Article

Canopy Height Integration for Precise Forest Aboveground Biomass Estimation in Natural Secondary Forests of Northeast China Using Gaofen-7 Stereo Satellite Data

Caixia Liu [1], Huabing Huang [2,3], Zhiyu Zhang [1], Wenyi Fan [4,5] and Di Wu [6,7,*]

1   International Research Center of Big Data for Sustainable Development Goals, State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; liucx@radi.ac.cn (C.L.); zhangzhiyu@aircas.ac.cn (Z.Z.)
2   School of Geospatial Engineering and Science, Sun Yat-Sen University, Guangzhou 510275, China; huanghb55@mail.sysu.edu.cn
3   Key Laboratory of Comprehensive Observation of Polar Environment (Sun Yat-Sen University), Ministry of Education, Zhuhai 519082, China
4   School of Forestry, Northeast Forestry University, Harbin 150040, China; fanwy@nefu.edu.cn
5   Key Laboratory of Sustainable Forest Ecosystem Management-Ministry of Education, Northeast Forestry University, Harbin 150040, China
6   School of Surveying and Mapping Engineering, Heilongjiang Institute of Technology, Harbin 150050, China
7   Heilongjiang Geomatics Centre of Ministry of Natural Resources, Harbin 150081, China
*   Correspondence: wud@hlsm.mnr.gov.cn

Abstract: Accurate estimates of forest aboveground biomass (AGB) are necessary for the accurate tracking of forest carbon stock. Gaofen-7 (GF-7) is the first civilian sub-meter three-dimensional (3D) mapping satellite from China. It is equipped with a laser altimeter system and a dual-line array stereoscopic mapping camera, which enables it to synchronously generate full-waveform LiDAR data and stereoscopic images. The bulk of existing research has examined how accurate GF-7 is for topographic measurements of bare land or canopy height. The measurement of forest aboveground biomass has not received as much attention as it deserves. This study aimed to assess the GF-7 stereo imaging capability, displayed as topographic features for aboveground biomass estimation in forests. The aboveground biomass model was constructed using the random forest machine learning technique, which was accomplished by combining the use of in situ field measurements, pairs of GF-7 stereo images, and the corresponding generated canopy height model (CHM). Findings showed that the biomass estimation model had an accuracy of $R^2 = 0.76$, RMSE = 7.94 t/ha, which was better than the inclusion of forest canopy height ($R^2 = 0.30$, RMSE = 21.02 t/ha). These results show that GF-7 has considerable application potential in gathering large-scale high-precision forest aboveground biomass using a restricted amount of field data.

Keywords: Gaofen-7 (GF-7); stereophotogrammetry; canopy height; aboveground biomass

1. Introduction

Forests play a crucial role in the terrestrial carbon sink system, acting as the largest carbon store among terrestrial ecosystems [1]. They absorb carbon dioxide from the atmosphere through photosynthesis, storing it in biomass (trees, plants) and soil, thus mitigating climate change [2]. The accurate estimation of forest biomass is crucial for ecological studies and global change research [3,4]. The advancements in remote sensing technologies, particularly LiDAR [5], multi-angle photogrammetry [6], and synthetic aperture radar

(SAR) [7], have significantly enhanced the capabilities of high-resolution mapping for forest aboveground biomass (AGB).

The evolution of spaceborne remote sensing for forest AGB monitoring began in the 1980s and 1990s, characterized by the emergence of improved sensors and platforms such as SPOT and RADARSAT, which provided enhanced vegetation monitoring capabilities compared with earlier satellites like Landsat [8–11]. As a result, spaceborne remote sensing has thus become crucial for accurately estimating forest AGB. In the 2000s, the launch of high-resolution commercial satellites like QuickBird and IKONOS brought about a revolution in sub-meter resolution imagery [6,12]. At the same time, advances in LiDAR technology led to a significant improvement in biomass estimation accuracy through the integration of optical, radar, and LiDAR data using sophisticated algorithms and machine learning [13,14]. Such modern remote sensing capabilities are essential for effective forest conservation, climate change studies, and management practices.

The application of advanced remote sensing technologies significantly enhances the measurement of forest three-dimensional structures, thereby improving the estimation of forest AGB. Satellite observation currently employs the following three primary techniques to achieve this: LiDAR technology [5,15], synthetic aperture radar (SAR) [7], and stereophotogrammetry [6]. One of the most significant advancements in this domain has been the development of spaceborne LiDAR systems, which can acquire terrain and height data for extensive forest areas at reduced costs. Historically, satellite-based measurements of forest canopy height were largely dependent on NASA's Geoscience Laser Altimeter System (GLAS) [16–18]. However, recent innovations, including the NASA Ice, Cloud, and land Elevation Satellite-2 (ICESat-2) LiDAR instrument and the Global Ecosystem Dynamics Investigation (GEDI) system, have markedly enhanced global forest monitoring capabilities [15,19]. The upcoming BIOMASS mission, scheduled for launch in 2025, exemplifies the integration of advanced P-band synthetic aperture radar (SAR) technology to monitor changes in forest AGB and production [20]. Stereophotogrammetry has also seen considerable improvements through several satellite missions [21]. Notable examples include the Chinese ZY-3 [22], the Japanese ALOS PRISM [23], SkySat satellites from Planet [24], and the Worldview series from Maxar [25]. These missions have enhanced forest monitoring by leveraging parallax range-dependent stereophotogrammetry. The launch of China's domestically developed Gaofen-7 (GF-7) satellite, which is equipped with advanced stereo imaging capabilities, presents new opportunities for forest parameter investigation [26]. With a superior spatial resolution of 0.8 m for panchromatic imagery compared with ZY-3's 2 m, GF-7 enables more precise assessments of forest structures and health [22,27].

Stereo imaging for forest height and biomass estimation presents distinct advantages over spaceborne LiDAR systems like GEDI and ICESat. While GEDI and ICESat offer precise measurements, they are costly to develop, launch, and operate. In contrast, stereo imaging provides similar data at a fraction of the cost, making forest monitoring more financially viable, especially for resource-constrained regions. Puliti et al. [28] demonstrated the feasibility of predicting aboveground biomass in Norway by leveraging spaceborne stereogrammetric digital surface models, specifically ArcticDEM data. Their findings likely showcased a high degree of accuracy in biomass estimation, highlighting the potential of stereogrammetric data for forest monitoring and management. Similarly, Zhang et al. [29] illustrated the accuracy of estimating developing stem volume using forest canopy height derived from ZY-3 stereoscopic data. While specific accuracy metrics may vary depending on the methodologies employed, both studies likely reported promising results, indicating the reliability and precision of using stereoscopic satellite data for forest parameter estimation. These findings underscore the importance and efficacy of spaceborne stere-

ogrammetric data in advancing our understanding of forest ecosystems and supporting sustainable resource management practices.

Additionally, stereo imaging achieves higher spatial resolutions, enabling detailed characterization of forest structure and better delineation of individual trees and canopy gaps [6,30]. This finer resolution facilitates more accurate estimates of forest height and biomass at a finer spatial resolution. Moreover, satellites with stereo imaging sensors can cover large areas in a single pass, providing comprehensive coverage of forested landscapes, unlike the narrower swath widths of spaceborne LiDAR systems. The temporal consistency of stereo imaging also enables continuous monitoring of forest dynamics over time, crucial for accurately tracking changes. Overall, stereo imaging emerges as a cost-effective, high-resolution, and globally accessible solution for monitoring and managing forest ecosystems, complementing the capabilities of spaceborne LiDAR systems.

The GF-7 satellite's high spatial resolution enables the creation of detailed three-dimensional (3D) models of forested areas, providing valuable insights into forest structure at a level of detail previously unseen. Although initially intended for other applications, such as terrestrial elevation modeling [31] and urban planning [32–34], the rich information captured by GF-7 imagery presents an opportunity to extract relevant forest structural parameters. According to Du et al. [35], there has been a significant improvement in the accuracy of canopy height estimation using stereo imagery from the GF-7 satellite compared with data obtained from airborne laser scanning (ALS). Ni et al. [27] presented a technique to accurately extract forest height using GF-7 very high-resolution stereoscopic data, demonstrating the possibility for efficient regional and worldwide forest height assessment without extensive fieldwork. The high spatial resolution of the GF-7 satellite makes it easier to derive specific forest heights, which improves large-scale forest structures.

Despite these advancements, the potential of canopy height models (CHMs) derived from GF-7 in conjunction with spectral data for biomass estimation remains largely unexamined. While existing research has predominantly focused on evaluating the accuracy of GF-7 for topographic measurements on bare land or canopy height, there has been insufficient attention to its application in estimating forest aboveground biomass. Therefore, this study aims to address the following questions: (1) How effectively can GF-7 stereo imaging capture topographic features for the estimation of aboveground biomass in forests? (2) How can an aboveground biomass model be constructed using the random forest machine learning technique by integrating in situ field measurements, GF-7 stereo images, and the corresponding generated digital surface model (DSM)? (3) What is the accuracy of the biomass estimation model, in terms of R^2 and RMSE, and how does GF-7 demonstrate its potential for gathering large-scale high-precision forest aboveground biomass data? These questions aim to determine the significant promise of GF-7 for large-scale and precise forest AGB estimation, thereby enhancing our capabilities in forest monitoring and management.

2. Methods and Data

2.1. Study Area

The forest in this study is over 500 km^2 in size, with dimensions of roughly 21.66 km east to west and 24.14 km north to south. It is in Shangzhi City, Heilongjiang Province, China (127°20′ to 127°50′E, 45°05′ to 45°30′N) (Figure 1). With a frost-free period of approximately 125 days and an average temperature of 2.4 °C, the area displays traits of a subarctic monsoon climate. Rainfall is mostly concentrated from May to September, with an average of 700 mm per year.
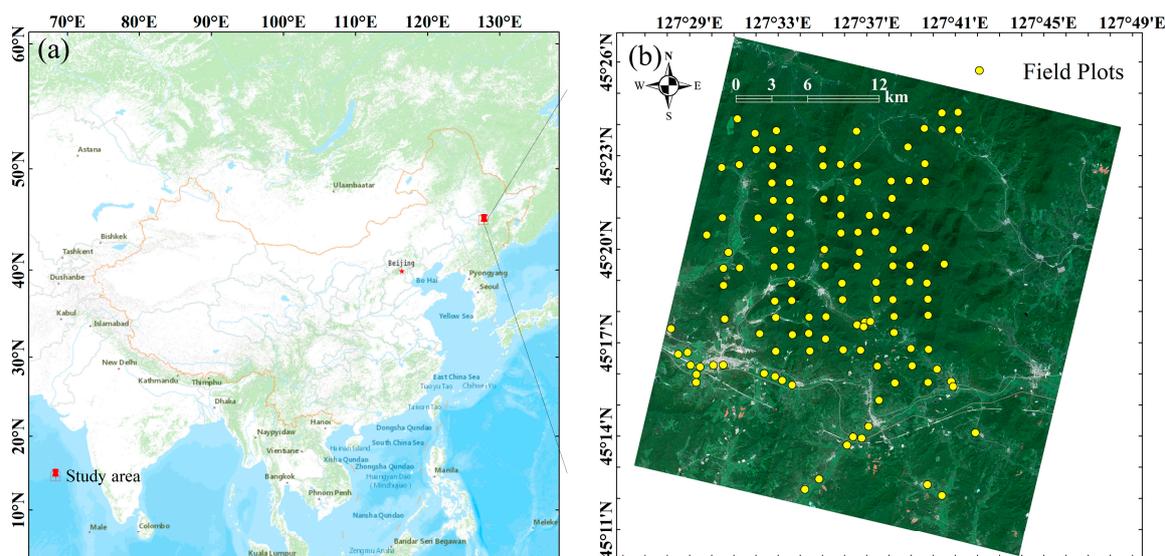
Figure 1. Location map of the study area (Shangzhi, Heilongjiang, China). (**a**) The location of the study area; (**b**) field plots over the GF-7 multispectral image on 20 August 2020 (R—red, G—green, B—blue).

The study area's forest resources make up a significant portion of northeastern China's forest resources, making it an important ecological resource area. The mixed coniferous and broad-leaved forests that make up most of the vegetation belt are home to pine trees. Spruce, larch, walnut, oak, elm, yew, birch, and poplar are among the major tree species. The forest's landform is part of a low-lying hilly region. The land has an average elevation of 300 m and rises gradually from the south to the north. Maoer Mountain, at 805 m above sea level, is the highest peak.

2.2. Field Measurements

Within this study region's GF-7 image coverage area, 74 plots in total were created in the year of 2019. The forest type and structure in the research region remained steady, even though the plot data and GF-7 satellite data were gathered one year apart. As a result, the change in the forest AGB during this time was not seen as significant. All the plots were randomly sampled as closely as possible to reflect the true distribution of the forest given the remote access to the field plots. The plots had a 0.06-hectare area and were square in shape. The GPS coordinates and relative positions within the stand were noted. A Vertex IV device was utilized to measure the height of trees in each plot that had a diameter at breast height (DBH) greater than 5 cm. For live standing trees, measurements of the trees' species, DBH, crown width, and height were made. The AGB of each plot was determined by adding the values of the relevant species-specific allometric equations, which were derived using tree height and DBH. The DBH-based allometric equations for calculating AGB of main tree species are listed in Table S2. AGB was measured in the plot with an average value of roughly 128.2 t/ha, a minimum of 29.4 t/ha, and a maximum of 276.5 t/ha.

2.3. Data and Methods

2.3.1. GF-7 Stereo Images and CHM Retrieval

GF-7, China's first sub-meter high-resolution Earth observation and remote sensing satellite, is meant to be used for industrial monitoring, natural resource monitoring, and land surveying. It was launched on 3 December 2019, from Taiyuan, China, and it was put into a 500 km sun-synchronous orbit, with a temporal resolution of 59 days. Unlike traditional optical remote sensing satellites, the GF-7 satellite contains double-line cameras

(DLCs) and could offer elevation data via stereo mapping technology. The GF-7 satellite images used in this study were provided by the Chinese Land Satellite Remote Sensing Application Center (http://www.lasac.cn/ access on 12 April 2024). To explore the relationship between DSM and forest canopy structure in different seasonal stereo pairs, we selected two periods of images in August and November, 2020, respectively. Three images make up each GF-7 imaging scene: two panchromatic images (with viewing angles of $-5°$ in the backward direction and $+26°$ in the forward direction) and one multispectral image containing bands of blue, green, red, and near-infrared. Furthermore, the four-band backward multispectral imagery has a spatial resolution of 2.6 m, and the GF-7 forward panchromatic and backward panchromatic images have respective spatial resolutions of 0.65 m and 0.8 m (see Table 1).

Table 1. Basic information of GF-7 DLC imagery.

GF-7 DLC Imagery	Name	Spectral Range (nm)	Spatial Resolution (m)	Viewing Direction	Viewing Angle (°)
Multispectral image	MUX	Blue: 450–520 Green: 520–590 Red: 630–690 NIR: 770–890	2.60	Backward	−5.00
Panchromatic image	FWD BWD	450–900	0.80 0.65	Forward Backward	26.00 −5.00

The preprocessing steps for each GF-7 image include radiometric correction, atmospheric correction, orthorectification, and BRDF correction. Initially, the raw images, provided as digital numbers (DNs), are converted to physical quantities of radiance or reflectance using the sensor's calibration coefficients, with gain and offset values detailed in the accompanying table. Atmospheric correction is performed to remove effects such as aerosol and molecular scattering, allowing us to obtain surface reflectance using the FLAASH module in ENVI 4.8 software. Orthorectification addresses geometric errors due to terrain effects or sensor distortions by correcting the images to a uniform geographic coordinate system through the RPC orthorectification module in ENVI. Lastly, to mitigate BRDF effects caused by variations in solar elevation, azimuth, and observation angles, we normalize the reflectance to standardized observation conditions using the Ross–Li model, incorporating NASA's MODIS MCD43A1 data in SNAP software. We will clarify in the manuscript that both GF-7 images underwent these preprocessing steps, as they are critical for enhancing the accuracy of model predictions.

Using the stereo image product in the rational polynomial coefficients (RPCs) file, which provides the image coordinates and the ground field coordinates of the conversion relationship, we used the image point with the same name using the matching method. The RPC model was then corrected using ground control points (10–15 points/scene, this study), effectively removing systematic errors from the model. Finally, we generated the epipolar image by acquiring feature point and performing intensive matching using the semi-global matching (SGM) algorithm, resulting in the production of the digital surface model (DSM). Forest heights are then calculated as the difference between the stereoscopic DSMs and the digital terrain model (DTM), which is a 1:10,000 digital elevation model acquired by airborne LiDAR and is adequate for the requirements of our research. The retrieval procedure of CHM can be referenced in Figure 2a.
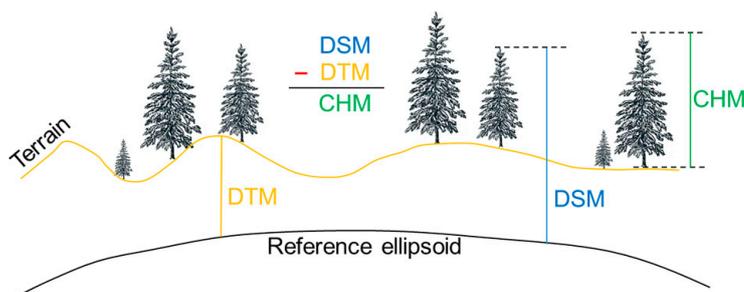
(a) Diagram of DSM and CHM
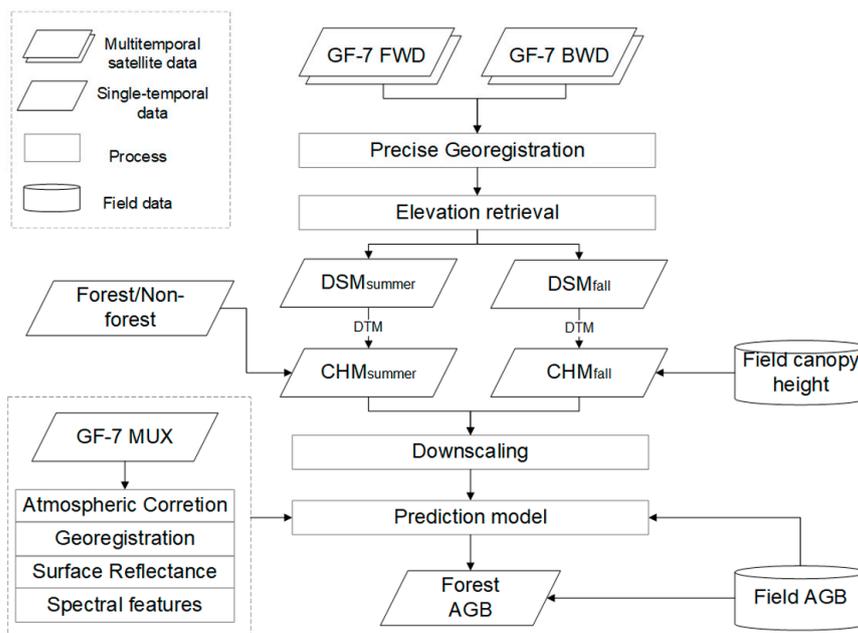
(b) The general structure of this study

Figure 2. The procedure for calculating forest canopy height and biomass from GF-7 stereoscopic imagery.

2.3.2. Random Forest Model for AGB Estimation

To investigate whether DSM or CHM can be capitalized on to improve mixed forest AGB retrieval, we mapped regional AGB maps in 2020 with four competing scenarios that varied the incorporation of DSM or CHM as the model input, as illustrated in Table 2.

Table 2. Overview of scenarios for AGB mapping using different inputs from GF-7 data.

Scenario	Description	Inputs
S1	Traditional prediction algorithm for AGB mapping using spectral features from GF-7 MUX images obtained during the growing season.	Original GF-7 spectral bands + 33 multispectral indices from GF-7 MUX bands (see Supplement Table S1).
S2	Same as S1, but incorporates CHM derived from digital terrain model (DTM) and GF-7 stereo image pairs in August.	Inputs from S1 + CHM from DTM and GF-7 stereo image pairs (August).

Table 2. *Cont.*

Scenario	Description	Inputs
S3	Same as S1, includes CHM derived from DTM and GF-7 stereo image pairs in November.	Inputs from S1 + CHM from DTM and GF-7 stereo image pairs (November).
S4	Uses the full set of inputs from S1 and different digital surface model (DSM) derived from GF-7 stereo image pairs in August and November.	Inputs from S1 + DSM from GF-7 stereo image pairs (August and November).

We selected the random forest (RF) regression model [36] implemented in a Matlab environment to create all AGB maps. Compared with conventional regression algorithms, the RF model possesses several advantages that make it well suited for our study. Other than being computationally efficient and sophisticated at handling a high dimensionality of input data, the RF demonstrates high prediction accuracy across a wide range of forest ecosystems [13,16]. All the prediction algorithms (S1–S4) were trained and validated with the same reference data. In addition to reserving seven independent samples for AGB model comparison, all remaining field plots were utilized for model development. The prediction procedure is briefly illustrated in Figure 2b. When building the model, the dataset was randomly divided into two subsets, with 90% allocated for training and 10% for model testing. We set the number of decision trees to 500 to obtain an unbiased estimate of the generalization error and used the default number of variables to be tested for each split (i.e., the square root of the number of input features) for all scenarios. The prediction results were evaluated using the coefficient of determination (R^2) for the relationship between field-measured and predicted AGB, and the root mean square error (RMSE). This process was repeated 50 times for each input scenario, with sample indexing randomized each time to ensure that the division between training and testing samples was not influenced by the original order of samples. We recorded R^2 and RMSE values during each iteration, selecting the model with the highest R^2 and the lowest RMSE as the final inversion model. It should be noted that the feature values of the plots, including spectral indices and canopy height, were obtained by calculating the mean of all pixels covering the plots based on the geographic coordinates of the plot's center.

The flowchart in Figure 2 presents a detailed methodology for estimating forest aboveground biomass (AGB) using digital surface model (DSM) and canopy height model (CHM) data. The process begins with the collection of multitemporal satellite data and field data, both of which are critical for ensuring accurate analysis. It encompasses two primary stages: GF-7 forward (FWD) and GF-7 backward (BWD), with an emphasis on precise geo-registration to facilitate effective elevation retrieval. This procedure produces seasonal DSM and CHM products for both summer and fall, while field canopy height measurements are utilized to validate the remote sensing data. Subsequently, the data undergoes downscaling to align with the resolution of MUX imagery. Finally, this refined dataset informs the development of a prediction model aimed at estimating forest AGB. It should be noted that the CHM was downscaled to 2.6 m to ensure it matches the resolution of the MUX images, facilitating subsequent analyses.

3. Results

3.1. Forest Height Prediction

The August surface model and canopy height model displayed higher topographic characteristics and more accurate estimations of forest height. Figure 3 displays the DSM and CHM for August and November. The DSM for August displays more topographic characteristics, such as ridges, valleys, and hilltops, than the DSM for November. We can also see that the ground elevation is higher in the north and lower in the south. August CHM better captures the distribution of forest height. The scatter plot in Figure 4 illustrates the relationship between canopy heights predicted using a canopy height model and field-measured heights for two different time points: August and November. For August, the regression analysis yielded an R^2 value of 0.52 and an RMSE of 2.14 m, indicating that 51.8% of the variance in canopy heights predicted by the August canopy height model is explained by the field measurements. The August regression line closely follows the 1:1 line, suggesting a reasonable fit with a slight underestimation at higher tree heights. In contrast, the November data exhibited a lower R^2 value of 0.45 and a higher RMSE of 2.96 m, showing that 44.9% of the variance in predictions from the November canopy height model is explained by the field data. The November regression line showed more deviation from the 1:1 line, particularly underestimating tree heights more significantly.
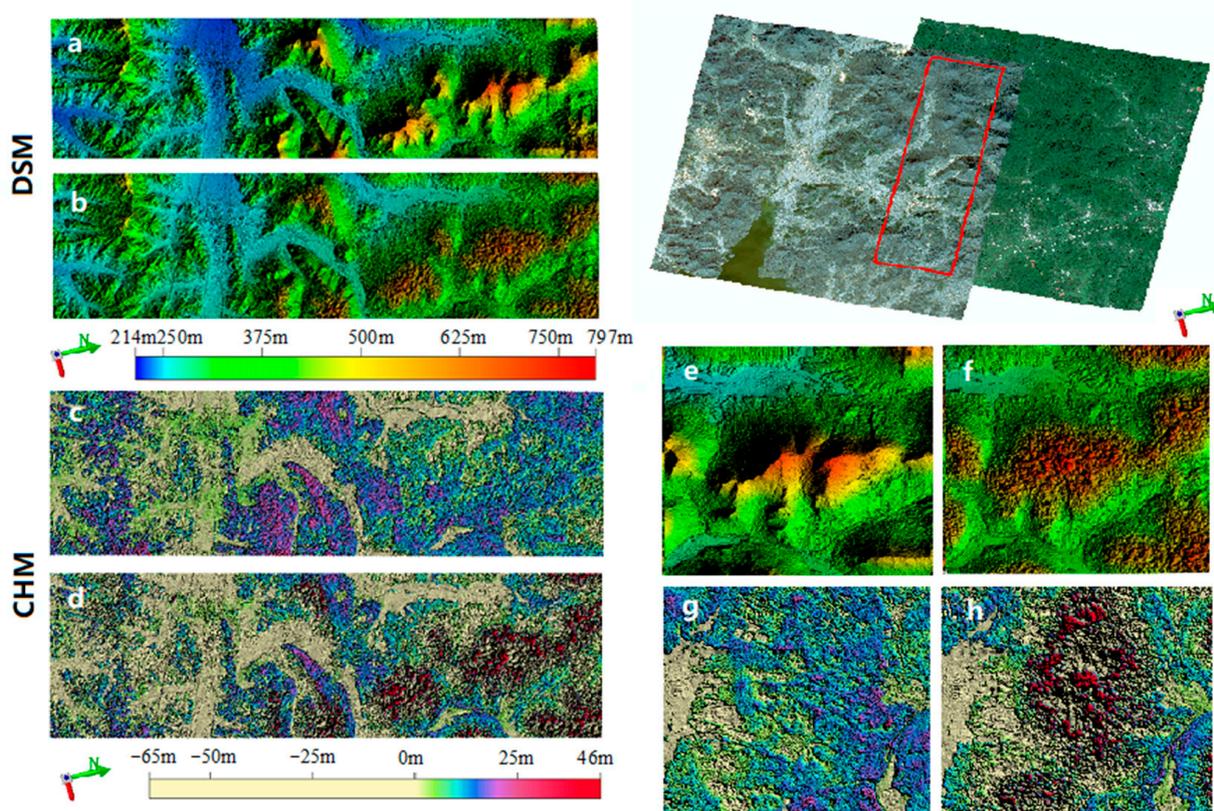

Figure 3. The August and November DSM and CHM. This figure shows only the DSM and CHM for the common regions between August and November, highlighted by the read box. (**a**,**b**) DSMs for August and November, respectively, and (**e**,**f**) show the larger detail plots in the red boxes. (**c**,**d**) CHMs for August and November, respectively, and (**g**,**h**) show the larger detail plots in the red boxes.
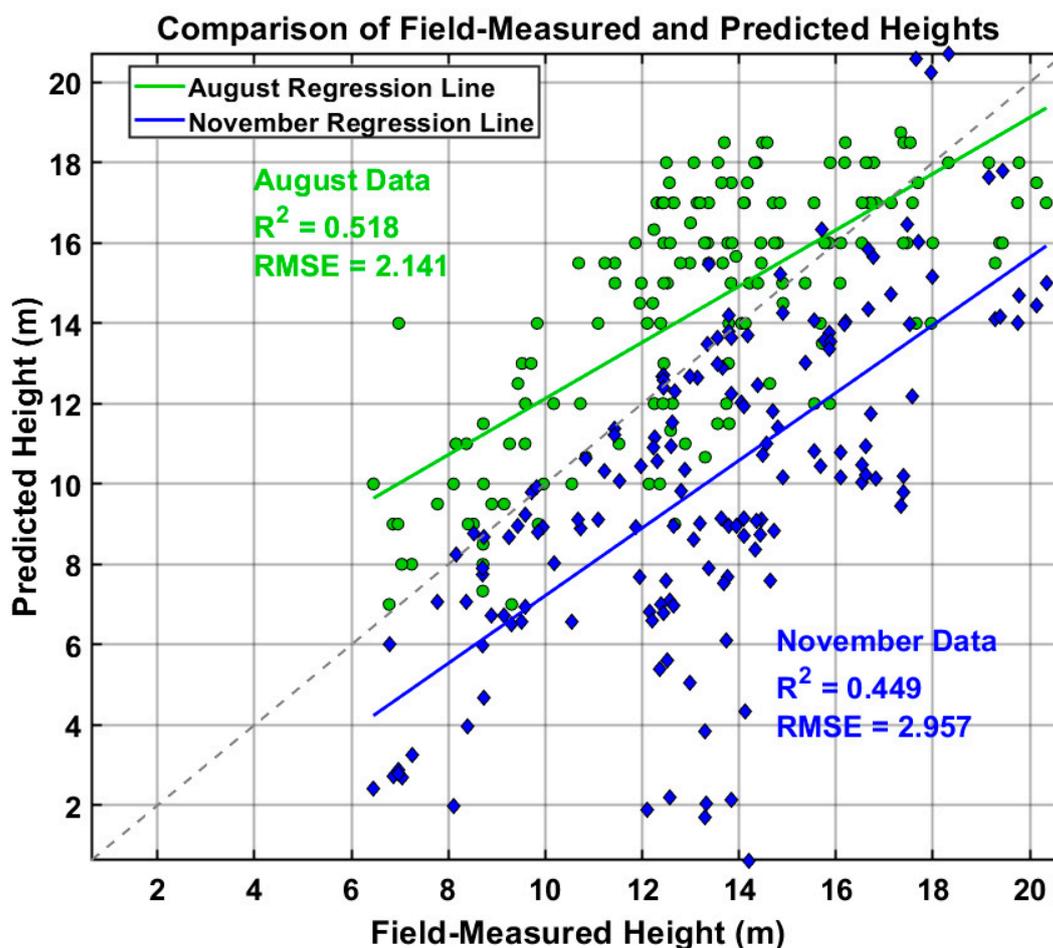
Figure 4. The scatter plot illustrates the relationship between canopy heights predicted using a canopy height model and field-measured heights for two different time points: August and November. The light green points and corresponding regression line represent the August data, while the light blue points and their regression line represent the November data. The 1:1 line (grey dashed) indicates perfect concordance between predicted and measured heights.

These results highlight the seasonal variability in the predictive performance of the canopy height model, indicating the need for further calibration to improve accuracy, particularly for November data. The overall model, however, demonstrates practical utility for estimating canopy heights in different seasons, informing forest management practices and ecological studies.

3.2. Feature Importance in Biomass Prediction

Random forests were employed to assess the relative importance of various variables for predicting AGB, using the "mean decrease in MSE" metric (see Figure 5). The S1 scenario included multispectral indices computed from GF-7 MUX bands and the original GF-7 spectral bands, while the S2 scenario added CHM data derived from DTM and GF-7 stereo images collected in August. In Scenario S1, TGI showed the highest importance, followed by SGI and class, while BAI was the least sensitive variable. The final predictive variables selected were EVI, GEMI, GARI, GLI, IOR, LAI, RGRI, SGI, TGI, VARI, and class, resulting in an R^2 of 0.71 and an RMSE of 47.60 t/ha. In Scenario S2, CHM emerged as the most critical variable, along with class and SGI, with BAI again being the least important. The final variable set included EVI, GLI, IOR, LAI, RGRI, RDVI, SGI, TGI, VARI, CHM, and class, significantly improving the model's accuracy to an R^2 of 0.90 and an RMSE of 20.01 t/ha. These findings highlight the substantial role of CHM in enhancing biomass

prediction accuracy, emphasizing the need to incorporate structural information alongside spectral indices, and the consistent low importance of BAI across both scenarios.
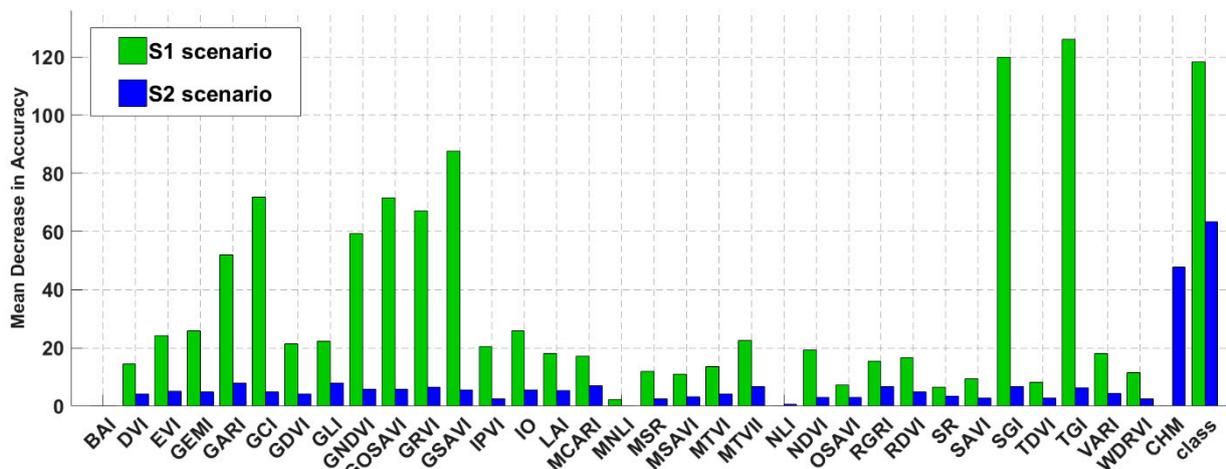

Figure 5. Feature importance scores for predicting AGB using random forest models under two scenarios: S1 and S2. Features are ranked in decreasing order of importance based on the mean decrease in mean squared error (MSE). The feature "class" refers to land cover classification data, distinguishing between forested and non-forested areas, derived from geographic national condition data.

3.3. Continuous AGB Mapping

Incorporating CHM and DSM data significantly enhances biomass prediction accuracy. The spatial distribution of the predicted biomass is shown in Figure 6. The scatter plot in Figure 7 highlights the relationship between field-measured biomass and model-predicted biomass across four scenarios (S1–S4). Regression analysis indicates a marked improvement in model performance metrics (R^2 and RMSE) when CHM and DSM data are included alongside traditional spectral indices. Specifically, scenarios S2 and S3, which incorporate CHM derived from different periods, demonstrate enhanced accuracy over scenario S1, which only utilizes spectral data. Scenario S4, integrating DSM, also shows significant predictive improvements. These results underscore the potential of CHM and DSM in refining AGB retrieval models, suggesting that structural parameters from remote sensing data critically augment conventional spectral approaches.

The models across the four scenarios are summarized in Table 3, which presents the slopes, intercepts, minimum (Min) and maximum (Max) values, and percentiles (Q1, median, Q3) of each model. This table highlights that the incorporation of the CHM in models S2 and S3 significantly improves biomass estimates. Specifically, Model S2 utilizes CHM derived from DTM and GF-7 stereo images captured in August, achieving enhanced prediction accuracy with a slope of 0.86 and a stable intercept of 15.58. Looking at the results from different models, Model S2 exhibits a minimum value of 120.81 t/ha and a maximum of 171.31 t/ha, with percentiles (Q1: 133.06, median: 138.48, Q3: 156.33) reflecting relatively tight clustering around the median, which highlights consistent predictions across the dataset. In contrast, Model S3, despite also including CHM, yields a lower slope of 0.38, with a minimum of 110.71 t/ha and a maximum of 158.13 t/ha. Its percentiles (Q1: 139.11, median: 155.46, Q3: 157.23) suggest increased variability in predictions, likely influenced by seasonal changes affecting biomass assessment. Interestingly, Model S4, which incorporates a comprehensive set of inputs, shows a notable decline in predictive capability, with a slope of 0.19. Its minimum value of 129.67 t/ha and maximum of 152.84 t/ha, along with percentiles indicating more uniform predictions (Q1: 137.16, median: 140.83, Q3: 142.85), suggest that, while it captures a broader range, the complexity of inputs may dilute

predictive power. Overall, the findings emphasize the critical role of CHM in enhancing the accuracy of forest biomass estimation.
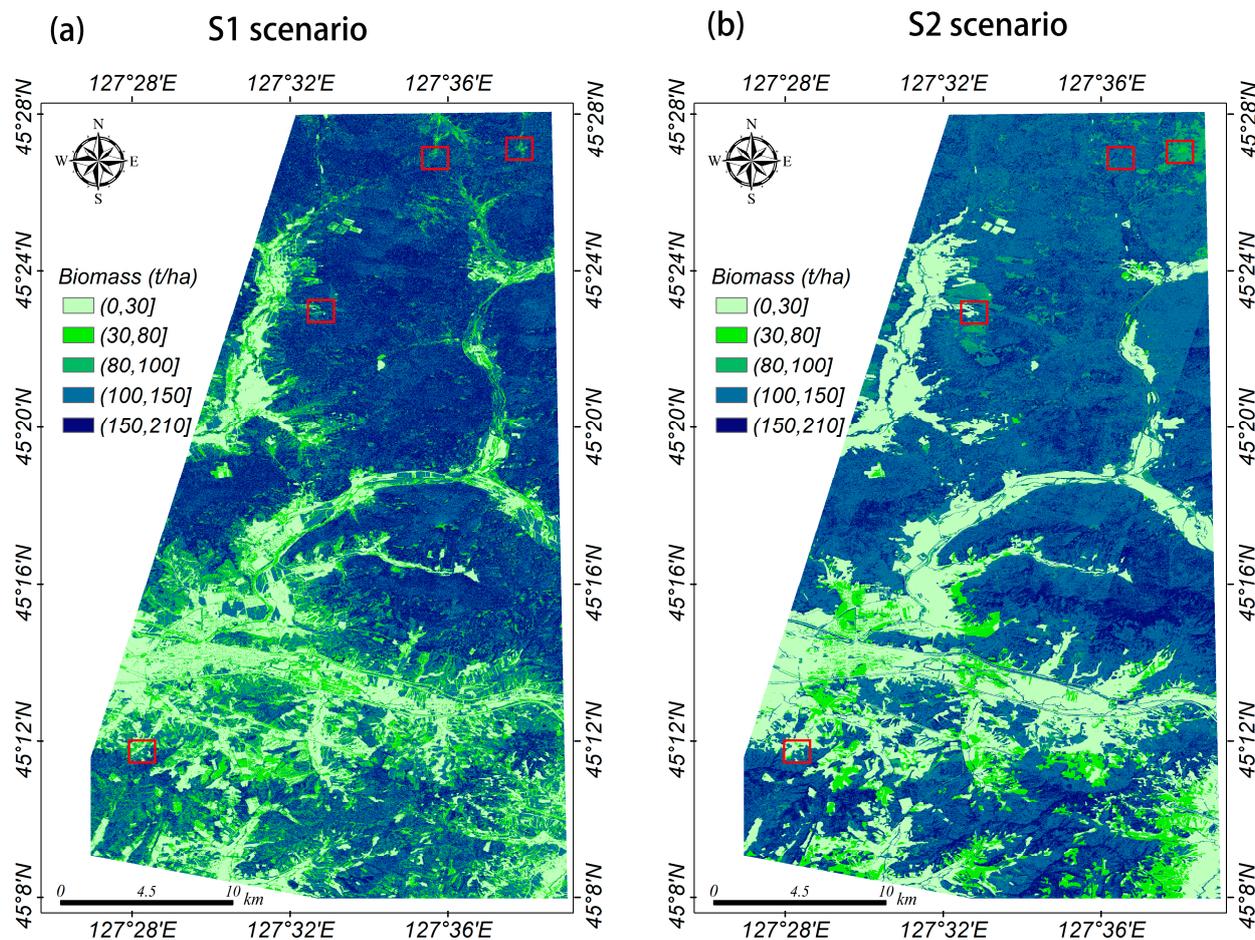

Figure 6. Predicted biomass maps in different scenarios: (**a**) for S1 scenario and (**b**) for S2 scenario. Detailed drawings of the red-framed area are shown in Figure 8.

Table 3. Statistical metrics of biomass estimation models.

Scenario	Slope	Intercept	Min	Max	Q1	Median	Q3
S1	0.84	21.23	92.79	171.97	136.36	154.14	165.68
S2	0.86	15.58	120.81	171.31	133.06	138.48	156.33
S3	0.38	88.72	110.71	158.13	139.11	155.46	157.23
S4	0.19	112.14	129.67	152.84	137.16	140.83	142.85

Figure 8 illustrates the impact of CHM on biomass estimation across various scenarios. The results highlight the significance of incorporating CHM in improving biomass predictions. Scenario (a) shows that disturbances from water bodies and soil moisture in river valley deltas lead to spectral confusion, causing biomass underestimation when relying solely on spectral properties. Scenario (b) demonstrates that, in the flat eastern region with lower canopy heights and predominant coniferous species, CHM results incorporate more accurate biomass estimates. Scenario (c) reveals that, during the growing season, spectral features alone cause an overestimation of farmland vegetation height, with the average farmland biomass being less than 5 t/ha; however, predictions improve with CHM inclusion. Scenario (d) emphasizes that at logging sites with low biomass, the use of CHM leads to more precise biomass distribution predictions. These findings underscore the criti-

cal role of canopy height information in enhancing the accuracy of biomass assessments, particularly in areas with topographic and spectral variabilities.
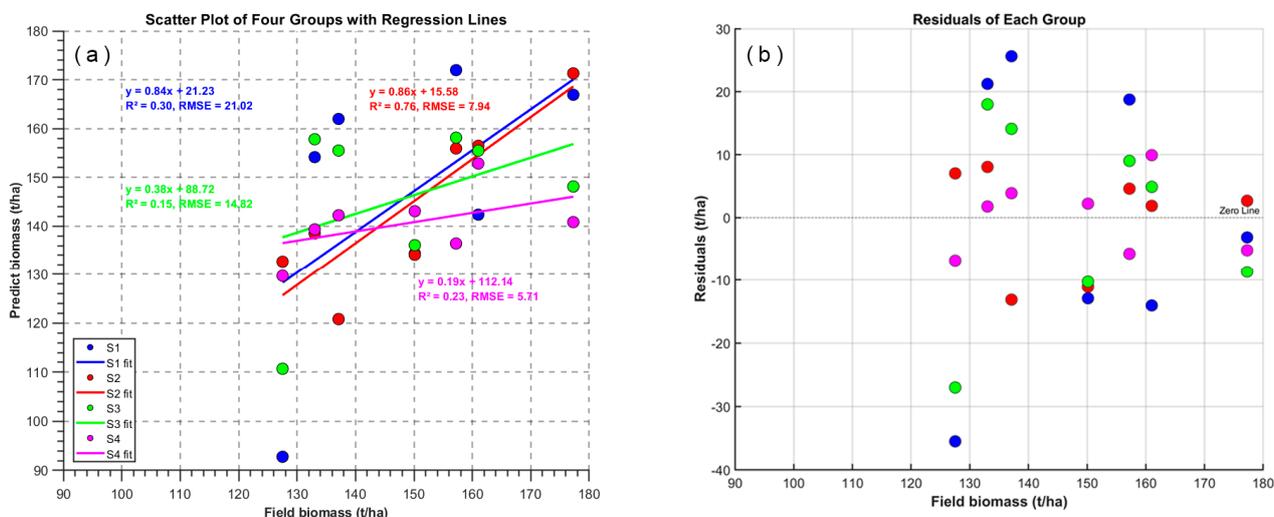

Figure 7. Scatter plots depicting the relationship between predicted biomass (t/ha) and field-measured biomass (t/ha) for four different scenarios (S1, S2, S3, and S4) in 2020. Detailed scenario descriptions are provided in Table 2. (**a**) Scatter plot includes a regression line, with annotations displaying the regression equation, coefficient of determination (R^2), and root mean square error (RMSE) to quantitatively assess model performance. (**b**) Residuals for each model's prediction compared with field biomass. The results demonstrate incremental improvements in biomass prediction accuracy from S1 to S4, highlighting the significant impact of incorporating CHM and DSM data. Scenarios S2 and S3 show enhanced prediction accuracy due to the inclusion of detailed canopy height information.
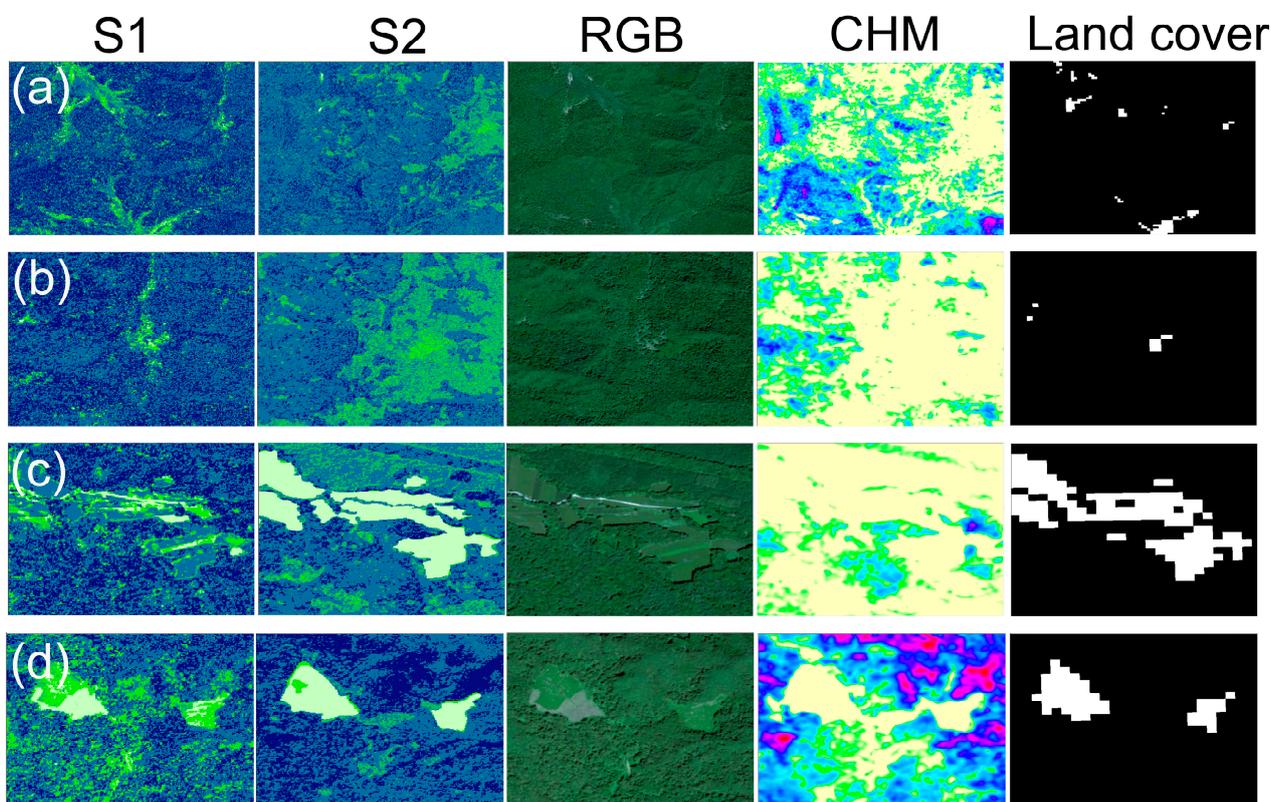

Figure 8. Biomass, spectrum, and canopy height spatial features at the same location. The biomass detail map on the far left shows the result of the S1 scenario, while the one on the right shows the result

of the S2 scenario. The biomass ramp is consistent with that shown in Figure 6, and the RGB channel denotes the real color channel display of GF-7. The CHM ramp is similar to those shown in Figure 3. The last column shows land cover, with black indicating forested area and white representing non-forest land. Figure (**a**): the disturbance of water bodies and soil moisture on the river valley delta causes the forest vegetation spectra to be mistaken for bare soil and water bodies, which leads to an underestimating of biomass forecast based solely on spectral properties. The regional variability of forest species and height under various topographic circumstances is depicted in Figure (**b**). Because of the region's eastern side's relative flatness, low forest heights, and predominance of coniferous tree species, biomass estimations that consider CHM factors are more accurate in reflecting the real distribution. Figure (**c**): due to the overestimation of the height of the farmland vegetation caused by spectral features alone (August during the growing season), the farmland's spectrum is viewed as being spectral like the forest. In agriculture, the average biomass is less than 5 t/ha, although biomass is more precisely anticipated because of the constraint that CHM is approximately 0 t/ha. A logging site is in the region of figure (**d**), with low biomass. A more accurate prediction of the biomass distribution is made by taking canopy height features into account.

4. Discussion

4.1. The Importance of Canopy Height in Biomass Estimation

The results presented in Figure 7 highlight the critical importance of incorporating CHM data for accurate AGB predictions. Our analysis compares four scenarios with varying input datasets: traditional spectral-only models (S1), models incorporating CHM derived from August (S2) and November (S3), and models utilizing digital surface models (DSMs) from different times (S4). Among these, Scenario S2, which includes CHM data from August, exhibits the highest model performance, as evidenced by the highest R^2 and the lowest root mean square error (RMSE). This indicates that the addition of precise canopy height information, particularly during the growing season when the canopy is fully developed, substantially enhances the accuracy of biomass predictions. Conversely, Scenario S3, which uses CHM derived from November (a leaf-off period), shows reduced performance, underscoring the influence of seasonal conditions on the effectiveness of CHM. Although Scenario S4, incorporating DSM data, improves biomass prediction compared with the spectral-only approach (S1), its performance does not surpass that of S2. This comparison suggests that, while DSM data are beneficial, CHM from summer conditions is more valuable in capturing canopy structure for biomass estimation.

These findings align with previous research. For example, Ni et al. [22] examines the seasonal effects on AGB estimation in mountainous deciduous forests using ZY-3 stereoscopic imagery, demonstrating that data collected during the growing season provide better AGB estimation due to a fully developed canopy structure. Our results similarly emphasize the superior accuracy of CHM derived from summer imagery. Moreover, the study by Gong et al. [37] on improving an oak canopy model using digital photogrammetry further supports these findings, highlighting the enhancement in canopy structure estimation when precise height information is used. While Ni et al. [22] focus specifically on mountainous deciduous forests and the ZY-3 platform, and Gong et al. [37] address oak canopy models with photogrammetry, our study extends these findings by comparing multiple scenarios and incorporating data from the GF-7 satellite. Additionally, our study highlights the relative benefits of CHM versus DSM data, underscoring that CHM collected during optimal conditions (summer) provides more accurate biomass predictions than DSM when included alongside traditional spectral indices. In summary, both our study and the referenced works underscore the importance of seasonal timing and precise canopy height information in remote sensing for biomass estimation. The inclusion of CHM during the growing season is significantly more effective, as it captures detailed canopy structure

crucial for accurate AGB retrieval. These findings collectively enhance our understanding of the optimal methodologies for AGB estimation in diverse environments.

4.2. Seasonal Variability in Canopy Height Model Performance

The comparison between the predicted canopy heights and field-measured heights for both August and November reveal several key aspects of model behavior and accuracy. The regression analysis for August data yield an R^2 value of 0.52 and an RMSE of 2.14 m, whereas the November data exhibit an R^2 value of 0.45 and an RMSE of 2.96 m. The higher R^2 and lower RMSE for the August data suggest that the canopy height model performs better during this month. The light green points and their regression line for August are more closely aligned with the 1:1 line, indicating a relatively higher accuracy in height predictions during this period. Conversely, the light blue points and regression line for November deviate more significantly from the 1:1 line, particularly underestimating canopy heights. This seasonal variability in model performance may result from denser foliage and more defined canopy structures in August, providing clearer signals for the model, while the partial defoliation and structural changes in November introduce noise and variability, degrading the model's predictive performance. Similar seasonal effects on remote sensing accuracy have been documented by Ni et al. [38], who found substantial differences in ground surface elevation extraction over deciduous forests between seasons. Ni et al. [22] further explain that the tops of the forest canopy are easily apparent in closed forest stands throughout the leaf-on season; in open forest stands, the ground surface is visible in mid-summer through canopy gaps, but it may be obscured in late summer or early fall.

These results underscore the necessity for careful selection of stereo satellite images when deriving canopy height models. The demonstrated seasonal variability suggests that images captured during periods with fully developed canopy structures (e.g., late summer) yield more accurate predictions. Therefore, for critical applications requiring high precision, such as forest carbon stock assessment or habitat monitoring, it is advisable to select stereo pairs from seasons that minimize foliage variability. Despite this variability, the canopy height model shows practical utility for large-scale canopy height estimation, critical for forest management and ecological monitoring. The promising performance in August suggests prioritizing this period for high-accuracy surveys. For periods like November, where the model's performance diminishes, incorporating more field measurements can validate and complement model outputs. Further integration of multi-temporal datasets and advanced remote sensing technologies, such as LiDAR, could significantly improve model robustness and accuracy across different seasons, facilitating better forest assessments and ecological studies.

4.3. Enhancing AGB Mapping: The Role of Optical Stereo Images

The integration of optical stereo images for AGB mapping offers a promising complement to traditional methods like LiDAR [13,39] and SAR [10,40], providing several advantages and some challenges. Optical stereo images are cost-effective, accessible, and provide high spatial resolution, which enhances the detail and accuracy of biomass estimates by capturing fine-scale vegetation variations. They also offer high temporal resolution, allowing for frequent monitoring and understanding of seasonal dynamics. Additionally, when combined with LiDAR and SAR, optical images can significantly improve biomass estimation accuracy [13]. However, they are dependent on weather conditions, such as cloud cover, and require complex data processing to derive accurate CHM. Moreover, data saturation in dense forests can limit their effectiveness [41,42]. Prospects for optical stereo images in biomass mapping include integrating multi-source data to leverage

the strengths of each method, advancing image processing algorithms, and employing machine learning and AI to refine analysis and predictive models. Establishing global monitoring programs utilizing optical stereo imagery can provide consistent up-to-date biomass data, supporting efforts in climate change mitigation, biodiversity conservation, and sustainable forest management. Despite the challenges, rapid advancements are likely to enhance the utility of optical stereo images, making them a crucial component in the future of forest monitoring.

4.4. Accuracy Comparisons with LiDAR and Other Remote Sensing Methods

In this study, we recognize the importance of discussing the accuracy of canopy height and biomass estimates derived from remote sensing techniques. While our results demonstrate that the incorporation of the CHM from GF-7 significantly enhances biomass estimation accuracy (R^2 = 0.76, RMSE = 7.94 t/ha), it is crucial to consider how these results align with accuracy standards set by other technologies. Previous studies have indicated that LiDAR systems, such as NASA's GEDI, typically achieve high accuracy levels, with R^2 values often exceeding 0.8 in forest environments [43]. Furthermore, studies have shown that airborne LiDAR can provide detailed canopy structure information, often resulting in lower RMSE values compared with non-LiDAR approaches [44]. However, we must recognize that geographical variability, the specific forest types examined, and the limited number of ground truth measurements may introduce uncertainties into our biomass estimates, potentially limiting the broader applicability of our findings. To frame our results within a broader context, we refer to the literature, indicating that "good" accuracy in AGB prediction is typically characterized as an R^2 value of 0.70 or higher [15,41]. Although direct comparisons at the plot level may not be feasible with our current dataset, exploring aggregate comparisons at larger spatial scales can yield valuable insights into the relative efficiency and accuracy of satellite photogrammetry versus LiDAR methods such as ICESat-1/2 [16,39] or GEDI [15], which provide point sampling data for AGB prediction.

Regarding the differences in results between spaceborne photogrammetry and airborne methods, while our study focuses primarily on spaceborne stereo imagery for estimating AGB, we acknowledge that understanding these differences is important. Airborne methods for predicting AGB offer several advantages, including higher spatial resolution and detailed canopy structure information, which enhance accuracy in AGB estimates. However, these methods are limited by smaller coverage areas and often higher costs associated with data acquisition [45]. By situating our findings within the larger framework of remote sensing technologies that integrate spaceborne LiDAR for sampling with stereo photogrammetry or SAR imagery for comprehensive AGB mapping, we aim to offer a clearer understanding of the strengths and limitations inherent to each method. This knowledge will ultimately support informed decision making in forest management and carbon stock assessments.

5. Conclusions

In this study, we have shown that the GF-7 very high-resolution stereo satellite is a useful tool for AGB and forest canopy height estimation. Our findings demonstrate that the accuracy of biomass prediction is much increased by the incorporation of canopy height model (CHM) data, especially when the data are collected from summer images when the canopy is fully developed. When compared with spectral-only models, the random forest model performs better in scenarios involving CHM, producing an R^2 = 0.76, RMSE = 7.94 t/ha. This high degree of accuracy highlights the value of GF-7 in recording specific forest structural metrics, offering a high-resolution, globally accessible, and reasonably priced tool for extensive forest monitoring. Overall, the GF-7 satellite demonstrates

significant potential in advancing remote sensing applications in forestry. Our approach offers significant improvements in the precision of AGB estimates by incorporating structural information, which is crucial for ecological studies, forest management, and climate change mitigation efforts.

References

1. Bonan, G.B. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science* **2008**, *320*, 1444–1449. [CrossRef] [PubMed]
2. Shugart, H.; Saatchi, S.; Hall, F. Importance of structure and its measurement in quantifying function of forest ecosystems. *J. Geophys. Res. Biogeosci.* **2010**, *115*. [CrossRef]
3. Fang, J.-y.; Wang, G.G.; Liu, G.-h.; Xu, S.-l. Forest biomass of China: An estimate based on the biomass–volume relationship. *Ecol. Appl.* **1998**, *8*, 1084–1091.
4. Yang, H.; Ciais, P.; Frappart, F.; Li, X.; Brandt, M.; Fensholt, R.; Fan, L.; Saatchi, S.; Besnard, S.; Deng, Z. Global increase in biomass carbon stock dominated by growth of northern young forests over past decade. *Nat. Geosci.* **2023**, *16*, 886–892. [CrossRef]
5. Dubayah, R.O.; Drake, J.B. Lidar remote sensing for forestry. *J. For.* **2000**, *98*, 44–46. [CrossRef]
6. St-Onge, B.; Hu, Y.; Vega, C. Mapping the height and above-ground biomass of a mixed forest using lidar and stereo Ikonos images. *Int. J. Remote Sens.* **2008**, *29*, 1277–1294. [CrossRef]
7. Minh, D.H.T.; Le Toan, T.; Rocca, F.; Tebaldini, S.; d'Alessandro, M.M.; Villard, L. Relating P-band synthetic aperture radar tomography to tropical forest biomass. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 967–979. [CrossRef]
8. Sader, S.A.; Waide, R.B.; Lawrence, W.T.; Joyce, A.T. Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat TM data. *Remote Sens. Environ.* **1989**, *28*, 143–198.
9. Castillo-Santiago, M.A.; Ricker, M.; de Jong, B.H. Estimation of tropical forest structure from SPOT-5 satellite images. *Int. J. Remote Sens.* **2010**, *31*, 2767–2782. [CrossRef]
10. Stelmaszczuk-Górska, M.A.; Urbazaev, M.; Schmullius, C.; Thiel, C. Estimation of above-ground biomass over boreal forests in siberia using updated in situ, ALOS-2 PALSAR-2, and RADARSAT-2 data. *Remote Sens.* **2018**, *10*, 1550. [CrossRef]
11. Shi, Y.; Wang, Z.; Zhang, G.; Wei, X.; Ma, W.; Yu, H. Evaluating the Research Status of the Remote Sensing-Mediated Monitoring of Forest Biomass: A Bibliometric Analysis of WOS. *Forests* **2024**, *15*, 524. [CrossRef]
12. Chen, G.; Hay, G.J.; St-Onge, B. A GEOBIA framework to estimate forest parameters from lidar transects, Quickbird imagery and machine learning: A case study in Quebec, Canada. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *15*, 28–37. [CrossRef]
13. Huang, H.; Liu, C.; Wang, X.; Zhou, X.; Gong, P. Integration of multi-resource remotely sensed data and allometric models for forest aboveground biomass estimation in China. *Remote Sens. Environ.* **2019**, *221*, 225–234. [CrossRef]
14. Simard, M.; Pinto, N.; Fisher, J.B.; Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res. Biogeosci.* **2011**, *116*. [CrossRef]
15. Duncanson, L.; Kellner, J.R.; Armston, J.; Dubayah, R.; Minor, D.M.; Hancock, S.; Healey, S.P.; Patterson, P.L.; Saarela, S.; Marselis, S. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sens. Environ.* **2022**, *270*, 112845. [CrossRef]
16. Huang, H.; Liu, C.; Wang, X.; Biging, G.S.; Chen, Y.; Yang, J.; Gong, P. Mapping vegetation heights in China using slope correction ICESat data, SRTM, MODIS-derived and climate data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 189–199. [CrossRef]

17. Wang, X.; Huang, H.; Gong, P.; Liu, C.; Li, C.; Li, W. Forest canopy height extraction in rugged areas with ICESAT/GLAS data. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 4650–4657. [CrossRef]

18. Liu, C.; Wang, X.; Huang, H.; Gong, P.; Wu, D.; Jiang, J. The importance of data type, laser spot density and modelling method for vegetation height mapping in continental China. *Int. J. Remote Sens.* **2016**, *37*, 6127–6148. [CrossRef]

19. Tang, H.; Huang, H.; Zheng, Y.; Qin, P.; Xu, Y.; Ding, S. Improved GEDI canopy height extraction based on a simulated ground echo in topographically undulating areas. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [CrossRef]

20. Quegan, S.; Le Toan, T.; Chave, J.; Dall, J.; Exbrayat, J.-F.; Minh, D.H.T.; Lomas, M.; D'alessandro, M.M.; Paillou, P.; Papathanassiou, K. The European Space Agency BIOMASS mission: Measuring forest above-ground biomass from space. *Remote Sens. Environ.* **2019**, *227*, 44–60. [CrossRef]

21. Ni, W.; Zhang, Z.; Sun, G.; Liu, Q. Modeling the stereoscopic features of mountainous forest landscapes for the extraction of forest heights from stereo imagery. *Remote Sens.* **2019**, *11*, 1222. [CrossRef]

22. Ni, W.; Yu, T.; Pang, Y.; Zhang, Z.; He, Y.; Li, Z.; Sun, G. Seasonal effects on aboveground biomass estimation in mountainous deciduous forests using ZY-3 stereoscopic imagery. *Remote Sens. Environ.* **2023**, *289*, 113520. [CrossRef]

23. Huang, H.; Chen, P.; Xu, X.; Liu, C.; Wang, J.; Liu, C.; Clinton, N.; Gong, P. Estimating building height in China from ALOS AW3D30. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 146–157. [CrossRef]

24. Bhushan, S.; Shean, D.; Alexandrov, O.; Henderson, S. Automated digital elevation model (DEM) generation from very-high-resolution Planet SkySat triplet stereo and video imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 151–165. [CrossRef]

25. Meddens, A.J.; Vierling, L.A.; Eitel, J.U.; Jennewein, J.S.; White, J.C.; Wulder, M.A. Developing 5 m resolution canopy height and digital terrain models from WorldView and ArcticDEM data. *Remote Sens. Environ.* **2018**, *218*, 174–188.

26. Xie, J.; Huang, G.; Liu, R.; Zhao, C.; Dai, J.; Jin, T.; Mo, F.; Zhen, Y.; Xi, S.; Tang, H. Design and data processing of China's first spaceborne laser altimeter system for earth observation: GaoFen-7. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1034–1044.

27. Ni, W.; Li, Z.; Wang, Q.; Zhang, Z.; Liu, Q.; Pang, Y.; He, Y.; Li, Z.; Sun, G. Forest heights extraction using GF-7 very high-resolution stereoscopic imagery and Google Earth multi-temporal historical imagery. *J. Remote Sens.* **2024**, *4*, 0158. [CrossRef]

28. Puliti, S.; Hauglin, M.; Breidenbach, J.; Montesano, P.; Neigh, C.; Rahlf, J.; Solberg, S.; Klingenberg, T.; Astrup, R. Modelling above-ground biomass stock over Norway using national forest inventory data with ArcticDEM and Sentinel-2 data. *Remote Sens. Environ.* **2020**, *236*, 111501. [CrossRef]

29. Zhang, T.; Lin, H.; Long, J.; Zhang, M.; Liu, Z. Analyzing the Saturation of Growing Stem Volume Based on ZY-3 Stereo and Multispectral Images in Planted Coniferous Forest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *15*, 50–61. [CrossRef]

30. Maack, J.; Kattenborn, T.; Fassnacht, F.E.; Enßle, F.; Hernández, J.; Corvalán, P.; Koch, B. Modeling forest biomass using Very-High-Resolution data—Combining textural, spectral and photogrammetric predictors derived from spaceborne stereo images. *Eur. J. Remote Sens.* **2015**, *48*, 245–261. [CrossRef]

31. Zhu, X.; Tang, X.; Zhang, G.; Liu, B.; Hu, W. Accuracy comparison and assessment of DSM derived from GFDM satellite and GF-7 satellite imagery. *Remote Sens.* **2021**, *13*, 4791. [CrossRef]

32. Chen, P.; Huang, H.; Liu, J.; Wang, J.; Liu, C.; Zhang, N.; Su, M.; Zhang, D. Leveraging Chinese GaoFen-7 imagery for high-resolution building height estimation in multiple cities. *Remote Sens. Environ.* **2023**, *298*, 113802. [CrossRef]

33. Chen, P.; Huang, H.; Ye, F.; Liu, J.; Li, W.; Wang, J.; Wang, Z.; Liu, C.; Zhang, N. A benchmark GaoFen-7 dataset for building extraction from satellite images. *Sci. Data* **2024**, *11*, 187. [CrossRef]

34. Wu, J.; Meng, Q.; Gao, L.; Zhang, L.; Zhao, M.; Su, C. A deep learning framework for 3D vegetation extraction in complex urban environments. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *129*, 103798. [CrossRef]

35. Du, L.; Pang, Y.; Ni, W.; Liang, X.; Li, Z.; Suarez, J.; Wei, W. Forest terrain and canopy height estimation using stereo images and spaceborne LiDAR data from GF-7 satellite. *Geo-Spat. Inf. Sci.* **2024**, *27*, 811–821. [CrossRef]

36. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

37. Gong, P.; Mei, X.; Biging, G.S.; Zhang, Z. Improvement of an oak canopy model extracted from digital photogrammetry. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 919–924.

38. Ni, W.; Sun, G.; Ranson, K.J.; Pang, Y.; Zhang, Z.; Yao, W. Extraction of ground surface elevation from ZY-3 winter stereo imagery over deciduous forested areas. *Remote Sens. Environ.* **2015**, *159*, 194–202. [CrossRef]

39. Travers-Smith, H.; Coops, N.C.; Mulverhill, C.; Wulder, M.A.; Ignace, D.; Lantz, T.C. Mapping vegetation height and identifying the northern forest limit across Canada using ICESat-2, Landsat time series and topographic data. *Remote Sens. Environ.* **2024**, *305*, 114097. [CrossRef]

40. Musthafa, M.; Singh, G. Improving Forest Above-Ground Biomass Retrieval Using Multi-Sensor L- and C- Band SAR Data and Multi-Temporal Spaceborne LiDAR Data. *Front. For. Glob. Change* **2022**, *5*, 822704. [CrossRef]

41. Cao, L.; Coops, N.C.; Sun, Y.; Ruan, H.; Wang, G.; Dai, J.; She, G. Estimating canopy structure and biomass in bamboo forests using airborne LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2019**, *148*, 114–129. [CrossRef]

42. Su, H.; Shen, W.; Wang, J.; Ali, A.; Li, M. Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. *For. Ecosyst.* **2020**, *7*, 64. [CrossRef]

43. Shendryk, Y. Fusing GEDI with earth observation data for large area aboveground biomass mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *115*, 103108. [CrossRef]

44. Du, L.; Pang, Y.; Wang, Q.; Huang, C.; Bai, Y.; Chen, D.; Lu, W.; Kong, D. A LiDAR biomass index-based approach for tree-and plot-level biomass mapping over forest farms using 3D point clouds. *Remote Sens. Environ.* **2023**, *290*, 113543. [CrossRef]

45. Tebaldini, S.; Ho Tong Minh, D.; Mariotti d'Alessandro, M.; Villard, L.; Le Toan, T.; Chave, J. The status of technologies to measure forest biomass and structural properties: State of the art in SAR tomography of tropical forests. *Surv. Geophys.* **2019**, *40*, 779–801. [CrossRef]