



Feng Xu *, Wenjing Chen, Rui Xie, Yihui Wu and Dongming Jiang

College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; wjchen@njfu.edu.cn (W.C.); ruixie@njfu.edu.cn (R.X.); wuyihui@njfu.edu.cn (Y.W.); 17660975577@njfu.edu.cn (D.J.)

* Correspondence: ramboxufeng@njfu.edu.cn; Tel.: +86-139-1391-7970

Abstract: Vegetation classification, biomass assessment, and wildfire dynamics are interconnected wildfire-ecosystem components. The Chongli District, located in Zhangjiakou City, was the venue for skiing at the 2022 Winter Olympics. Its high mountains and dense forests create a unique environment. The establishment of alpine ski resorts highlighted the importance of comprehensive forest surveys. Understanding vegetation types and their biomass is critical to assessing the distribution of local forest resources and predicting the likelihood of forest fires. This study used satellite multispectral data from the Sentinel-2B satellite to classify the surface vegetation in the Chongli District through K-means clustering. By combining this classification with a biomass inversion model, the total biomass of the survey area can be calculated. The biomass inversion equation established based on multispectral remote sensing data and terrain information in the Chongli area have a strong correlation (shrub forest $R^2 = 0.811$, broadleaf forest $R^2 = 0.356$, coniferous forest $R^2 = 0.223$). These correlation coefficients are key indicators for our understanding of the relationship between remote sensing data and actual vegetation biomass, reflecting the performance of the biomass inversion model. Taking shrubland as an example, a correlation coefficient as high as 0.811 shows the model's ability to accurately predict the biomass of this type of vegetation. In addition, through multiple linear regression, the optimal shrub, broadleaf, and coniferous forest biomass models were obtained, with the overall accuracy reaching 93.58%, 89.56%, and 97.53%, respectively, meeting the strict requirements for survey accuracy. This study successfully conducted vegetation classification and biomass inversion in the Chongli District using remote sensing data. The research results have important implications for the prevention and control of forest fires.

Keywords: vegetation classification; biomass; remote sensing data; sentinel data

1. Introduction

Wildfires are uncontrolled fires caused by vegetation that spread rapidly across natural landscapes [1]. Their ignition and spread are affected by weather conditions, topography, and combustible vegetation [2]. The intricate interactions between vegetation classification, biomass, and wildfire are critical to understanding wildfire dynamics, which underscores the need for effective wildfire management and prevention. In recent years, the frequency and severity of forest fires have increased significantly over much of the world [3]. Large-scale forest fires can result in considerable environmental damage [4–6], disrupting the composition and structure of ecosystems significantly [7,8]. Forest fire risk in China shows an increasing trend, with more areas under the high-risk zone [9], highlighting the urgency of responding to this escalating threat.

The Chongli District, renowned as the venue for China's esteemed 2022 Winter Olympics skiing program, necessitates prompt forest resource surveys. The region's mountainous terrain and distinctive microclimate contribute to a diverse array of natural conditions, rendering the assessment of forest stand structures and surveying forest resources a formidable task.



Citation: 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. https:// doi.org/10.3390/fire7020058

Academic Editor: Grant Williamson

Received: 23 November 2023 Revised: 13 February 2024 Accepted: 14 February 2024 Published: 17 February 2024



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/). Traditional ground surveys reliant on manual labor are not only costly but also time consuming. To address these challenges, remote sensing technology offers a more efficient and accurate approach to swiftly identify vegetation types and discern surface vegetation by analyzing the spectral characteristics of remote sensing data [10].

Earlier studies on vegetation remote sensing classification primarily relied on visual interpretation. However, the accuracy of interpretation often relied upon image quality and manual expertise, resulting in lengthy and time-sensitive operations [11]. The professionals must obtain precise target feature data from remote sensing images through direct assessment or by utilizing auxiliary decoding instruments. In recent years, rapid advancements in remote sensing and sensor technologies have yielded a wealth of abundant and precise multispectral and hyper-spectral images, enabling vegetation classification based on remote sensing [12–14] data. Hyperspectral remote sensing employs imaging spectrometers to capture high-resolution spectral image data across various regions of the electromagnetic spectrum, including the visible, near-infrared, mid-infrared, and thermal-infrared bands, within a narrow and continuous wavelength range [15]. This advanced technology not only captures target images but also obtains precise spectral data, providing the capability to create hyperspectral "maps in one". Compared to traditional remote sensing methods, hyperspectral imaging offers more comprehensive data due to its exceptional spectral resolution. Hyperspectral images with narrow-band data extract more information than multispectral images with wide-band data. Therefore, it is better suited for quantitatively distinguishing feature components and improving accuracy in feature inversion.

At the hyperspectral image scale, the primary focus in studying vegetation identification and classification lies in examining the impact of algorithms on classification accuracy. In a study by Buddenbaum et al. [16], hyperspectral images were employed to classify both the species and age classes of coniferous forests. The results demonstrated that the combination of spectral information with dry density or texture information yielded comparable and more accurate classification outcomes compared to using spectral information alone. Zhu Honglei et al. [17] conducted a study on the diversity of riverbank plant communities in the Henan section of the Yellow River, utilizing UAV remote sensing and artificial neural networks. The overall classification accuracy achieved an impressive 61.42%. Li Chan et al. [18] employed plant leaf spectral data and applied three machine learning algorithms, namely k-nearest neighbor, support vector machine, and random forest, to classify plant species in the agricultural region of Yixing, Jiangsu Province. The maximum classification accuracy attained was 94.74%. Du Xin et al. [19] conducted research based on Pleiades 1A/1B remote sensing data, integrating spectral, topographic, and texture information. By employing a combination of projection tracing and machine learning techniques, plant classification in the eastern part of Shenzhen City was achieved with an accuracy exceeding 70%.

Existing vegetation classification studies typically focus on flat terrains like river mudflats, agricultural regions, and suburbs, characterized by limited vegetation species. Commonly used methods include supervised machine learning algorithms like artificial neural networks, k-nearest neighbor, support vector machines, and random forests. However, traditional methods achieve only modest accuracy, reaching a maximum of 70%. In the case of Chongli, with its vast area, high altitudes, and diverse vegetation, a different approach is needed. Using the unsupervised K-Means algorithm for remote vegetation classification eliminates the need for manually annotating data, making data processing more efficient in the large study area.

The results of vegetation classification have a direct impact on forest biomass [20], which serves as an indicator of energy accumulation in the production and metabolic processes of an ecosystem within a specific area. It is a crucial indicator that reflects the structural and functional characteristics of the forest ecosystem, as well as its development potential [21]. These results hold significant practical significance for the scientific management and economic development of forested lands [22]. Analyzing the spatial and temporal dynamics of biomass allows for a comprehensive understanding of long-term trends in forest ecosystem carbon stock changes and the affected areas [23]. This knowledge

is invaluable for studying terrestrial carbon cycles and forest fire prevention in the Chongli District. Among the various methods for estimating biomass [24], the field survey method is labor intensive, costly, and time consuming. Additionally, it lacks spatial and temporal coverage, making it challenging to apply in large or remote areas. The use of airborne LiDAR [25] systems, although effective, requires substantial resources and manpower and is limited in its applicability to large-scale study areas. Therefore, exploring alternative approaches, such as remote sensing and machine learning techniques, can provide a more efficient and cost-effective means of estimating biomass in large study areas like Chongli District. These methods leverage the advantages of remote sensing data and advanced algorithms to overcome the limitations of traditional approaches.

Yali Zhu et al. [26] developed a model to estimate the crown and aboveground biomass of poplar trees in the vicinity of the Tarim Basin in the Xinjiang Uygur Autonomous Region, leveraging UAV remote sensing data. The model achieved an impressive accuracy of 95.63%. In a similar vein, Ye Pingquan et al. [27] employed three machine learning regression algorithms to estimate the aboveground biomass of acacia plantation forests, incorporating slope-direction variables. This approach led to an enhancement in the accuracy of the aboveground biomass estimation model. Guo Chaofan et al. [28] established a relationship model between 21 representative vegetation indices and biomass, focusing on the grassland area in Haiyan County, Qinghai Province. Moving on, Liu Yanhui et al. [29] devised a vegetation biomass inversion model for the reclamation area of a surface coal mine in the Inner Mongolia grassland. They accomplished this by integrating Worldview-3 and Sentinel-1 SAR data, achieving a remarkable accuracy of $R^2 = 0.7983$. Lastly, Li Tianchi et al. [30] developed a biomass estimation model for winter wheat, utilizing correlation and stepwise regression analyses in conjunction with digital images captured during its flowering period and hyperspectral data. The model attained a peak accuracy of 90%.

Previous research on biomass inversion has predominantly focused on individual vegetation species, neglecting areas with a more intricate distribution of forest plant species. In this study, we aim to address this gap by constructing a multiple linear regression model that integrates the outcomes of remote sensing-based plant classification and data from the regional forest resources type II survey. This step-by-step approach enhances the practicality and feasibility of developing a biomass inversion model for complex study areas. Investigating surface vegetation classification and biomass inversion in Chongli serves as both a foundation for understanding the local ecosystem and a means of providing technical support for real-time monitoring of vegetation changes in the region.

2. Materials and Methods

2.1. Overview of the Study Area

The study was conducted in the Chongli District, located in Zhangjiakou City, Hebei Province (40°47′–41°17′ N, 114°17′–115°34′ E) [31]. This region lies within the transitional zone between the Inner Mongolia Plateau and the North China Plateau [31]. It is characterized by the presence of the Inner Mongolia Grassland to the north and the downtown area of Zhangjiakou to the south. The topography exhibits a general pattern of higher elevations in the northeast and lower elevations in the southwest, with a natural inclination. The study area is encompassed by mountains, with undulating terrain in the east, middle, and west, and three large ditches traversing the entire region. The total area of the study area is approximately 2300 km², with elevations ranging from 812 m to 2169 m [32]. Geomorphologically, the region can be divided into two types: tectonic denudation plateau areas and eroded mountain areas. Figure 1 is a true-color satellite imagery of the Chongli District.

The climate in Chongli County belongs to the temperate subarid zone of the East Asian continental monsoon climate [31]. Due to its geographical location and topography, the air activity is frequent in winter, the temperature rises quickly in spring, but fluctuates greatly, the frost period is late, the rainfall is low, and the number of windy days is high. Summer is cool and short, the temperature is relatively stable, the temperature difference between day and night is large, and the rainfall is concentrated. Due to the topography of the mountainous areas, there are hailstorms and rainstorms from time to time. Autumn temperatures drop rapidly, and the first frost appears earlier.

Chongli District has an average summer temperature of 19 °C, an average winter temperature of -12 °C [33], an average wind speed of only two degrees, early snowfall, thick snow, and a long snow-storage period. The average precipitation in a calendar year is 488 mm [33], the total precipitation is 1.13 billion cubic meters, the total annual runoff is 100.69 million cubic meters, and the average temperature of the calendar year is 3.7 °C. Chongli District is rich in vegetation resources, with an afforestation area of 3867 hectares. The area of forest conservation is 10,220 hectares. The total area of forest land is 2.35 million mu. By 2021, the forest coverage rate will reach 67% [33].



Figure 1. True-color satellite imagery of the Chongli District.

The vegetation in the Chongli area is primarily composed of deciduous broad-leaved forests. Various types of forests are present, including shrub forests, economic forests, broad-leaved forests (both artificial and natural), and coniferous forests. Based on the survey results, it is evident that shrub forests predominantly constitute the broad-leaved forests, with a significant number of them located in parks and green spaces. Moreover, ornamental shrubs, such as octagonal golden discs, contribute to the landscape. In contrast, nurseries and sparse forests cover a smaller area, primarily situated within parks or designated green spaces. The prevailing tree species in the region encompass birch, cypress, poplar, hazel wood, larch, and others.

2.2. The Source and Processing of Data

The Sentinel-2B [34] satellite plays a crucial role in the Copernicus program, which is led by the European Space Agency (ESA) and focuses on global monitoring for environment and security (GMES). Launched on 7 March 2017, and placed into orbit by the ESA, the Sentinel-2B satellite is an integral part of this program. It captures images in three different resolution bands: 10 m, 20 m, and 60 m. In this study, remote sensing data from the L1C satellite were obtained by downloading from the European Space Agency website (https://scihub.copernicus.eu/dhus/#/home, accessed on 16 June 2021).

To enhance the accuracy of the remote sensing data, atmospheric corrections were performed on the 10 m and 20 m resolution bands of the L1C-level data. This correction process was carried out using the Sen2Cor plug-in, which is provided by the European Space Agency (ESA). After the corrections, the data were transformed into L2A level. Subsequently, the L2A-level data were resampled to a 10 m resolution using the nearest

neighbor method in the SNAP (Sentinel Application Platform) software, also provided by the ESA. The resampled data were then exported to file types that are compatible with

ENVI software and projected to the Beijing 54 coordinate system for further analysis. The vector boundary data of Chongli District were obtained using the National Geographic Information Resource Service Catalog System. The specific dataset used was the new map number K50, which was produced in 2017. The geodetic datum utilized for this dataset is the 2000 National Geodetic Coordinate System. Once the original data were obtained, the vector boundary of Chongli District in Zhangjiakou City was extracted individually using ArcGIS software. This extraction process resulted in the acquisition of the vector boundary data for the study area. Finally, the data were reprojected to the Beijing 54 coordinate system for consistency and compatibility.

The field survey data used in this study were obtained from the Class II survey of forest resources in Chongli District, Zhangjiakou City. It was assumed that the sample plots selected during the survey represented a single vegetation type, meaning there was no mixing of different types of vegetation. The Digital Elevation Model (DEM) [35] used in this research is a digital representation of the land surface morphology, providing information about factors such as slope, slope direction, and slope change rate across space. The specific DEM used in this study was obtained from NASA EARTH SEARCH and has a horizontal spatial resolution of 30 m and a vertical spatial resolution of 1 m. To facilitate subsequent calculations, the horizontal spatial resolution was oversampled to 10 m using the nearest neighbor method. Additionally, the Beijing 54 coordinate system was selected as the coordinate system for this study.

The Normalized Difference Vegetation Index (NDVI) [36] can be calculated using the following equation:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$
(1)

where NIR, RED are the infrared band reflectance values, respectively.

There is a high correlation between the green plant covered area and NDVI, so the study area can be masked using NDVI to remove the interference of non-plant components in the study. The images of the study area were masked using NDVI greater than 1, 2, 3, 4, 5, 6, 7 and 8 as thresholds and the results obtained are shown in Figure 2.



Figure 2. Different NDVI threshold mask effects.

The red part of the figure shows the vegetation cover area, and the non-vegetation cover area has been removed by the mask. By combining the visual observation with the true color satellite image (R:G:B = Band4:Band3:Band2), we can see that, when NDVI > 0.7,

the area covered by vegetation and the area obtained from the mask have a higher degree of conformity, and therefore the pixels obtained from the mask when NDVI > 0.7 are taken

3. Method

3.1. Vegetation Classification

as the effective pixels in the subsequent studies.

The K-Means Clustering Algorithm is an iterative cluster analysis algorithm, categorized as an unsupervised classification algorithm. It operates on the principle of iteratively identifying the centroid of each cluster and reallocating objects within the clusters based on their mean values. This process facilitates the classification of data points into distinct clusters.

Cluster analysis was predominantly carried out in this study using ENVI and Python. From the previously obtained field survey data, 500 groups of three vegetation types, namely shrub forest, coniferous forest, and broad-leaved forest, were randomly selected for cluster analysis. The training set consisted of the initial 80% (400 groups) of each vegetation type, while the remaining 20% (100 groups) formed the test set. Remote sensing images were initially classified into 10 classes using ENVI, and the preliminary classification results were summarized as shown in Figure 3.

Following the substitution of sample points from the training set into the classification result map using Python, the proportions of the three vegetation types in each class were computed. The dominant vegetation within the region was selected to represent the current classification. The statistical findings are presented in Table 1.

The preliminary classification results of K-Means clustering can be merged to further obtain the merged classification results as shown in Figure 4.

From Table 2, it is evident that the number of valid points (points located within areas other than the non-study area and non-vegetation cover area) is smaller compared to the points used in the test, regardless of whether it pertains to shrub forest, coniferous forest, or broad-leaved forest. Approximately 21% to 29% of the points fell within the non-study area or the non-vegetation cover area. The reason for points falling within the non-study area could be attributed to imprecise vector boundaries of the Chongli District or some data collection points exceeding the administrative boundaries during the field survey. Similarly, points falling within the non-vegetation cover area can be attributed to the NDVI threshold being set higher, resulting in the masking of effective pixels with lower NDVI values. For points falling in these two places, we do not calculate the number of valid points, the number of correct points and the accuracy, so we use '/' as placeholders.



Figure 3. K-Means preliminary classification results.

Classes	Shrub	Coniferous Forest	Broadleaf Forest	Dominant Vegetation	Emark
class 0	0.00%	0.00%	0.00%	none	non-research area
class 1	27.11%	31.50%	41.39%	none	non-vegetation
class 2	0.00%	100.00%	0.00%	coniferous forest	alea
class 3	0.00%	0.00%	100.00%	broadleaf forest	
class 4	0.00%	0.00%	0.00%	none	
class 5	0.00%	100.00%	0.00%	coniferous forest	
class 6	16.00%	56.00%	28.00%	coniferous forest	
class 7	31.07%	48.54%	20.39%	coniferous forest	
class 8	39.68%	29.70%	30.63%	shrub	
class 9	32.69%	32.41%	34.90%	broadleaf forest	

Table 1. K-Means preliminary classification results.



Figure 4. K-Means merged classification results.

Table 2. Classification accuracy.

Classes	Valid Point	Correct Point	Accuracy
non-research area	/	/	/
non-vegetation area	/	/	/
shrub	74	44	59.46%
coniferous forest	79	25	31.65%
broadleaf forest	71	43	60.56%

The classification results indicate that broadleaf forests achieved the highest accuracy at 60.56%. The difference in accuracy between shrub forests and broad-leaf forests was relatively small, with a rate of 59.49%. Coniferous forests exhibited the lowest classification accuracy at 31.65%. It is worth noting that the accuracy of classification using unsupervised methods is not particularly high for shrub forests, coniferous forests, or broadleaf forests. However, based on these classification results, it can be observed that the K-Means classification method can be utilized for rough classification of broadleaf forests and shrub forests in scenarios where high accuracy requirements are not essential.

3.2. Biomass Inversion

The vegetation indices such as GNDVI [37], OSAVI [38], SR2 [39], SR3 [40], and GI [41] in Table 3 can be obtained directly by substituting the raw band data as in Equations (2),

(3), (4), (5) and (6), respectively. Texture features such as COR(*), VAR(*), and CONT(*) can be calculated using the Gray-level Co-occurrence Matrix (GLCM) [42].

Table 3. Independent variables involved in the construction & calculation of biomass inversion models.

Types	Independent Variables	Number
original band	CA, BLUE, GREEN, RED, VRE1, VRE2, VRE3 WV, NIR, N_NIR, SWIR1, SWIR2	12
vegetation abundance	LOW, CLF	2
terrain factor	DEM	1
vegetation index	GNDVI, OSAVI, SR2, SR3, GI	5
texture feature	Cor(*), Var(*), Cont(*), Mean(*), Homo(*) Diss(*), Entr(*), Sec(*)	96

$$GNDVI = \frac{N_NIR - GREEN}{N_NIR + GREEN}$$
(2)

$$OSAVI = \frac{NIR - RED}{NIR + RED + 0.16}$$
(3)

$$SR2 = \frac{NIR}{VRE1}$$
(4)

$$SR3 = \frac{N_{NIR}}{BLUE}$$
(5)

$$GI = \frac{GREEN}{RED}$$
(6)

3.2.1. Portability Studies of Existing Biomass Models

The multiple linear regression method is employed to calculate biomass by considering various independent variables such as original band, band index, texture features, elevation, and other data [43]. The biomass is treated as the dependent variable, and multiple linear regression equations are constructed to represent the linear relationships between variables that exhibit high correlation with biomass.

In a previous study, Zhou Xisheng [44] optimized a biomass inversion model for low vegetation, coniferous forests, and broad-leaved forests with high accuracy by building stepwise regression models.

The low vegetation inversion equation is as follows:

$$B_{s} = 10 \times \begin{bmatrix} -171.896 - 49.335 \times LOW + 76.406 \times CLF + 316.404 \times GNDVI \\ -13.710 \times SR2 - 0.365 \times COR(VRE2) + 1.807 \times DEM \end{bmatrix}$$
(7)

In this study, it is assumed that there is a single type of vegetation in the samples selected for the field survey, so for the low vegetation, LOW is taken as 1 and CLF is taken as 0. The inverse equation for broadleaf forest is as follows.

$$B_{b} = 10 \times \begin{bmatrix} 660.327 - 16.739 \times COR(VRE2) - 3601.606 \times GREEN \\ +9.944 \times COR(SWIR1) - 695.210 \times OSAVI - 196.861 \times VAR(VRE2) \\ +98.126 \times CONT(SWIR1) \end{bmatrix}$$
(8)

The coniferous forest inversion equation is as follows.

$$B_{c} = 10 \times \begin{bmatrix} 183.909 - 473.034 \times SWIR1 - 0.016 \times SR3 - 0.232 \times DEM \\ +0.299 \times GI + 14.747 \times COR(VRE2) \end{bmatrix}$$
(9)

- 11

1.

The results obtained from the biomass measurements of the study area by the above inverse equations are shown in Table 4.

Table 4. The biomass calculation results of the existing inversion models in Chongli District.	Table 4.	The biomass ca	lculation results	s of the existin	ng inversion m	odels in Chong	lı District.
--	----------	----------------	-------------------	------------------	----------------	----------------	--------------

Classes	Maximum Biomass (kg/m ²)	Minimum Biomass (kg/m ²)	Average Biomass (kg/m ²)
shrub	523.73	-1.71	275.75
coniferous forest	-655.36	-139,715.28	-62,239.59
broadleaf forest	2830.66	-583,657.11	-167,978.80

Based on the table provided, it is evident that the biomass data obtained from the aforementioned model includes negative numbers. This outcome clearly deviates from the typical pattern of biomass values, leading to the conclusion that the biomass inversion model is invalid. Therefore, it is not advisable to directly apply this inversion model for biomass calculation in the Chongli area.

3.2.2. Construction of a Biomass Inversion Model

Stepwise Regression (SR) is a multivariate linear model that follows a specific construction method. It involves the iterative addition or removal of variables based on their significance in the equations, using forward, backward, and stepwise approaches. The model is constructed by evaluating the changes in equation significance and incorporating or eliminating variables accordingly.

The stepwise regression model was formulated utilizing the data acquired from the Class II survey of forest resources in the Chongli District. In this survey, 100 sample points were randomly selected from each vegetation type, and any invalid sample points were excluded. The construction of the inversion model involved using the initial 80% of the sample points, while the biomass of each vegetation type was directly calculated using the respective biomass model and a specific formula. The remaining 20% of the sample points were reserved for evaluating the inversion model. The model employed for calculating the biomass of the initial 80% of the sample points can be represented as shown in Table 5.

The biomass per unit (or per unit area) of the plants calculated above can be further estimated using the average biomass per image W_X using the following equation.

The formula for broadleaf forests versus coniferous forests is as follows.

$$\overline{W_X} = 100 \times \frac{\rho}{A} \sum W_i, i = S, B, L, R, T1$$
(10)

where ρ is the vegetation density (plants per acre), A is the conversion factor, here, A = 666.67, and W_i is the biomass of each part of the vegetation.

The formula for shrubland is as follows.

$$\overline{W_{X}} = 100 \times W_{T2} \tag{11}$$

where W_{T2} is the whole plant biomass of a typical shrub.

The above sample points were subjected to regression modeling using Both-SRA and R^2 was used as an evaluation of the fitting results, which are shown in the following Table 6.

1.

Species	Biomass Model	Correlation Coefficient
	$W_s = 0.0573 \times \left(D^2 \times H\right)^{0.8657}$	0.97
Chinese cypress	$W_B = 0.0043 \times (D^2 \times H)^{1.1085}$	0.89
	$W_{L} = 0.0038 \times \left(D^{2} \times H\right)^{1.0385}$	0.84
	$W_{R} = 0.0485 \times \left(D^{2} \times H\right)^{0.6886}$	0.80
	$W_{s} = 0.0146 imes \left(D^{2} imes H ight)^{0.9835}$	0.993
Cunninghamia (taxus)	$W_{B} = 0.0243 \times \left(D^{2} \times H\right)^{0.7359}$	0.993
	$W_L = 0.0949 imes \left(D^2 imes H ight)^{0.4795}$	0.982
	$W_{ m R}=0.0102 imes \left({ m D}^2 imes { m H} ight)^{0.8745}$	0.975
	$W_s = 0.006 \times (D^2 \times H)^{1.098}$	0.995
poplar tree	$W_B = 0.001 \times (D^2 \times H)^{1.157}$	0.984
	$W_{ m L}=0.012 imes \left({ m D}^2 imes { m H} ight)^{0.685}$	0.955
	$W_{R} = 0.083 imes \left(D^{2} imes H ight)^{0.636}$	0.915
	$W_s = \exp(-2.8319 + 0.8379 \times \ln(D^2 \times H))$	0.9996
lanch tree (Dinus lanis)	$W_{\rm B} = \exp(-3.9021 + 0.8822 \times \ln(D^2 \times H))$	0.9015
farch tree (Finus farix)	$W_{L} = \exp(-4.0174 + 0.7659 \times \ln(D^{2} \times H))$	0.9007
	$W_{R} = \exp(-3.6497 + 0.8247 \times \ln(D^{2} \times H))$	0.9994
Other hard broad	$W_{T1} = 0.07112 \times (D^2 \times H)^{0.910358078}$	/
Other typical shrubs	$W_{T2} = 0.157759 \times (V_c)^{0.881}$	0.932

Table 5. Biomass model of typical tree species in Chongli District.

where D is the diameter at breast height, H is the tree height, V_c is the volume of plant crown per unit area, and W_s , W_B , W_L , W_R , W_{T1} , and W_{T2} are the trunk biomass, branch biomass, leaf biomass, below-ground biomass, above-ground biomass, and whole-plant biomass, respectively.

Table 6. Stepwise regression model fitting results.

Classes	R ²	Independent Variable	Coefficient
		constant	-2.014
		Cont(Red)	0.517
		SR2	6.029
		Mean(SWIR2)	2.465
shrub	0.811	Mean(Green)	-0.610
		Cor(N_NIR)	0.001
		Cor(VRE3)	0.001
		Mean(N_NIR)	0.133
		Sec(SWIR2)	1.258
		constant	220.571
1 11 (()	0.254	Entr(VRE2)	89.329
broadleaf forest	0.356	WV	-0.087
		Sec(VRE1)	213.875
coniferous forest	0.223	constant	-414.570

The inverse equation for shrubland biomass is as follows.

$$B_{S} = \begin{bmatrix} -2.014 + 0.517 \times \text{Cont}(\text{Red}) + 6.029 \times \text{SR2} \\ +2.456 \times \text{Mean}(\text{SWIR2}) - 0.610 \times \text{Mean}(\text{Green}) + 0.001 \times \text{Cor}(\text{N}_{\text{NIR}}) \\ +0.001 \times \text{Cor}(\text{VRE3}) + 0.133 \times \text{Mean}(\text{N}_{\text{NIR}}) + 1.258 \times \text{Sec}(\text{SWIR2}) \end{bmatrix}$$
(12)

The broadleaf forest biomass inverse equation is as follows.

$$B_{B} = 220.571 + 89.329 \times Entr(VRE2) - 0.087 \times WV + 213.875 \times Sec(VRE1)$$
(13)

The inverse equation for coniferous forest biomass is as follows:

$$B_{\rm C} = -414.570 + 406.537 \times \text{Entr}(\text{Green}) - 123.457 \times \text{Var}(\text{CA})$$
(14)

Combining the vegetation species obtained from the previous classification, the biomass inversion equation for any image element in the study area is as follows.

$$B = \begin{cases} B_{s}, \text{ low wood} \\ B_{B}, \text{broad} - \text{leaved forest} \\ B_{C}, \text{coniferous forest} \\ 0, \text{ other} \end{cases}$$
(15)

4. Results

For the three vegetation types, the accuracy of the inversion model was evaluated by the coefficient of determination R_{yz}^2 , the root mean square error (RMSE_{yz}), and the mean relative error (MRE) using the last 20% of the data. The final evaluation results are shown in Figure 5 and Table 7.



Figure 5. Verification of inversion accuracy.

Table 7. Inversion accuracy o	f different types of	vegetation
-------------------------------	----------------------	------------

Classes	R_{yz}^2	RMSE _{yz}	MRE
shrub	0.28	1.20	16.0%
broadleaf forest	0.49	139.13	42.8%
coniferous forest	0.14	315.63	65.0%

Based on Figure 5 and Table 7, it is evident that the multiple linear regression model constructed using the stepwise approach exhibits a high level of accuracy when estimating the biomass of broadleaf forests. However, it demonstrates poor estimation capabilities for both broadleaf and coniferous forests. Moreover, the model tends to underestimate biomass in high-biomass woodlands and overestimate it in low-biomass woodlands. In general, the relationship between remote sensing images, texture characteristics, and other data established by the multiple linear regression model does not achieve a particularly precise estimation of biomass. This discrepancy may arise from the inherent challenge faced by the multiple linear regression model in capturing the nonlinear relationship between biomass and these variables, which aligns with the findings of previous studies [45].

5. Discussion

Vegetation classification refers to the categorization of plant species and their distribution in a specific area. Different types of vegetation have varying characteristics, including plant density, moisture content, and flammability. These classifications are typically based on factors like plant species, growth form (e.g., grass, shrub, tree), and ecological characteristics. Vegetation classification is fundamental to understanding the potential fuel sources for wildfires [46].

Biomass represents the total mass of living or dead organic material in a specific ecosystem. In the context of wildfires, the biomass of vegetation serves as the primary fuel for the fire [47]. The amount and type of biomass in an area directly impacts the intensity and spread of wildfires. Denser and more abundant vegetation often results in higher biomass, which can lead to more intense and destructive wildfires when ignited.

Vegetation remote sensing classification and biomass estimation play a vital role in the development of forest fire prevention projects and the sustainable management of forest resources [48]. This study focuses on conducting comprehensive research on vegetation remote sensing classification and biomass inversion in the Chongli District of Zhangjiakou City.

The results demonstrate that utilizing the K-Means unsupervised classification method, based on Sentinel-2B satellite data, enables the initial categorization of vegetation in the Chongli area into shrub forests, broad-leaf forests, and coniferous forests. The accuracy of classification is higher for shrub forests and broad-leaf forests, while it is relatively lower for coniferous forests. Consequently, the K-Means algorithm can be applied to achieve a preliminary classification of shrub forests and broadleaf forests in Chongli.

In the biomass inversion study of the three vegetation types, it is observed that vegetation biomass shows a strong correlation with the texture characteristics of remote sensing images. However, the correlation with original band data, vegetation-related band indices, and topographic factors is not as pronounced. During the accuracy test of the inversion model, all three vegetation types tend to be underestimated in high-biomass forests and overestimated in low-biomass forests, leading to substantial errors in biomass estimation for a small number of samples. However, further experiments indicate that, as the estimated area expands, the discrepancy between the sum of biomass measurement decreases. The overall accuracy of the biomass model for shrub, broad-leaf, and coniferous forests reaches 93.58%, 89.56%, and 97.53%, respectively, meeting the requirements for forest surveys in the most ideal scenario. Nevertheless, future research should concentrate on constructing more accurate linear or nonlinear biomass estimation.

It is noteworthy that this study did not consider the influence of climate, depression, vegetation abundance, soil water content, and other factors on biomass in the study area during the biomass inversion study. The selection of models was also confined to multiple linear regression models, without incorporating nonlinear regression models. Therefore, an important area of future research lies in integrating the aforementioned factors and employing scientific and rational modeling techniques to construct a more precise model.

Nature displays a close correlation between biomass and fire, which is particularly evident in forest ecosystems [49]. The interconnection between biomass and fire is complex and influenced by various factors, holding significant implications for ecosystem stability and sustainability. Fires can cause direct and severe harm to organisms, leading to mass mortality [50]. For instance, forest fires can consume vegetation, destroy animal habitats, and drive wildlife to the brink of extinction [47]. In extreme cases, high-intensity fires can completely incinerate biomass, with ecosystems requiring an extensive period of recovery [51]. On the other hand, moderate fires can foster the growth and renewal of biomass. Within particular ecosystems, fire serves as a natural successional process that eliminates accumulated dry matter and plant residues, thus facilitating opportunities for growth [52]. Notably, certain plant seeds require the presence of fire to stimulate germination [53], while specific plants and animals have developed selective adaptations to take advantage of fire's benefits.

Observing and monitoring changes in biomass can help predict fire risk [54]. For example, if biomass remains high in an area for an extended period, the site may be at a higher risk of fire. Conversely, if biomass continues to decline, it may indicate that the

vegetation has been affected by disease, drought, or other unusual factors, reducing the likelihood of fire spread.

The biomass inversion model presented in this study has the capacity to accurately depict variations in biomass. Consequently, it facilitates the examination of pertinent data for discerning the patterns of alterations in fire risk. Moreover, it enables swift and effective spatial identification through remote sensing, thereby aiding scientists in proactively managing biomass fluctuations to regulate fire severity. Additionally, this model can serve as a foundation for developing standardized fire treatments rooted in prioritized eco-efficiency.

6. Conclusions

This paper establishes a biomass inversion model to obtain information on vegetation species distribution in the Chongli District, using remote sensing images and machine learning. This is significant in guiding forest fire prevention projects and the rational development of forest resources. The study classifies vegetation into three categories: shrub forest, broadleaf forest, and coniferous forest, based on the K-Means unsupervised classification method of Sentinel-2B satellite data. The classification accuracy of shrub and broadleaf forests is higher than that of coniferous forests. The study finds that vegetation biomass is highly correlated with texture features of remote sensing images, but not closely related to original band data or the vegetation-related band index. The correlation with the terrain factor is not significant either. In the accuracy test of the inversion model, the difference between the sum of biomass estimation and measurement decreases with the expansion of the estimated area. However, further research is needed to construct a more accurate linear or nonlinear biomass estimation model. Future research should consider the effects of climate, depressions, vegetation richness, and soil moisture content on biomass in the study area. In conclusion, biomass and vegetation classification are critical to the accuracy of fire behavior simulation, which is important for forest fire prevention and control on a large scale.

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

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

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

References

- 1. Bowman, D.M.; Kolden, C.A.; Abatzoglou, J.T.; Johnston, F.H.; van der Werf, G.R.; Flannigan, M. Vegetation fires in the Anthropocene. *Nat. Rev. Earth Environ.* 2020, *1*, 500–515. [CrossRef]
- Ganteaume, A.; Camia, A.; Jappiot, M.; San-Miguel-Ayanz, J.; Long-Fournel, M.; Lampin, C. A review of the main driving factors of forest fire ignition over Europe. *Environ. Manag.* 2013, *51*, 651–662. [CrossRef] [PubMed]
- Shao, Y.; Wang, Z.; Feng, Z.; Sun, L.; Yang, X.; Zheng, J.; Ma, T. Assessment of China's forest fire occurrence with deep learning, geographic information and multisource data. J. For. Res. 2023, 34, 963–976. [CrossRef]
- Shao, Y.; Feng, Z.; Cao, M.; Wang, W.; Sun, L.; Yang, X.; Ma, T.; Guo, Z.; Fahad, S.; Liu, X.; et al. An Ensemble Model for Forest Fire Occurrence Mapping in China. Forests 2023, 14, 704. [CrossRef]
- 5. Zhang, L.; Li, J.; Zhang, F. An Efficient Forest Fire Target Detection Model Based on Improved YOLOv5. *Fire* **2023**, *6*, 291. [CrossRef]
- Shi, C.; Zhang, F. A Forest Fire Susceptibility Modeling Approach Based on Integration Machine Learning Algorithm. *Forests* 2023, 14, 1506. [CrossRef]
- Shao, Y.; Feng, Z.; Sun, L.; Yang, X.; Li, Y.; Xu, B.; Chen, Y. Mapping China's Forest Fire Risks with Machine Learning. *Forests* 2022, 13, 856. [CrossRef]
- 8. Xu, Y.; Sun, Y.; Zhang, F.; Jiang, H. Modeling Fire Boundary Formation Based on Machine Learning in Liangshan, China. *Forests* **2023**, *14*, 1458. [CrossRef]

- Shao, Y.; Fan, G.; Feng, Z.; Sun, L.; Yang, X.; Ma, T.; Li, X.; Fu, H.; Wang, A. Prediction of forest fire occurrence in China under climate change scenarios. J. For. Res. 2023, 34, 1217–1228. [CrossRef]
- 10. Xie, Y.; Sha, Z.; Yu, M. Remote sensing imagery in vegetation mapping: A review. J. Plant Ecol. 2008, 1, 9–23. [CrossRef]
- 11. Yang, C.; Wu, G.; Li, Q.; Wang, J.; Qu, L.; Ding, K. Research Progress on Remote Sensing Classification of Vegetation. Geogr. Geo. *Inf. Sci.* **2018**, *10*, 191–196.
- 12. Gao, D.; Liu, Y.; Hu, B.; Wang, L.; Chen, W.; Chen, Y.; He, T. Time Synchronization based on Cross-Technology Communication for IoT Networks. *IEEE Internet Things J.* 2023, 19753–19764. [CrossRef]
- 13. Gao, D.; Wang, H.; Guo, X.; Wang, L.; Gui, G.; Wang, W.; Yin, Z.; Wang, S.; Liu, Y.; He, T. Federated learning based on CTC for heterogeneous internet of things. *IEEE Internet Things J.* 2023, 10, 22673–22685. [CrossRef]
- 14. Gao, D.; Wang, L.; Hu, B. Spectrum efficient communication for heterogeneous IoT networks. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 3945–3955. [CrossRef]
- 15. Song, W.; Zhang, X.; Chen, Y.; Xu, H.; Wang, L.; Wang, Y. Dimensionality Reduction and Research of Hyperspectral Remote Sensing Images Based on Manifold Learning. *arXiv* **2024**, arXiv:1812.09530v1.
- 16. Buddenbaum, H.; Schlerf, M.; Hill, J. Classification of coniferous tree species and age classes using hyperspectral data and geostatistical methods. *Int. J. Remote Sens.* **2005**, *26*, 5453–5465. [CrossRef]
- 17. Zhu, H.; Huang, Y.; Li, Y.; Yu, F.; Tu, T.; Wang, W.; Zang, Y.; Li, J.; Luo, Y. Diversity of Plant Community in Flood Land of Henan Section of the Lower Yellow River based on Unmanned Aerial Vehicle Remote Sensing. *Wetl. Sci.* **2021**, *19*, 17–26.
- 18. Li, C.; Wang, J.; Wu, G.; Li, Q. Classification of agricultural plants based on leaf spectral features. J. Shenzhen Univ. Sci. Eng. 2018, 35, 307–315. [CrossRef]
- 19. Xin, D.; Huang, X.; Hongga, I.; Shen, L. Research on classification of plant community using projection pursuit learning network algorithm on high resolution remote sensing images. *J. Geo-Inf. Sci.* **2016**, *18*, 124–132.
- 20. Shi, L.; Liu, S. Methods of estimating forest biomass: A review. Biomass Vol. Estim. Valorization Energy 2017, 10, 65733.
- Li, X.; Wang, X.; Gao, Y.; Wu, J.; Cheng, R.; Ren, D.; Bao, Q.; Yun, T.; Wu, Z.; Xie, G. Comparison of Different Important Predictors and Models for Estimating Large-Scale Biomass of Rubber Plantations in Hainan Island, China. *Remote Sens.* 2023, 15, 3447. [CrossRef]
- 22. Wu, H.; Xu, H. A review of the application of sampling techniques in forest biomass inventory. J. Southwest For. Univ.(Nat. Sci.) 2021, 41, 183–188.
- Pelletier, F.; Cardille, J.A.; Wulder, M.A.; White, J.C.; Hermosilla, T. Inter-and intra-year forest change detection and monitoring of aboveground biomass dynamics using Sentinel-2 and Landsat. *Remote Sens. Environ.* 2024, 301, 113931. [CrossRef]
- 24. Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Li, G.; Moran, E. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *Int. J. Digit. Earth* **2016**, *9*, 63–105. [CrossRef]
- 25. Guo, Q.; Su, Y.; Hu, T. LiDAR Principles, Processing and Applications in Forest Ecology; Academic Press: Cambridge, MA, USA, 2023.
- Yali, Z.; Huifang, Z.; Jinglu, Z. Establishment of the Model for Estimating Aboveground Biomass of Populus euphratica Based on UAV Remote Sensing. *For. Resour. Wanagement* 2019, 80. [CrossRef]
- 27. Ye, Q.; Yu, S.; Liu, J.; Zhao, Q.; Zhao, Z. Aboveground biomass estimation of black locust planted forests with aspect variable using machine learning regression algorithms. *Ecol. Indic.* **2021**, *129*, 107948. [CrossRef]
- Guo, C.; Chen, Z.; Zhang, Z. Research on Remote Sensing Estimation of Forage Above-ground Biomass Based on Optimal Model Selection. *Acta Agrestia Sin.* 2021, 29, 946.
- 29. Yanhui, L.; Xiaoyu, Y.; Nisha, B.; Xiaowei, G. Estimating biomass of reclaimed vegetation in prairie mining area: Inversion method based on Worldview-3 and Sentinel-1 SAR data. *Earth Sci. Front.* **2021**, *28*, 219.
- Li, T.; Feng, H.; Zhu, B.; Fan, Y.; Jin, L.; Cheng, Q.; Li, Q. Winter Wheat Biomass Inversion Based on UAV Hyperspectral and Digital Image Data. *Mod. Agric. Sci. Technol.* 2020, 1–5.
- Miao, X.; Li, J.; Mu, Y.; He, C.; Ma, Y.; Chen, J.; Wei, W.; Gao, D. Time Series Forest Fire Prediction Based on Improved Transformer. Forests 2023, 14, 1596. [CrossRef]
- Zhang, G.; Zhu, S.; Zhang, N.; Zhang, G.; Xu, Y. Downscaling hourly air temperature of WRF simulations over complex topography: A case study of Chongli District in Hebei Province, China. J. Geophys. Res. Atmos. 2022, 127, e2021JD035542. [CrossRef]
- Liang, X.; Yang, T.; Niu, J.; Zhang, L.; Wang, D.; Huang, J.; Yang, Z.; Berndtsson, R. Quality Assessment and Rehabilitation of Mountain Forest in the Chongli Winter Olympic Games Area, China. *Forests* 2022, 13, 783. [CrossRef]
- 34. Segarra, J.; Buchaillot, M.L.; Araus, J.L.; Kefauver, S.C. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. *Agronomy* **2020**, *10*, 641. [CrossRef]
- 35. Zhou, Q. Digital elevation model and digital surface model. In *International Encyclopedia of Geography: People, the Earth, Environment and Technology;* Wiley: New York, NY, USA, 2017; pp. 1–17.
- 36. Pettorelli, N. The Normalized Difference Vegetation Index; Oxford University Press: New York, NY, USA, 2013.
- 37. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]
- Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 1996, 55, 95–107. [CrossRef]

- Sims, D.A.; Gamon, J.A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* 2002, *81*, 337–354. [CrossRef]
- 40. Blackburn, G.A. Quantifying chlorophylls and caroteniods at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sens. Environ.* **1998**, *66*, 273–285. [CrossRef]
- 41. Le Maire, G.; François, C.; Dufrene, E. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sens. Environ.* **2004**, *89*, 1–28. [CrossRef]
- 42. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man. Cybern.* 1973, 610–621. [CrossRef]
- Sarker, L.R.; Nichol, J.E. Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sens. Environ.* 2011, 115, 968–977. [CrossRef]
- 44. Zhou, X. Urban Vegetation Classification and Biomass Inversion Based on Sentinel-2A Data in Xuzhou; China University of Mining and Technology: Xuzhou, China, 2019.
- 45. Xu, Z.; Li, Y.; Li, M.; Li, C.; Wang, L. Forest biomass retrieval based on Sentinel-1A and Landsat 8 image. J. Cent. South. Univ. For. Technol. 2020, 40, 147–155.
- Weise, D.R.; Wright, C.S. Wildland fire emissions, carbon and climate: Characterizing wildland fuels. *For. Ecol. Manag.* 2014, 317, 26–40. [CrossRef]
- 47. Gajendiran, K.; Kandasamy, S.; Narayanan, M. Influences of wildfire on the forest ecosystem and climate change: A comprehensive study. *Environ. Res.* 2023, 240, 117537. [CrossRef]
- Nandasena, W.; Brabyn, L.; Serrao-Neumann, S. Using remote sensing for sustainable forest management in developing countries. In *The Palgrave Handbook of Global Sustainability*; Springer: Berlin, Germany, 2023; pp. 487–508.
- 49. Martins, F.d.S.R.V.; Xaud, H.A.M.; dos Santos, J.R.; Galvão, L.S. Effects of fire on above-ground forest biomass in the northern Brazilian Amazon. *J. Trop. Ecol.* 2012, 28, 591–601. [CrossRef]
- Whiterod, N.S.; Lintermans, M.; Cramp, R.L.; Franklin, C.E.; Kennard, M.J.; McCormack, R.; Pearce, L.; Raadik, T.A.; Ward, M.; Zukowski, S. The impact of the 2019–2020 Australian wildfires on aquatic systems. In *Australia's Megafires: Biodiversity Impacts* and Lessons from 2019–2020; CSIRO Publishing: Clayton, Australia, 2023; p. 59.
- 51. Nolan, R.H.; Collins, L.; Leigh, A.; Ooi, M.K.; Curran, T.J.; Fairman, T.A.; Resco de Dios, V.; Bradstock, R. Limits to post-fire vegetation recovery under climate change. *Plant Cell Environ*. **2021**, *44*, 3471–3489. [CrossRef] [PubMed]
- McLauchlan, K.K.; Higuera, P.E.; Miesel, J.; Rogers, B.M.; Schweitzer, J.; Shuman, J.K.; Tepley, A.J.; Varner, J.M.; Veblen, T.T.; Adalsteinsson, S.A. Fire as a fundamental ecological process: Research advances and frontiers. *J. Ecol.* 2020, 108, 2047–2069. [CrossRef]
- 53. Lamont, B.B.; He, T.; Yan, Z. Evolutionary history of fire—Stimulated resprouting, flowering, seed release and germination. *Biol. Rev.* **2019**, *94*, 903–928. [CrossRef]
- Kaduyu, I. Wildfire prediction and monitoring in the rangeland areas of Botswana: A case study of Kgalagadi district. Master's Thesis, Botswana University of Agriculture & Natural Resources, Gaborone, Botswana, 2023.

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.