7, the predicted FL from the model was slightly lower than the observed values, indicating a certain degree of discrepancy between the predictions and actual measurements. Nevertheless, the overall distribution remained relatively stable. In terms of spatial distribution, low-FL areas (0–100 t/hm^2) are concentrated in the hilly regions of Central and Northeastern Sichuan, where forest fuel density is relatively sparse. Medium-FL areas (100–240 t/hm^2) are primarily located in the northwestern and southeastern parts of Sichuan, indicating a moderate fire risk in these regions. High-FL areas (240–310 t/hm^2) are concentrated in the high-altitude southwestern regions. In consideration of the climate and vegetation characteristics of these areas, FL is relatively high, suggesting a need for fire prevention measures to mitigate fire risk.

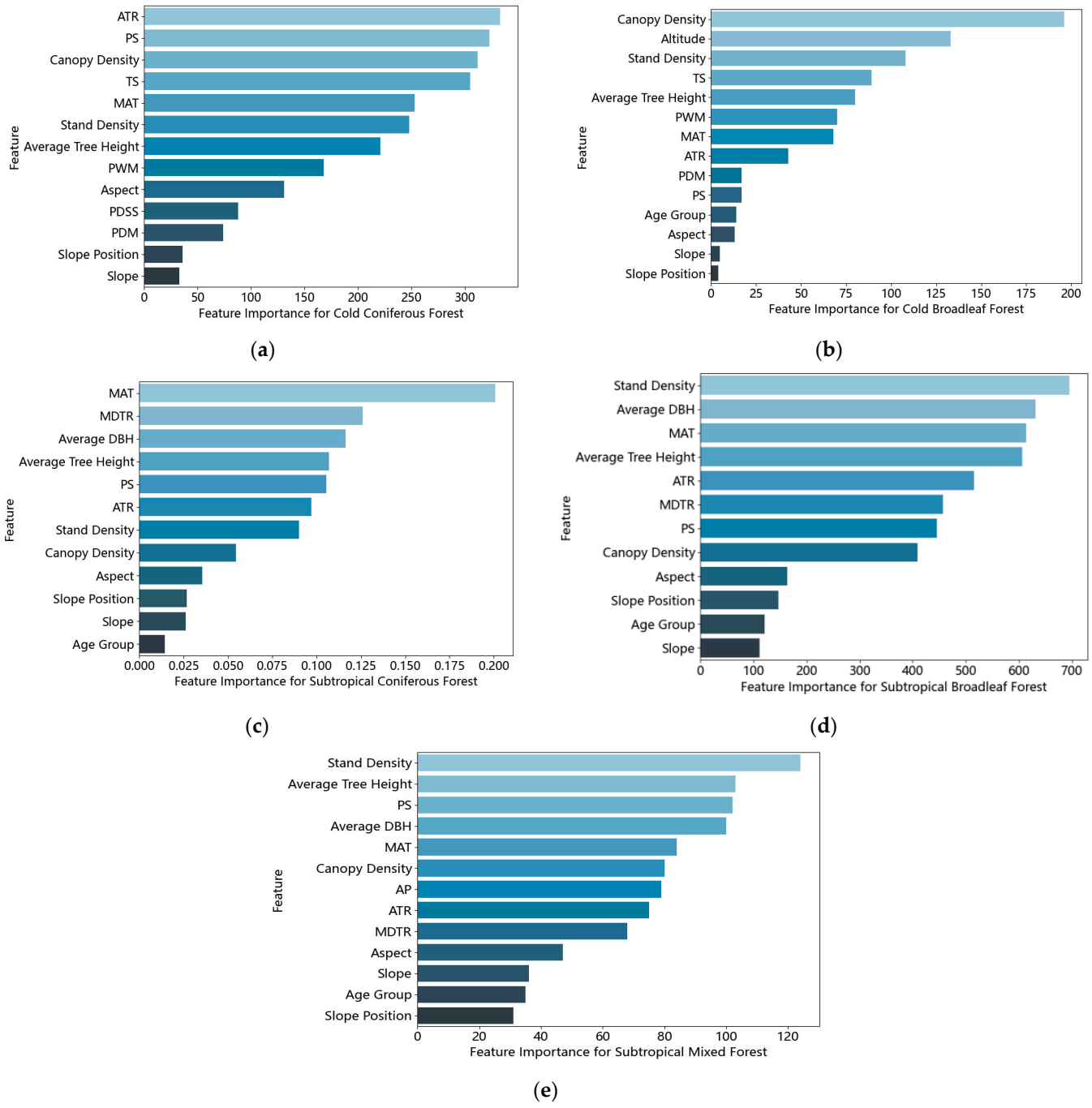


Figure 6. Evaluation of Variable Importance for the Optimal Model: (**a**) Cold Coniferous Forest. (**b**) Cold Coniferous Forest. (**c**) Cold Coniferous Forest. (**d**) Subtropical Coniferous Forest. (**e**) Subtropical Broadleaf Forest.

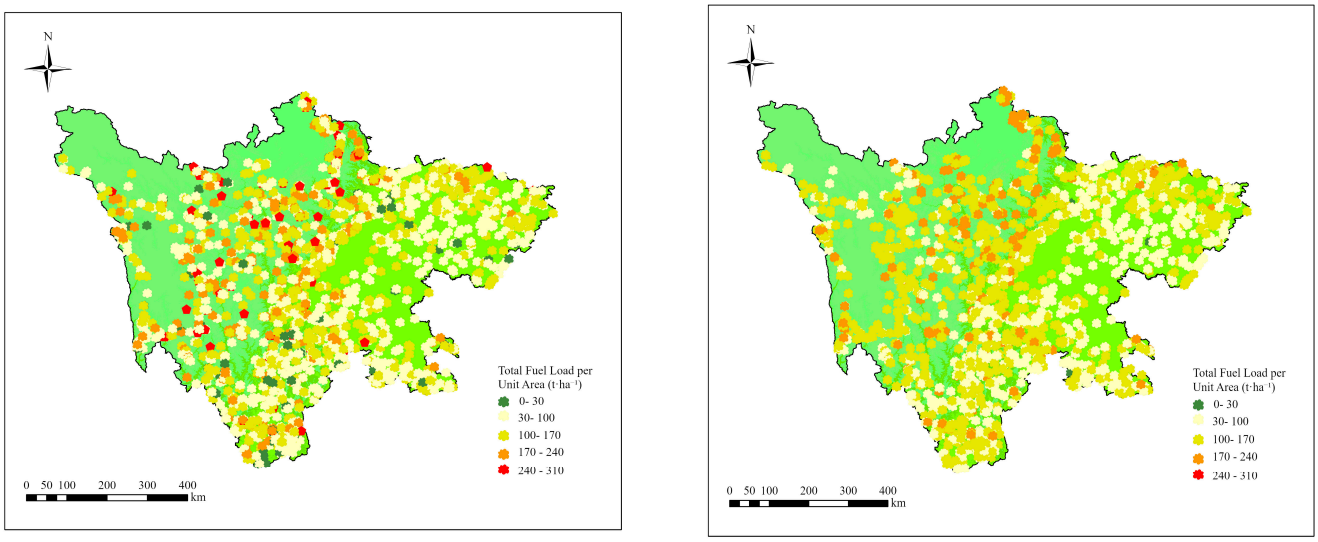


Figure 7. Predicted Forest FL Distribution in Sichuan Province. (**a**) Actual data. (**b**) Predicted data.

4. Discussion

4.1. Model Performance Analysis

XGBoost performed best in cold coniferous, cold broadleaf, subtropical broadleaf, and subtropical mixed forests, achieving its highest fit in cold broadleaf forests with an R^2 of 0.85. This result highlighted XGBoost's adaptability and stability when handling complex, nonlinear relationships in the data. The Stacking model showed strong performance in subtropical coniferous forests, with an R^2 of 0.82. These findings suggest that the suitability of different models varies across forest types, emphasizing the importance of selecting models tailored to specific vegetation characteristics.

4.2. Feature Importance of Influencing Factors

This study reveals that biological structural factors, such as canopy density and tree height, play a dominant role in cold broad-leaved forests, subtropical broad-leaved forests, and subtropical mixed forests. In contrast, climatic factors, including mean annual temperature and temperature seasonality, are the primary drivers in cold coniferous forests and subtropical coniferous forests. This disparity likely arises from the structural complexity and higher species diversity of broad-leaved and mixed forests, which reduce their dependence on climatic conditions. Conversely, coniferous forests, characterized by their simple canopy structure and limited species diversity, are more directly influenced by climatic variables such as temperature and precipitation. These findings align with those of Penman et al. [57], who identified climatic factors as significant drivers of fuel accumulation in coniferous forests, while biological structural factors, such as canopy species composition and vegetation complexity, primarily govern fuel accumulation in more complex forest types. Similarly, Daniels et al. [58] demonstrated that climatic factors had a particularly pronounced influence on fuel accumulation in single-species forests, especially under changing climatic conditions. Furthermore, Schwilk et al. [59] highlighted significant differences in the response mechanisms of various forest types to climatic and biological factors. These differences emphasize the need for forest management strategies that are tailored to the specific characteristics and ecological dynamics of each forest type.

4.3. Superiority of Machine Learning Models over Traditional Regression Methods

This study compared the predictive performance of machine learning models with multiple linear stepwise regression models. The results clearly demonstrated that machine learning models outperformed linear regression models across all evaluation metrics. Specifically, the XGBoost model addressed overfitting effectively through its regularization strategy while enhancing model stability and predictive accuracy by optimizing variable selection via feature importance ranking. Additionally, the Stacking method combined the predictive outputs of multiple base models, capitalizing on the complementary strengths of different algorithms in capturing data features. This advantage was particularly evident in subtropical coniferous forests, where multi-factor interactions and complex data distributions posed significant challenges. These findings align with the research of D'Este et al. [60], which emphasized that machine learning methods were well-suited for handling complex data characteristics and achieving precise estimations of forest fuel loads.

4.4. Forest Fire Risk Zoning

Using the optimal model, a spatial inversion of forest fuel loads in Sichuan Province was conducted. The results revealed that low-fuel-load areas were primarily concentrated in Central and Northeastern Sichuan, medium-fuel-load areas were mainly distributed in the northwest and southeast, and high-fuel-load areas were predominantly located in the high-altitude regions of Southwestern Sichuan. These findings align with the study by Wu et al. [61], which demonstrated that high-fuel-load areas are often concentrated in high-altitude regions, likely due to the combined influences of vegetation types and climatic conditions. To mitigate potential wildfire risks, it is recommended to enhance monitoring in high-risk areas, regularly remove combustible materials, and establish firebreaks to reduce the likelihood and severity of wildfires.

5. Conclusions

Based on data from 6848 arbor sample plots in Sichuan Province, this study developed fuel load prediction models by incorporating stand characteristics, site conditions, and climatic factors. The predictive performance of machine learning methods was compared to traditional multiple linear stepwise regression models. The findings demonstrated that the XGBoost and Stacking models significantly outperformed traditional regression models in terms of both predictive accuracy and stability. Specifically, XGBoost proved most effective for cold coniferous forests, cold broadleaf forests, subtropical broadleaf forests, and subtropical mixed forests, while the Stacking model excelled in subtropical coniferous forests. Feature importance analysis revealed that fuel loads in different vegetation types were influenced by a combination of biological, structural, and climatic factors. In broadleaf and mixed forests, fuel loads were primarily driven by biological structural factors, such as canopy density, tree height, and diameter at breast height (DBH). Conversely, coniferous forests were predominantly influenced by climatic factors, including mean annual temperature and temperature seasonality. Spatial inversion results indicated that high fuel load areas were concentrated in the high-altitude regions of Southwestern Sichuan, while low and medium fuel load areas were distributed in central and Northeastern Sichuan. This study provides a scientific foundation and technical support for forest fire risk assessment and prevention in Sichuan Province.

Author Contributions: Conceptualization, H.W.; methodology, H.W. and H.C.; validation, H.W. and H.S.; formal analysis, H.W. and H.C.; resources, H.W.; data curation, H.S.; writing—original draft preparation, H.W. and H.C.; writing—review and editing, C.D.; visualization, C.D. and Z.M.; supervision, K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research project of the Sichuan Forestry and Grassland Bureau (grant number G202206012, G202206012—2).

Data Availability Statement: The data are available upon request from the corresponding author.

Acknowledgments: We are grateful to the Sichuan Forestry and Grassland Survey and Planning Institute for supplying the valuable model data. We would like to express our gratitude to the English language editors, who helped in checking the grammar.

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

References

1. Seidl, R.; Fernandes, P.M.; Fonseca, T.F.; Gillet, F.; Jönsson, A.M.; Merganičová, K.; Netherer, S.; Arpaci, A.; Bontemps, J.-D.; Bugmann, H.; et al. Modelling Natural Disturbances in Forest Ecosystems: A Review. *Ecol. Model.* **2011**, *222*, 903–924. [CrossRef]
2. McManamay, R.A.; Vernon, C.R.; Chen, M.; Thompson, I.; Khan, Z.; Narayan, K.B. Dynamic Urban Land Extensification Is Projected to Lead to Imbalances in the Global Land-Carbon Equilibrium. *Commun. Earth Environ.* **2024**, *5*, 70. [CrossRef]
3. Liu, J.; Wang, Y.; Guo, H.; Lu, Y.; Xu, Y.; Sun, Y.; Gan, W.; Sun, R.; Li, Z. Spatial and Temporal Patterns and Driving Factors of Forest Fires Based on an Optimal Parameter-Based Geographic Detector in the Panxi Region, Southwest China. *Fire Ecol.* **2024**, *20*, 27. [CrossRef]
4. Song, Y.; Xu, C.; Li, X.; Oppong, F. Lightning-Induced Wildfires: An Overview. *Fire* **2024**, *7*, 79. [CrossRef]
5. Zong, X.; Tian, X.; Yao, Q.; Brown, P.M. An Analysis of Fatalities from Forest Fires in China, 1951–2018. *Int. J. Wildland Fire* **2022**, *31*, 507–517. [CrossRef]
6. Rifai, S.W.; De Kauwe, M.G.; Gallagher, R.V.; Cernusak, L.A.; Meir, P.; Pitman, A.J. Burn Severity and Post-Fire Weather Are Key to Predicting Time-To-Recover from Australian Forest Fires. *Earth's Future* **2024**, *12*, e2023EF003780. [CrossRef]
7. Marques, L. Crossing Tipping Points in the Amazon Rainforest: The Decisive Decade. *Highl. J.* **2024**, *3*, 4–33. [CrossRef]
8. Clelland, A.A.; Marshall, G.J.; Baxter, R.; Potter, S.; Talucci, A.C.; Rady, J.M.; Genet, H.; Rogers, B.M.; Natali, S.M. Annual and Seasonal Patterns of Burned Area Products in Arctic-Boreal North America and Russia for 2001–2020. *Remote Sens.* **2024**, *16*, 3306. [CrossRef]
9. Wang, J.; Azam, W. Natural Resource Scarcity, Fossil Fuel Energy Consumption, and Total Greenhouse Gas Emissions in Top Emitting Countries. *Geosci. Front.* **2024**, *15*, 101757. [CrossRef]
10. Pyne, S.J.; Andrews, P.L.; Laven, R.D. *Introduction to Wildland Fire: Fire Management in the United State*; Wiley: New York, NY, USA, USA, 1984.
11. Younes, N.; Yebra, M.; Boer, M.M.; Griebel, A.; Nolan, R.H. A Review of Leaf-Level Flammability Traits in Eucalypt Trees. *Fire* **2024**, *7*, 183. [CrossRef]
12. Lindenmayer, D.; Taylor, C.; Blanchard, W. Empirical Analyses of the Factors Influencing Fire Severity in Southeastern Australia. *Ecosphere* **2021**, *12*, e03721. [CrossRef]
13. Gordon, C.E.; Nolan, R.H.; Boer, M.M.; Bendall, E.R.; Williamson, J.S.; Price, O.F.; Kenny, B.J.; Taylor, J.E.; Denham, A.J.; Bradstock, R.A. Severe and Short Interval Fires Rearrange Dry Forest Fuel Arrays in South-Eastern Australia. *Fire* **2024**, *7*, 130. [CrossRef]
14. Geng, D.; Yang, G.; Ning, J.; Li, A.; Li, Z.; Ma, S.; Wang, X.; Yu, H. Modification of the Rothermel Model Parameters—The Rate of Surface Fire Spread of *Pinus koraiensis* Needles under No-Wind and Various Slope Conditions. *Int. J. Wildland Fire* **2024**, *33*, WF23118. [CrossRef]
15. Roberts, M.; Lareau, N.P.; Juliano, T.W.; Shamsaei, K.; Ebrahimian, H.; Kosovic, B. Sensitivity of simulated fire-generated circulations to fuel characteristics during large wildfires. *J. Geophys. Res.* **2024**, *129*, e2023JD040548. [CrossRef]
16. Stephens, S.L.; Foster, D.E.; Battles, J.J.; Bernal, A.A.; Collins, B.M.; Hedges, R.; Moghaddas, J.J.; Roughton, A.T.; York, R.A. Forest Restoration and Fuels Reduction Work: Different Pathways for Achieving Success in the Sierra Nevada. *Ecol. Appl.* **2024**, *34*, e2932. [CrossRef]
17. Faris, W.F.; Rakha, H.A.; Kafafy, R.I.; Idres, M.; Elmoselhy, S. Vehicle Fuel Consumption and Emission Modelling: An In-Depth Literature Review. *IJVSMT* **2011**, *6*, 318. [CrossRef]
18. Nolan, R.H.; Gibson, R.K.; Cirulis, B.; Holyland, B.; Samson, S.A.; Jenkins, M.; Penman, T.; Boer, M.M. Incorporating Burn Heterogeneity with Fuel Load Estimates May Improve Fire Behaviour Predictions in South-East Australian Eucalypt Forest. *Int. J. Wildland Fire* **2024**, *33*, WF22179. [CrossRef]
19. Ning, J.; Liu, H.; Yu, W.; Deng, J.; Sun, L.; Yang, G.; Wang, M.; Yu, H. Comparison of Different Models to Simulate Forest Fire Spread: A Case Study. *Forests* **2024**, *15*, 563. [CrossRef]
20. Baron, J.N.; Hessburg, P.F.; Parisien, M.-A.; Greene, G.A.; Gergel, S.E.; Daniels, L.D. Fuel Types Misrepresent Forest Structure and Composition in Interior British Columbia: A Way Forward. *Fire Ecol.* **2024**, *20*, 15. [CrossRef]

21. Alizadeh, M.R.; Adamowski, J.; Entekhabi, D. Land and Atmosphere Precursors to Fuel Loading, Wildfire Ignition and Post-Fire Recovery. *Geophys. Res. Lett.* **2024**, *51*, e2023GL105324. [CrossRef]
22. Fajardo-Cantos, Á.; Peña, E.; Plaza-Álvarez, P.; González-Romero, J.; Moya, D.; González-Camuñas, H.; Díaz, A.; Botella, R.; Lucas-Borja, M.E.; De Las Heras, J. Soil Response in a Mediterranean Forest Ecosystem of Southeast Spain Following Early Prescribed Burning. *Heliyon* **2024**, *10*, e37948. [CrossRef] [PubMed]
23. Vega, J.A.; Álvarez-González, J.G.; Arellano-Pérez, S.; Fernández, C.; Cuiñas, P.; Jiménez, E.; Fernández-Alonso, J.M.; Fontúrbel, T.; Alonso-Rego, C.; Ruiz-González, A.D. Developing Customized Fuel Models for Shrub and Bracken Communities in Galicia (NW Spain). *J. Environ. Manag.* **2024**, *351*, 119831. [CrossRef]
24. Chávez-Durán, Á.A.; García, M.; Olvera-Vargas, M.; Aguado, I.; Figueroa-Rangel, B.L.; Trucíos-Caciano, R.; Rubio-Camacho, E.A. Forest Canopy Fuel Loads Mapping Using Unmanned Aerial Vehicle High-Resolution Red, Green, Blue and Multispectral Imagery. *Forests* **2024**, *15*, 225. [CrossRef]
25. Kim, H.K.; Kim, Y.; Lee, S.M.; Ahn, K.Y. Studies on Combustion Characteristics and Flame Length of Turbulent Oxy−Fuel Flames. *Energy Fuels* **2007**, *21*, 1459–1467. [CrossRef]
26. Sheng, F.; Diao, J.; Wang, B.; Tang, X. Forest Fuel Load Models of Tree Layers in Different Forest Types in Shanghai. *Guangxi For. Sci.* **2023**, *52*, 384–393. [CrossRef]
27. Wang, Y.; Suo, A.; Gao, Y.; Wang, B.; Chen, F.; Liu, X. Surface Fuel Load and Influencing Factors of Yunnan Pine Forest in Xichang, Sichuan. *J. Beijing For. Univ.* **2023**, *45*, 100–108. [CrossRef]
28. Xiao, J. Development of Fuel Load Estimation Models for *Pinus massoniana* Forests in Fujian. *Fujian For. Sci. Technol.* **2023**, *50*, 89–94, 105. [CrossRef]
29. Chojnacky, D.C.; Heath, L.S.; Jenkins, J.C. Updated Generalized Biomass Equations for North American Tree Species. *Forestry* **2014**, *87*, 129–151. [CrossRef]
30. Nguyen, T.H.; Jones, S.; Reinke, K.J.; Soto-Berelov, M. Estimating Fine Fuel Loads in Eucalypt Forests Using Forest Inventory Data and a Modelling Approach. *For. Ecol. Manag.* **2024**, *561*, 121851. [CrossRef]
31. Sydorenko, S.; Gumeniuk, V.; De Miguel-Díez, F.; Soshenskiy, O.; Budzinskyi, I.; Koren, V. Assessment of the Surface Forest Fuel Load in the Ukrainian Polissia. *Fire Ecol.* **2024**, *20*, 35. [CrossRef]
32. Li, B.; Zhong, J.; Shi, G.; Fang, J. Forest Fire Spread Prediction Method Based on BP Neural Network. In Proceedings of the 2022 9th International Conference on Dependable Systems and Their Applications (DSA), Ürümqi, China, 4–5 August 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 954–959.
33. Sibanda, J. Machine Learning Models to Predict Cracking on Steel Slabs During Continuous Casting. Master's Thesis, Department of Computer and Information Science, Linköping University, Linköping, Sweden, 2024.
34. Peng, Y.; Su, H.; Sun, M.; Li, M. Reconstructing Historical Forest Fire Risk in the Non-Satellite Era Using the Improved Forest Fire Danger Index and Long Short-Term Memory Deep Learning-a Case Study in Sichuan Province, Southwestern China. *For. Ecosyst.* **2024**, *11*, 100170. [CrossRef]
35. Li, S.; Liang, X.; Wang, Z.; Liu, X.; Li, Z.; Zhang, Y. Comprehensive Comparative Study of Forest Fire Emergency Management between China and the United States. In Proceedings of the 4th International Conference on Public Management and Intelligent Society, PMIS 2024, Changsha, China, 15–17 March 2024; EAI: Changsha, China, 2024.
36. Yang, Z.; Zhang, B.; Chen, J.; Hou, Y.; Wu, Y.; Xie, H. Characteristics of Spatial and Temporal Variation in Drought in the Sichuan Basin from 1963 to 2022. *Sustainability* **2024**, *16*, 8397. [CrossRef]
37. Sichuan Forestry and Grassland Bureau. *Implementation Guidelines for the First Forest and Grassland Fire Risk Survey of Sichuan Province: Forest Fuel Sample Plots, Grassland Fuel Sample Plots, and Large Plot Surveys*; Internal Material; Sichuan Forestry and Grassland Bureau: Chengdu, China, 2021.
38. Weise, D.R.; Wright, C.S. Wildland Fire Emissions, Carbon and Climate: Characterizing Wildland Fuels. *For. Ecol. Manag.* **2014**, *317*, 26–40. [CrossRef]
39. Xu, F.; Chen, W.; Xie, R.; Wu, Y.; Jiang, D. Vegetation Classification and a Biomass Inversion Model for Wildfires in Chongli Based on Remote Sensing Data. *Fire* **2024**, *7*, 58. [CrossRef]
40. Fang, Z.; Bailey, R.L. Height–Diameter Models for Tropical Forests on Hainan Island in Southern China. *For. Ecol. Manag.* **1998**, *110*, 315–327. [CrossRef]
41. Liu, Y.; Goodrick, S.; Heilman, W. Wildland Fire Emissions, Carbon, and Climate: Wildfire–Climate Interactions. *For. Ecol. Manag.* **2014**, *317*, 80–96. [CrossRef]
42. Sikkink, P.G.; Keane, R.E. A Comparison of Five Sampling Techniques to Estimate Surface Fuel Loading in Montane Forests. *Int. J. Wildland Fire* **2008**, *17*, 363. [CrossRef]
43. Northup, B.K.; Zitzer, S.F.; Archer, S.; McMurtry, C.R.; Boutton, T.W. Above-Ground Biomass and Carbon and Nitrogen Content of Woody Species in a Subtropical Thornscrub Parkland. *J. Arid. Environ.* **2005**, *62*, 23–43. [CrossRef]
44. Tsarouhas, V.; Gullberg, U.; Lagercrantz, U. Mapping of Quantitative Trait Loci (QTLs) Affecting Autumn Freezing Resistance and Phenology in Salix. *Theor. Appl. Genet.* **2004**, *108*, 1335–1342. [CrossRef] [PubMed]

45. Moritz, M.A.; Parisien, M.-A.; Batllori, E.; Krawchuk, M.A.; Van Dorn, J.; Ganz, D.J.; Hayhoe, K. Climate Change and Disruptions to Global Fire Activity. *Ecosphere* **2012**, *3*, 1–22. [CrossRef]
46. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. *Intl. J. Climatol.* **2017**, *37*, 4302–4315. [CrossRef]
47. Dahouda, M.K.; Joe, I. A Deep-Learned Embedding Technique for Categorical Features Encoding. *IEEE Access* **2021**, *9*, 114381–114391. [CrossRef]
48. Bujang, M.A. An Elaboration on Sample Size Determination for Correlations Based on Effect Sizes and Confidence Interval Width: A Guide for Researchers. *Restor. Dent. Endod.* **2024**, *49*, e21. [CrossRef]
49. Precious, A.O.; Aisha, A.; Peter, S. Modeling Economic Relationships: A Statistical Investigation of Trends and Relationships. *Soc. Sci. Humanit. J.* **2024**, *8*, 3778–3796. [CrossRef]
50. Segal, M.R. *Machine Learning Benchmarks and Random Forest Regression*; Wolters Kluwer: Alphen aan den Rijn, The Netherlands, 2003.
51. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.
52. Jakkula, V. *Tutorial on Support Vector Machine (SVM)*; School of EECS, Washington State University: Pullman, WA, USA, 2010.
53. Song, Y.; Liang, J.; Lu, J.; Zhao, X. An Efficient Instance Selection Algorithm for *k* Nearest Neighbor Regression. *Neurocomputing* **2017**, *251*, 26–34. [CrossRef]
54. Noviandy, T.R.; Idroes, G.M.; Hardi, I. Machine Learning Approach to Predict AXL Kinase Inhibitor Activity for Cancer Drug Discovery Using Bayesian Optimization-XGBoost. *JSCDM* **2024**, *5*, 46–56. [CrossRef]
55. Frazier, P.I. A Tutorial on Bayesian Optimization. *arXiv* **2018**, arXiv:1807.02811.
56. Alvarez, A.; Gracia, M.; Vayreda, J.; Retana, J. Patterns of Fuel Types and Crown Fire Potential in *Pinus halepensis* Forests in the Western Mediterranean Basin. *For. Ecol. Manag.* **2012**, *270*, 282–290. [CrossRef]
57. Penman, T.D.; York, A. Climate and Recent Fire History Affect Fuel Loads in Eucalyptus Forests: Implications for Fire Management in a Changing Climate. *For. Ecol. Manag.* **2010**, *260*, 1791–1797. [CrossRef]
58. Daniels, L.D.; Maertens, T.B.; Stan, A.B.; McCloskey, S.P.J.; Cochrane, J.D.; Gray, R.W. Direct and Indirect Impacts of Climate Change on Forests: Three Case Studies from British Columbia. *Can. J. Plant Pathol.* **2011**, *33*, 108–116. [CrossRef]
59. Schwilk, D.W.; Keeley, J.E.; Knapp, E.E.; McIver, J.; Bailey, J.D.; Fettig, C.J.; Fiedler, C.E.; Harrod, R.J.; Moghaddas, J.J.; Outcalt, K.W.; et al. The National Fire and Fire Surrogate Study: Effects of Fuel Reduction Methods on Forest Vegetation Structure and Fuels. *Ecol. Appl.* **2009**, *19*, 285–304. [CrossRef]
60. D'Este, M.; Elia, M.; Giannico, V.; Spano, G.; Lafortezza, R.; Sanesi, G. Machine Learning Techniques for Fine Dead Fuel Load Estimation Using Multi-Source Remote Sensing Data. *Remote Sens.* **2021**, *13*, 1658. [CrossRef]
61. Wu, X.; Zhang, G.; Yang, Z.; Tan, S.; Yang, Y.; Pang, Z. Machine Learning for Predicting Forest Fire Occurrence in Changsha: An Innovative Investigation into the Introduction of a Forest Fuel Factor. *Remote Sens.* **2023**, *15*, 4208. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.