remote sensing

MDPI

Article

Upscaling Forest Canopy Height Estimation Using Waveform-Calibrated GEDI Spaceborne LiDAR and Sentinel-2 Data

Junjie Wang, Xin Shen and Lin Cao *

Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China; wangjunjie@njfu.edu.cn (J.W.); xinshen@njfu.edu.cn (X.S.)
* Correspondence: lincao@njfu.edu.cn

Abstract: Forest canopy height is a fundamental parameter of forest structure, and plays a pivotal role in understanding forest biomass allocation, carbon stock, forest productivity, and biodiversity. Spaceborne LiDAR (Light Detection and Ranging) systems, such as GEDI (Global Ecosystem Dynamics Investigation), provide large-scale estimation of ground elevation, canopy height, and other forest parameters. However, these measurements may have uncertainties influenced by topographic factors. This study focuses on the calibration of GEDI L2A and L1B data using an airborne LiDAR point cloud, and the combination of Sentinel-2 multispectral imagery, 1D convolutional neural network (CNN), artificial neural network (ANN), and random forest (RF) for upscaling estimated forest height in the Guangxi Gaofeng Forest Farm. First, various environmental (i.e., slope, solar elevation, etc.) and acquisition parameters (i.e., beam type, Solar elevation, etc.) were used to select and optimize the L2A footprint. Second, pseudo-waveforms were simulated from the airborne LiDAR point cloud and were combined with a 1D CNN model to calibrate the L1B waveform data. Third, the forest height extracted from the calibrated L1B waveforms and selected L2A footprints were compared and assessed, utilizing the CHM derived from the airborne LiDAR point cloud. Finally, the forest height data with higher accuracy were combined with Sentinel-2 multispectral imagery for an upscaling estimation of forest height. The results indicate that through optimization using environmental and acquisition parameters, the ground elevation and forest canopy height extracted from the L2A footprint are generally consistent with airborne LiDAR data (ground elevation: R^2 = 0.99, RMSE = 4.99 m; canopy height: R^2 = 0.42, RMSE = 5.16 m). Through optimizing, ground elevation extraction error was reduced by 45.5% (RMSE), and the canopy height extraction error was reduced by 30.3% (RMSE). After training a 1D CNN model to calibrate the forest height, the forest height information extracted using L1B has a high accuracy (R^2 = 0.84, RMSE = 3.13 m). Compared to the optimized L2A data, the RMSE was reduced by 2.03 m. Combining the more accurate L1B forest height data with Sentinel-2 multispectral imagery and using RF and ANN for the upscaled estimation of the forest height, the RF model has the highest accuracy (R^2 = 0.64, RMSE = 4.59 m). The results show that the extrapolation and inversion of GEDI, combined with multispectral remote sensing data, serve as effective tools for obtaining forest height distribution on a large scale.

Keywords: GEDI Spaceborne LiDAR; waveform calibrate; upscale estimation; forest canopy height

check for updates

Citation: Wang, J.; Shen, X.; Cao, L. Upscaling Forest Canopy Height Estimation Using Waveform-Calibrated GEDI Spaceborne LiDAR and Sentinel-2 Data. *Remote Sens.* **2024**, *16*, 2138. https://doi.org/10.3390/rs16122138

Academic Editor: Nikolay S. Strigul

Received: 12 April 2024
Revised: 24 May 2024
Accepted: 31 May 2024
Published: 13 June 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/).

1. Introduction

Forests cover about 30% of the global land area, and are the largest component of terrestrial ecosystems [1], as well as an important carbon sink for terrestrial ecosystems. The exchange of carbon through processes like photosynthesis and respiration within forests accounts for a significant portion, ranging from 50% to 90%, of the total carbon fluxes in terrestrial ecosystems [2], playing a dominant role in the regional and global carbon cycle [3,4]. In order to increase forest carbon sinks, maintain ecological balance, meet the growing demand for timber, and reduce the consumption of natural forests

and the destruction of forest ecosystems [5,6], large-scale plantation forest construction began in the 1960s, and plantation forests accounted for about 6.95% of the world's forest area, which is an important component of the world's forest resources. China, as the country with the largest and fastest growing area of plantation forests in the world, is particularly important to monitor the forest resources of plantation forests quickly and accurately in such a background. The structure of forests, which influences the growth status and biodiversity of forests [7], determines the carbon sequestration capacity, biomass distribution, and carbon storage of forests [8]. Forest height is one of the fundamental and crucial parameters of forest structural characteristics. Accurate extraction of forest canopy height is highly significant for studying forest biomass distribution, carbon sink, forest productivity, and biodiversity in China [9,10].

The traditional method of measuring canopy height primarily uses an altimeter, which, while simple and highly precise for individual trees, is time-consuming, laborious, and inefficient for large forest areas. Remote sensing technology addresses these shortcomings, allowing for the efficient and accurate estimation of forest height over large areas. Optical, microwave, and laser remote sensing techniques can efficiently gather forest parameters on a large scale [11]. Optical remote sensing methods capture spatially continuous texture distribution and spectral information of the forest canopy, but are limited by factors such as slope, solar elevation angle, and shading, which impact accuracy [12,13]. Additionally, optical images provide horizontal distribution information, but not the vertical structure, and the spectral signal can become saturated [14]. Microwave remote sensing offers some penetration through the canopy to extract vertical structure parameters, but its signal can be affected by topographic relief and also experiences saturation [15]. To address these limitations, LiDAR, an active remote sensing technology, measures the time difference between emitted and returned laser pulses to acquire precise distance information. It can obtain three-dimensional data of forest canopies from ground-based, airborne, and spaceborne platforms, achieving vertical accuracy of 15 to 30 cm for terrain and vegetation [16]. This makes LiDAR a powerful tool for estimating forest structural parameters, including canopy height, vertical structure, and biomass [17,18]. With the development of UAV remote sensing technology, airborne LiDAR has become widely used for canopy height estimation, offering rapid acquisition of 3D terrain data and detailed descriptions of understory topography and canopy structure [19,20]. However, despite its advantages, airborne LiDAR's limited spatial coverage and high costs remain challenges for large-area data acquisition.

Spaceborne LiDAR has the advantages of large spatial coverage and repeatable observation, which can make up for the shortcomings of airborne LiDAR, and efficiently obtain the vertical information of vegetation over a wide range based on guaranteeing accuracy and time consistency. In December 2018, NASA successfully launched GEDI (Global Ecosystem Dynamics Investigation). When carried on the International Space Station, GEDI carries a laser altimeter dedicated to observing the high-resolution vertical structure of forests, each ground footprint the diameter of each ground footprint is about 25 m, and the beam dithering unit can produce eight tracks of data, with a distance of 600 m between the tracks and a footprint spacing of 60 m along the track. The returned full-waveform LiDAR data are observed, and the waveforms can be used to extract a variety of canopy parameters, such as the canopy height, the relative height of the canopy, and the ground level of the forest floor, etc. The GEDI plan follows the ISS during the planned work period of at least two years and continues to measure the forest heights in tropical, subtropical, and temperate regions within 51.6° north–south latitude [21], providing data support for the studies on the surface vegetation structure and carbon stock distribution. The data products of GEDI are divided into four levels: L1 is the raw geo-located waveform data, L2 is the footprint level canopy height and contour index product, L3 is the gridded canopy height product, and L4 is the footprint and gridded biomass estimation product [22], and the GEDI-related studies mainly use L2A and L1B. L2A and L1B data are mainly used in GEDI-related studies, with a high proportion of studies adopting L2A data. L1B is the

raw waveform data that has been geo-located, and L2A is the ground level, canopy height, and relative height data extracted from the waveform. In recent years, many scholars have studied the GEDI acquisition accuracy and its parameter characteristics. Quirós [23] took the southern region of Spain as the study area, and chose the relative height index RH100 to represent the height of the forest canopy. Compared to the digital elevation model extracted with the airborne LiDAR, the difference in the ground elevation obtained by both of them had a RMSE of 6.13 m, and it was found that the factors of the degree of cover and slope had a significant effect on the extraction. Adam et al. selected a study area in central Germany, and chose a variety of environmental and acquisition parameters to explore their effects on GEDI height measurements. Slope, vegetation height, and beam sensitivity had the greatest impact on accuracy [21]. It was found that GEDI and DEM derived from airborne LiDAR have a high correlation (R^2 = 0.93–0.99), while there was a weak correlation between GEDI and the airborne LiDAR-derived canopy height model (R^2 = 0.27–0.34). Fayad [24] used GEDI L2A data to estimate the accuracy of water levels in eight lakes in Switzerland, and the agreement with the local hydrological station data was good, with the average elevation deviation ranging from −13.8 to 9.8 cm, and it was also found that GEDI data collected in the morning or at night had a lower bias than that of the data collected during the daytime. The environmental and acquisition parameters are factors that must be considered when studying GEDI data, especially L2A data.

GEDI provides six different algorithms for processing waveform data and extracting terrain and forest canopy parameters to obtain high-quality observation data. Adam et al. [21] used airborne LiDAR as reference data to compare the differences between GEDI L2A forest understory terrain and canopy height data and airborne extraction under different algorithms, and found that algorithm setting group 5 gave significantly worse results compared to the other algorithms under different conditions. Each has its own advantages and disadvantages. Potapov [25] analyzed the correlation between the GEDI relative height metrics and the airborne LiDAR canopy height baseline data and found that RH95 had the highest correlation (RMSE = 6.5 m). Liu [26] evaluated six waveform processing algorithms of GEDI with five forests in the United States as the study area, and combined them with corresponding airborne LiDAR data, and found that the forest canopy height was relatively high in low coverage (greater than or equal to 80°) in Algorithm 4 and high coverage (less than or equal to 80°) in Algorithm 2, and Algorithm 1 had the highest accuracy in the rest of the coverage, with RMSEs of 6.98 m, 4.66 m, and 7.11 m. Lin [27] selected RH50, RH75, RH90, RH95, and RH100 to compare with the CHM products of airborne LiDAR at 30 m spatial resolution, respectively, to search for the attribute field closest to the forest canopy height, and finally chose RH95 (Mean Error = 1.43 m) as the value of forest height for the GEDI footprint. Zhu [28] found that the forest height RH95 extracted from GEDI data had the highest correlation with that extracted from airborne LiDAR data (R^2 = 0.95, RMSE = 2.22 m), and that GEDI extracted forest height with higher accuracy compared to ICESat-2/ATLAS. Taken together, the selection of appropriate preprocessing algorithms and filtering conditions can reduce the GEDIL2A data error, and the relative height RH95 is a representative indicator in reflecting forest height information.

In addition, some studies have attempted to utilize GEDI L1B raw positioning data to complete forest height information extraction and evaluate its accuracy. Lahssini [29] investigated the extraction accuracy of GEDI in the tropical forest region of Gabon using a single RH95 model, a random forest regression model using multiple GEDI waveform feature parameters, and a convolutional neural network model using L1B localized waveforms, respectively, and found that the estimation accuracy (RMSE) of the random forest model built using multiple waveform parameters was increased by 80% compared to the model based on a single parameter. Lang [30] proposed a Bayesian deep convolutional neural network approach to parameterize the complete GEDI L1B waveform input into a one-dimensional convolutional neural network model, which ultimately estimates global forest heights, resulting in the generation of global forest height estimates expected to have a low bias (RMSE of 4.4 m). The deep learning regression of forest height using GEDI

waveform modeling can avoid the process of modeling processing and feature engineering learning for atmospheric, environmental, and instrumental effects, and reduce the time for waveform processing, which has great potential for application. GEDI extracts data through discrete ground sampling. The discrete ground sampling mode brings the disadvantage of spatial sampling data discontinuity. To overcome this disadvantage, researchers introduced spatially continuous passive remote sensing imagery and fused GEDI spaceborne LiDAR to complete mapping larger-scale areas. Potapov [25] fused forest canopy height data from GEDI with Landsat8 OLI imagery and then regressed it using a bagged regression tree integration method to map global forest canopy height products at 30 m spatial resolution for 2019, with RMSE ranging from 6.6–9 m.

Combined with the previous studies, it can be observed that previous studies predominantly utilized GEDI L2A data, whereas investigations based on L1B waveform data seldom calibrated the waveform and mostly relied on the use of waveform parameters. There is little research that focuses on forest height estimation that combine calibrated waveform data with spatially continuous Sentinel-2 multispectral imagery. In addition, previous inversion studies on forest height and other parameters mainly focused on the global scale or national scale, and the study areas were mostly concentrated in the forests of Europe, the Americas, and Africa, while there were fewer studies on upscaling forest heights at the scale of forest plantations in Chinese regions. Our detailed objectives are as follows: (1) to calibrate the GEDI L1B waveform data combined using the pseudo-waveforms simulated from the airborne LiDAR point cloud and 1D CNN model; (2) to select and optimize the L2A footprint using the environmental and acquisition parameters, then compare and assess the forest height utilizing the CHM derived from the airborne LiDAR point cloud; and (3) to upscale estimate the forest height using GEDI footprint and Sentinel-2 multispectral imagery through machine learning algorithms. In order to achieve our research goals, we will focus on two issues. First, we will explore how airborne LiDAR data can be used to improve the accuracy of forest height parameters in GEDI L1B and L2A products. Second, we will investigate how to achieve the extrapolation mapping of forest height information by integrating discrete GEDI patches with continuous satellite optical imagery through the effective correction of GEDI data.

2. Materials and Methods

The workflow of the study is shown in Figure 1.

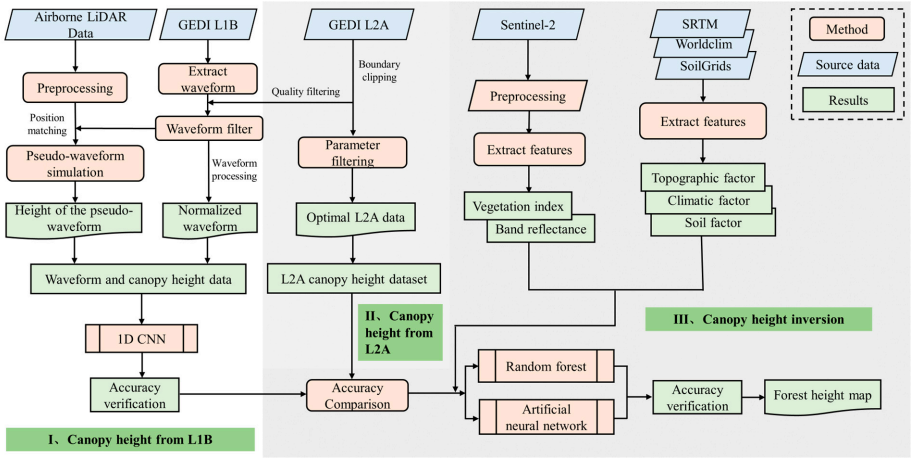


Figure 1. Workflow for upscale estimation of forest canopy height using a novel waveform-calibrated approach based on GEDI data and multispectral data. Part 1 calibrated the forest height based on the 1D CNN model using the GEDI L1B waveform and airborne LiDAR point cloud; Part 2 optimized the forest height of the L2A footprint using multiple parameters; Part 3 used the optimal forest height in combination with multispectral imagery and other data to estimate the forest height by upscaling the forest height using machine learning algorithms.

2.1. Study Area

Our study area is the Gaofeng State Owned Forest Farm in Nanning City, Guangxi (Figure 2), located in the south of China (longitude 108°08′~108°32′E, latitude 22°50′~23°04′N), which manages an area of about 1.5 million acres and has a forest volume of about 7 million m^3, making it the largest state-owned forest in Guangxi. The climate of the Forest Farm belongs to the southern subtropical monsoon climate type. The mean annual temperature is 21.6 °C, the mean annual precipitation is 1300.6 mm, and the mean annual sunshine time is 1827.0 h. The landscape type is mainly hilly, with partial low hills. The relative height is about 85–480 m above sea level. The soil is mainly red soil developed from sand shale. Gao Feng Forest Farm is dominated by eucalyptus (Eucalyptus robusta Smith), ponytail pine (Pinus massoniana Lamb.), and Chinese fir (Cunninghamia lanceolata (Lamb.) Hook.). In total, 96% of the total forest area is plantation forest.

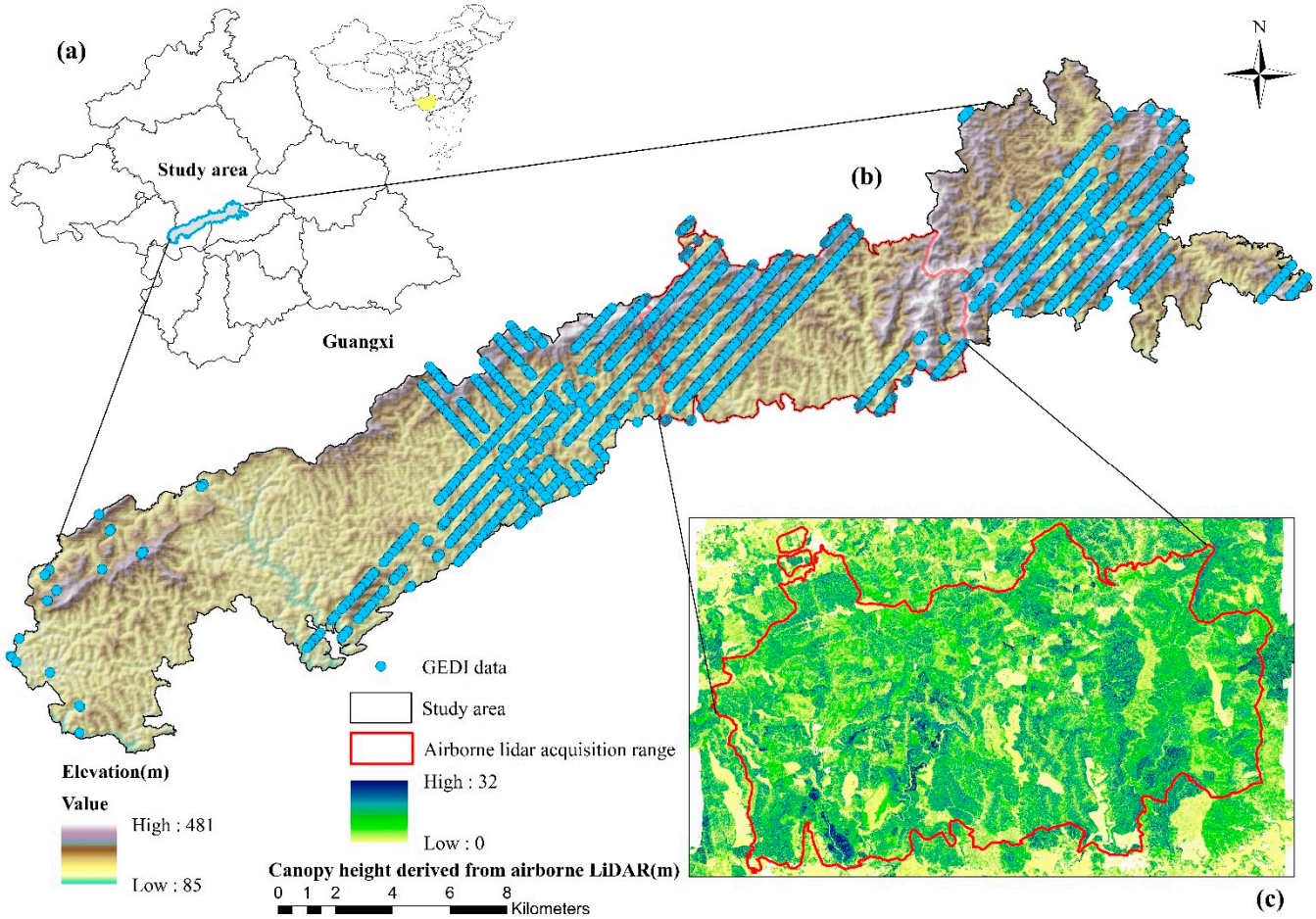


Figure 2. The map above primarily shows the geographic location of the study area. (**a**) Geographic location of the study area in Guangxi Zhuang Autonomous Region. (**b**) Distribution of GEDI footprints (blue dots) in the Gaofeng Forest Farm; the bottom map is the digital elevation model map of the Forest area and the boundary of the typical area of the boundary plate Dongsheng sub-field (red line). (**c**) Schematic of canopy height acquired using airborne LiDAR in a typical area.

2.2. Data and Pre-Processing

2.2.1. GEDI Spaceborne LiDAR

GEDI's laser ranging system accurately measures the vertical structure of forests in temperate and tropical regions. The instrument records waveform data at 1-ns (15-cm) intervals [22]. GEDI consists of three lasers, two of which are operated at full power, and one of which is split into two beams called coverage beams, totaling four beams. Each laser

is connected to a Beam Dithering Unit (BDU), which performs optical dithering along the beam path. The BDU rapidly changes the 1.5 mrad deflection angle of the output laser beam, producing a total of eight ground tracks spaced approximately 600 m apart. The total width of the track is 4.2 km, and the average distance between footprints on the track is 60 m. GEDI data products can be categorized into four classes based on the different stages of data processing. The first level, L1, consists of raw and localized waveforms, where L1A represents the raw waveform data directly collected with the sensors, and L1B represents the waveform data processed with GPS and star tracker information for precise localization. The second level, L2, consists of canopy height and profile data at the footprint level, where L2A mainly provides ground elevation and relative height index obtained through waveform processing, while L2B provides canopy cover, leaf area index, and other canopy parameters. The third level, L3, provides gridded products for canopy height indices and variability. These products are derived through spatial interpolation and grid calculations for parameters such as topography, canopy height, leaf area index, and vertical leaf profile. L4 includes aboveground biomass estimation products, with L4A and L4B representing footprint-level and gridded aboveground biomass estimation products, respectively. The GEDI data were accessed on 30 April 2022 and acquired from the NASA data website (https://search.earthdata.nasa.gov/), ensuring temporal alignment with the airborne LiDAR data to maintain consistency across the datasets, the V2 version of the GEDI L1B, and L2A data, which were collected between May 2019 and October 2020 and were downloaded (the V2 version of the horizontal localization accuracy was significantly better than that of V1 version), and the range of the downloaded data was the Gaofeng Forest Farm in Guangxi Province.

2.2.2. Airborne LiDAR

Airborne LiDAR data were collected in February 2018 using a Riegl LMS-Q680i laser scanner within the premises of Gaofeng Forest Farm. At the time of data collection, the aircraft was flying at an altitude of 750 m, at a speed of 180 km per hour, with a lateral overlap of 65%. The sensor captured complete laser pulse return waveforms at 3 ms intervals. The scanner pulse emission frequency was set to 300 kHz, and the scanning frequency was set to 80 Hz. The emission wavelength was 1550 nm, the scanning angle was $\pm 30^{\circ}$, and the field of view was 60°. The beam divergence was 0.5 mrad, and the ground footprint was 37.5 cm. The data showed that the average point spacing in the sample plot was 0.45 m, and the average point density was about 9.58 points/m^2. The final extracted point cloud was stored in LAS 1.2 format [31].

2.2.3. Sentinel-2

Sentinel-2 is a satellite equipped with high-resolution multispectral imaging capabilities. This series of satellites consists of Sentinel-2A and Sentinel-2B. The primary payload of Sentinel-2 is the Multispectral Imager (MSI), which uses a push-scan mode to capture ground data in the 400–2400 nanometer spectral range. This broad range is divided into 13 bands, including the visible (blue, green, red, and near-infrared), the red-edge bands (red-edge 1–4), and the short-wave infrared (Swir1, Swir2). The terrestrial resolution of the bands varies from 10 m, 20 m, and 60 m. The spectral resolution ranges from 15 nm to 180 nm. Spectral resolution ranges from 15 nanometers to 180 nanometers, and the imaging width reaches a staggering 290 km. In this study, Sentinel-2 imagery was processed using Google Earth Engine (GEE) using the "COPERNICUS/S2_SR" dataset from the GEE data catalog and accessed on 20 November 2022 (https://developers.google.com/earthengine/datasets/catalog/COPERNICUS_S2_SR). This dataset represents surface reflectance derived from L2A upper atmospheric reflectance data through radiometric correction and atmospheric correction. GEE screened the Sentinel-2 remotely sensed imagery was set spanning the period from March to September 2019.

2.2.4. Auxiliary Data

This study uses digital elevation modeling products from NASA JPL's Shuttle Radar Topography Mission (SRTM) as topographic data [32]. With a spatial resolution of 30 m, this dataset provides high-quality topographic variables, such as elevation, slope, and aspect. These variables are essential for understanding vegetation growth and distribution, especially in relation to vegetation height analysis [33]. The SRTM data of the study area were retrieved and processed through the GEE platform using the terrain dataset ID "SRTMGL1_003", which was used to compute elements such as digital elevation model, slope, and slope orientation topography. Climatic data, soil data, and land cover data were also used as ancillary data. Climate data were obtained using the WorldClim version 2 climate dataset with a spatial resolution of 1 km [34]. Meteorological elements such as mean annual temperature and mean annual precipitation were extracted from this dataset. Soil data were obtained from the SoilGrids product of the International Soil Reference and Information Center (ISRIC) at 250 m resolution [35]. Key soil factors extracted include cation exchange capacity, water pH, nitrogen content, and organic carbon levels, sourced from ISRIC accessed on 6 October 2022 (https://soilgrids.org/). Land cover data utilized in this research originate from a 30 m resolution product developed by Gong Peng's research team at Tsinghua University [36]. This dataset played a key role in distinguishing between different forest types and non-forested areas. The data were accessed on 6 October 2022 (https://data-starcloud.pcl.ac.cn/zh/resource/). The integration of these datasets facilitated a comprehensive analysis of the study area and supported the understanding of the role of topography, climate, soil, and land cover in influencing forest height information.

2.3. Methods

2.3.1. GEDI Data Processing

To achieve temporal alignment between GEDI data and airborne LiDAR data, the V2 version of L1B was acquired from the LP DACC website (https://lpdaac.usgs.gov/products, accessed on 30 April 2022) for the Guangxi Province Gaofeng Forest Farm study area. Data span the period between May 2019 and October 2020 for L1B and L2A products, respectively. Raw positioning waveforms from the L1B product and canopy height indicators from the L2A product were extracted.

For the L2A data, all GEDI data within the study area were preliminarily cropped using vector boundary data. Forest height values for GEDI footprint measurements were determined using the canopy height percentile 95 (RH95). This metric reflects the height between the ground position and the cumulative energy percentile. Although RH100 theoretically corresponds to the height of the top of the canopy, RH95 was used in this study due to factors such as noise and uncertainty in detecting the location of ground echoes, etc. In order to ensure high-quality L2A data, filtering metrics were set as follows. First, a sensitivity of footprint data greater than 0.9 [25] was used; the sensitivity estimates the relative minimum percentage of echoes for which ground echoes can be detected. This metric was used to measure whether the waveform energy was sufficient to penetrate the canopy and reach the ground [37]. Second, forest footprints with footprint RH95 values exceeding 2 m were selected. Third, footprints with a degradation flagging metric of 0 were used. Fourth, footprints with an absolute height difference of less than 50 m between the ground elevation (elev_lowestmode) and the Shuttle Radar Topography Map (SRTM) DEM were included [24]. Ultimately, 3890 L2A footprints that met the optimizing criteria were retained for further analysis.

For the L1B waveform data, the corresponding L2A footprint data were employed in combination with L1B for parsing. The Quality Flag indicator was extracted from L2A, serving as the quality metric for waveform optimization. The Quality Flag assesses waveform validity based on criteria related to energy, sensitivity, and amplitude. If these criteria are met, the waveform data are considered valid. To ensure compatibility with convolutional neural network learning and eliminate magnitude differences, this study preprocesses the original L1B waveforms. Initially, the original waveform amplitude

is normalized by subtracting the background average noise provided by the GEDI L1B product. This adjustment ensures that the amplitude value of non-existing feature echo waveform frames is approximately 0. Subsequently, all waveforms are standardized to a fixed length of 1420. For waveforms not reaching the required length, zero values are added at the end to reach the specified length. This step is crucial for enabling the neural network structure to learn the maximum height range of the canopy layer, which is 15 cm multiplied by 1420 [38]. Finally, the denoised and filled waveforms undergo normalization, resulting in a dataset of 2429 GEDI waveform records.

2.3.2. Airborne LiDAR Data Processing

The preprocessing of ALS data steps encompassed denoising, filtering, normalization, and the generation of elevation models [39]. Initially, the original point cloud of airborne LiDAR underwent denoising to address issues arising from flying objects and multipath effects, resulting in high and low roughness in the original data. The noise was effectively removed through point cloud clustering, yielding high-quality airborne LiDAR data. Following denoising, the point cloud data underwent filtering using an improved asymptotic irregular triangular mesh encryption method [40]. This method facilitated the separation of ground points from non-ground points. Subsequently, a digital elevation model (DEM) with a 2 m resolution was generated using the ordinary Kriging interpolation method. A Digital Surface Model (DSM) with a resolution of 2 m was then derived through ordinary Kriging interpolation. Finally, a CHM was obtained by performing a difference operation between the DSM and the DEM [41]. For both the DSM and the CHM, a buffer zone with a diameter of 25 m around the GEDI footprint's corresponding location was considered. The values within this buffer zone were utilized as baseline values for forest understory ground elevation and forest height information. This approach ensured a consistent and accurate reference for analyzing the GEDI footprint data in relation to the forest understory characteristics.

GEDI pseudo-waveforms are modeled from airborne LiDAR point cloud data [37]. The forest heights extracted from the simulated pseudo-waveforms were utilized as the true values for tree heights [38,39]. Numerous investigations have explored the transformation of discretized LiDAR point-cloud data into full-waveform LiDAR data [42,43], utilizing a waveform simulation technique developed by Blair and Hofton [34,35]. Our study employs an ALS point contribution weighting-based approach [44], where the laser footprint intensity is transformed into a Gaussian distribution and the contribution of each discrete return frame is weighted based on its distance from the center of the footprint. Parameters for waveform simulation were established in accordance with the GEDI strong-beam laser transmitter specifications. These parameters include σp as -1, pFWHM as 15, σf as 5.5, vertical resolution as 0.15 m, and a simulated maximum number of frames (maxBins) set at 1420. We utilize the Pearson correlation coefficient as a metric to align the positions of the GEDI-normalized waveforms with the simulated waveforms of the airborne LiDAR point cloud. First, we identify the end position of the GEDI waveform and regard this position as the initial point for moving the simulated waveform. Subsequently, we move the simulated waveform frame by frame and calculate the Pearson correlation coefficient between the two waveforms. Once the Pearson correlation coefficient reaches its peak value, we determine this as the optimal matching position of the simulated waveform. Out of the total number of footprints, 1013 footprints passed the Pearson correlation coefficient matching. The footprints were divided into 80% and 20% ratios for the subsequent construction of a 1D convolutional neural network model to estimate the forest height parameter. A total of 1013 simulated footprints were generated, with an 8:2 ratio division into 811 and 202 footprints for subsequent convolutional network training. This simulation approach ensures a comprehensive representation of GEDI waveforms, contributing to more accurate forest height estimations in the study. Figure 3 shows a schematic of an airborne LiDAR point cloud simulating a GEDI waveform using a GEDI simulator. Figure 3(a1–a3) depict

schematic diagrams of typical point clouds for low tree heights, while Figure 3(b1–b3) show schematic diagrams of typical point clouds for high tree heights.

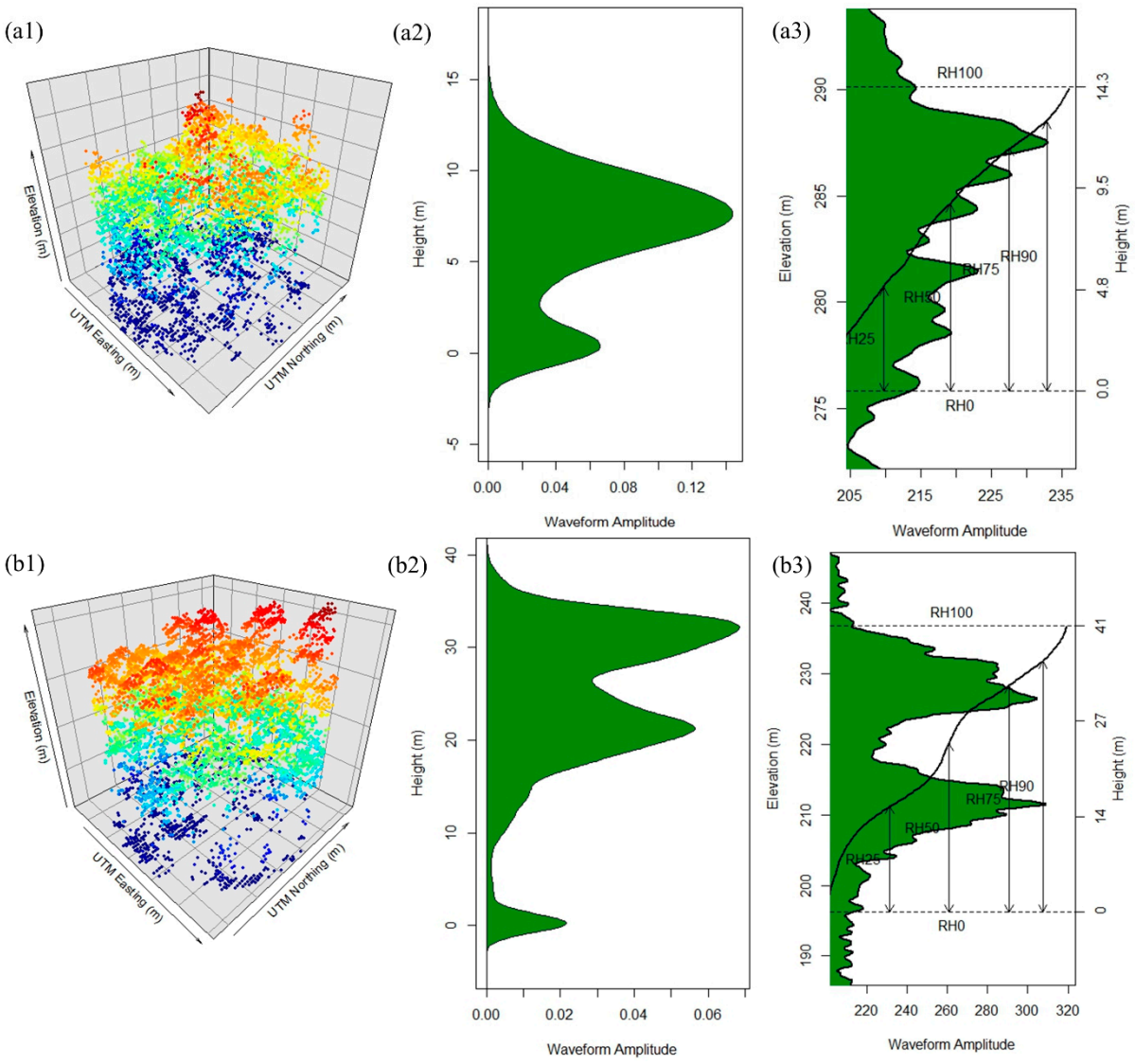


Figure 3. The GEDI waveforms are simulated through the GEDI simulator using an airborne LiDAR point cloud. Part a and part b represent the typical point cloud of low tree height (around 10 m) and the typical point cloud of high tree height (around 25 m), respectively. (**a1**,**b1**) Represent the 3D view of the point cloud corresponding to the 25 m resolution footprint; The point cloud color ranges from dark blue to dark red representing low to high heights; (**a2**,**b2**) represent the waveform obtained from the simulation of the point cloud, from which height information can be extracted; (**a3**,**b3**) represent the waveform corresponding to the extraction of the GEDI footprint.

2.3.3. Sentinel-2 Processing

GEE was used to screen the Sentinel-2 remotely sensed imagery set spanning the period March through September 2019. Cloud coverage was assessed using the method of calculating cloud fraction with a threshold set at 30. The image set was iteratively traversed and masked to exclude areas affected by cloud cover. Ultimately, an average value for each band of high-quality Sentinel-2 imagery was calculated to minimize the effects of cloud cover. In the realm of remote sensing, vegetation indices serve as vital tools for both qualitative and quantitative evaluations of vegetation cover and growth vitality. In this study, 12 commonly used vegetation indices were calculated from Sentinel-2 images, and

the selected vegetation indices are detailed in Table 1 [45]. In addition, six texture features were extracted from Sentinel-2 images. These features, including texture contrast (Contrast), texture entropy (Entropy), and texture variance (Variance), were obtained from the Sentinel-2 image by creating a grayscale image using visible light wavelengths. Principal component analysis was conducted on the Sentinel-2 image to extract five bands, encompassing 99.5% of the information from the original image as feature bands.

Table 1. Table of formulas for calculating the vegetation index.

Vegetation Index	Formula	Reference
Anthocyanin Reflectance Index (ARI)	$(1/Green) - (1/RedEdge1)$	[46]
Chlorophyll Index Green (CIG)	$(NIR/Green) - 1$	[47]
Chlorophyll Index Red Edge (CIRE)	$(NIR/RedEdge1) - 1$	[47]
Chlorophyll Vegetation Index (CVI)	$(NIR \times Red)/(Green \times Green)$	[48]
Enhanced Vegetation Index (EVI)	$2.5 \times (NIR - Red)/(NIR + 6 \times Red + 7.5 \times Blue + 1)$	[49]
Two-Band Enhanced Vegetation Index (EVI2)	$2.5 \times (NIR - Red)/(NIR + 2.4 \times Red + 1)$	[50]
Modified Soil-Adjusted Vegetation Index (MSAVI)	$0.5 \times (2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - Red)}$	[51]
Moisture Stress Index (MSI)	$(S1/NIR)$	[52]
Normalized Difference Vegetation Index (NDVI)	$(NIR - Red)/(NIR + Red)$	[50]
Normalized Difference Vegetation Index (705 and 750 nm) (NDVI705)	$(RedEdge2 - RedEdge1)/(RedEdge2 + RedEdge1)$	[53]
Ratio Vegetation Index (RVI)	$RedEdge2/Red$	[54]
Soil-Adjusted Vegetation Index (SAVI)	$1.5 \times (NIR - Red)/(NIR + Red + 0.5)$	[55]

2.3.4. Convolutional Neural Network Construction

CNN is a deep learning architecture primarily employed for image processing tasks, yet its applicability extends to other domains like natural language processing and sound analysis. CNN has demonstrated effectiveness in processing both one-dimensional (e.g., time series) and two-dimensional (image) signals [56]. The fundamental building blocks of this neural network architecture are constructed through convolutional operations, which are particularly suitable for signal data exhibiting spatial or temporal autocorrelation. Given that each GEDI footprint consists of a univariate waveform signal, a 1D CNN is aptly suited for learning autocorrelation features that are inherent in the waveform.

The 1D CNN can be directly applied to raw GEDI data when considering the GEDI signal as a wave signal [57]. Figure 4 illustrates the current design of the 1D-CNN network structure tailored for processing GEDI waveforms to estimate forest height information parameters. In this convolutional neural network, filters with a kernel size of 3 are utilized. The convolution operation consists of a nonlinear activation function (rectified linear unit ReLU), a MaxPooling layer, and a Dropout layer [58]. The number of filters increases with the network structure. The 1D-CNN executes several convolutional block operations, followed by a Flatten block that unfolds features extracted with the convolutional kernel. Subsequently, all features are combined using a fully connected layer, and regression is performed using an output neuron to estimate the forest height parameter. For waveform training, 70% of the samples are divided as the training dataset, and another 30% of the samples are used as the test dataset. Model parameter fitting is performed in the training set, and the validation set is used to evaluate the accuracy of the model. The Adam optimizer was employed to minimize the loss, and the mean square error (MSE) was chosen to optimize the cost function of the neural network model. The Adam optimizer, renowned for its efficiency, contributed to loss minimization, while the mean square error was instrumental in optimizing the neural network model's cost function.

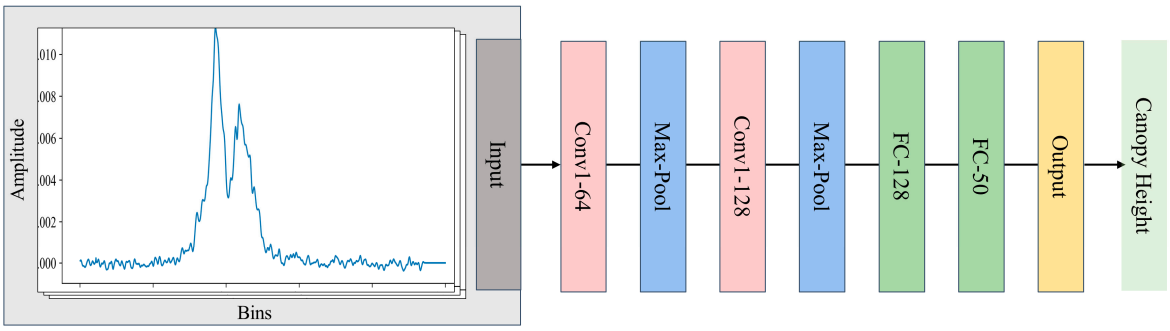


Figure 4. One-dimensional convolutional neural network model framework diagram.

2.3.5. Forest Height Extrapolation Model

Random Forest Algorithm

The RF algorithm combines the Bagging integrated learning theory and the random subspace approach, and is a machine learning algorithm proposed by Leo Breiman in 2001 [59]. The RF model uses a decision tree as the base classifier. Decision trees are widely used as a tree classifier, where each node in the tree keeps classifying by selecting the optimal split features until the termination condition for building the tree is reached. It works by constructing many decision trees during training, with no association between each decision tree in the forest.

RF regression is the process of generating many decision trees through the modeling dataset of sample observations and characteristics of the variables, which are randomly sampled. Each sampling result is a tree, and each tree generates rules and judgment values in line with their own attributes. Alongside the establishment of the final integration of all the decision trees of the rules and judgment values, to achieve the regression of the RF algorithm. The GEDI height parameter was used as the dependent variable, and feature variables such as vegetation index, terrain index, and texture features were used as the independent variables. Multiple subsets of bring-back data were constructed from the original dataset with random sampling, and each subset randomly selected a part of the features from all the features as the inputs, and then used the CART algorithm to generate the decision trees, and then averaged the predicted values of each decision tree to complete the regression modeling. Regarding the RF model parameters, maximum iterations were set to 50, n_estimators was set to 96, max_depth was set to 8, max_features was set to 5, min_samples_split was set to 4, and min_samples_leaf was set to 1.

Artificial Neural Network

ANN is a mathematical model inspired by the behavioral traits of animal neural networks. They implement distributed parallel information-processing algorithms, utilizing the system's complexity to adapt interconnections among internal nodes for information processing. The ANN exhibits characteristics such as nonlinearity, nonlocality, and nonconstancy. It is composed of multiple layers of neurons, and they accomplish regression tasks by learning intricate non-linear mapping relationships between inputs and outputs [60].

The GEDI height parameter serves as the dependent variable, while the independent variables used to construct the model dataset include the vegetation index, terrain index, texture features, and cofactors. The neural network model dataset is structured with defined input, hidden, and output layers, each specifying the number of neurons within. Then, the weights and biases of the neural network are initialized and the inputs in each neuron are weighted and summed, followed by an activation function to get the outputs. Finally, the outputs are processed, the MSE is used to compare the difference between the prediction and the actual output of the neural network, and then the gradient of the loss function for each parameter (weights and bias) is calculated by the back-propagation algorithm to obtain the direction of adjustment of the parameters. The iterations are repeated after updating the parameters until the loss function converges or a predetermined number of

iterations is reached. Finally, the neural network model is evaluated using a test dataset to check the predictive performance of the model and to predict the regression results on new input data. Regarding the parameters of the ANN model, net.trainParam.epochs is the number of training times. It is set to 1000 times in the code. net.trainParam.lr is the learning rate. It is set to 0.01 in the code. net. trainParam.goal is the minimum error of the training target, set to 0.000001 in the code.

2.4. Accuracy Assessment

In this study, R^2, RMSE, and MAE were used to evaluate the accuracy of the inversion results. The coefficient of determination, R^2, reflects the extent to which the regression equation explains the dependent variable. The root mean square error (RMSE) is calculated as the square root of the ratio of the square of the deviation of the predicted value from the true value to the number of observations (n), reflecting the difference between the predicted value and the true value. The mean absolute error (MAE) can effectively reflect the actual prediction error. The formulas for R^2, RMSE, and MAE are given below:

$$R^2 = 1 - \frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}{\Sigma_{i=1}^{n}(y_i - \overline{y})^2} \quad (1)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (2)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

where y_i is the true value of forest tree height in the study area obtained from GEDI measurements, $\hat{y}_i$ is the inverted forest height in the study area, and $\overline{y}$ is the mean forest height over all forest height measurements.

3. Results

3.1. GEDI L2A Product Extraction Accuracy

In this part of the study, Gaofeng Forest Farm Jiepai Dongsheng Sub-farm was selected as the study area, and airborne LiDAR data were used as the ground reference to compare and analyze the extraction accuracy of spatial LiDAR GEDI. It aims to understand the extraction accuracy of GEDI for understory elevation and forest height, and the main evaluation indexes include R^2, RMSE, and MAE.

Between April 2019 and October 2020, a total of 3890 GEDI footprints were collected within the Gaofeng Forest Farm, of which 928 fell within the sub-field and airborne LiDAR point cloud data coverage. The forest height data used for accuracy comparisons were obtained from the GEDI L2A product using RH95 as the forest height parameter obtained from the GEDI measurements. Six different values were obtained for each height parameter based on six different built-in algorithm sets, and out of the six parameters, four parameters with non-maximum and minimum values were retained, and the average of these four parameters was taken as the value of RH95.

The accuracy of ground elevation extracted from all GEDI footprints within the Gaofeng forest sub-farm has a high accuracy (R^2 = 0.99, RMSE = 4.99 m, and MAE = 3.97 m (Figure 5a)). After filtering the GEDI footprints with filtering conditions, the accuracy of the extractions was improved. The RMSE was reduced by 45.5% and the MAE by 43.4% after optimizing. These results indicate that setting the filtering conditions can effectively improve the accuracy of GEDI L2A in measuring understory elevation.

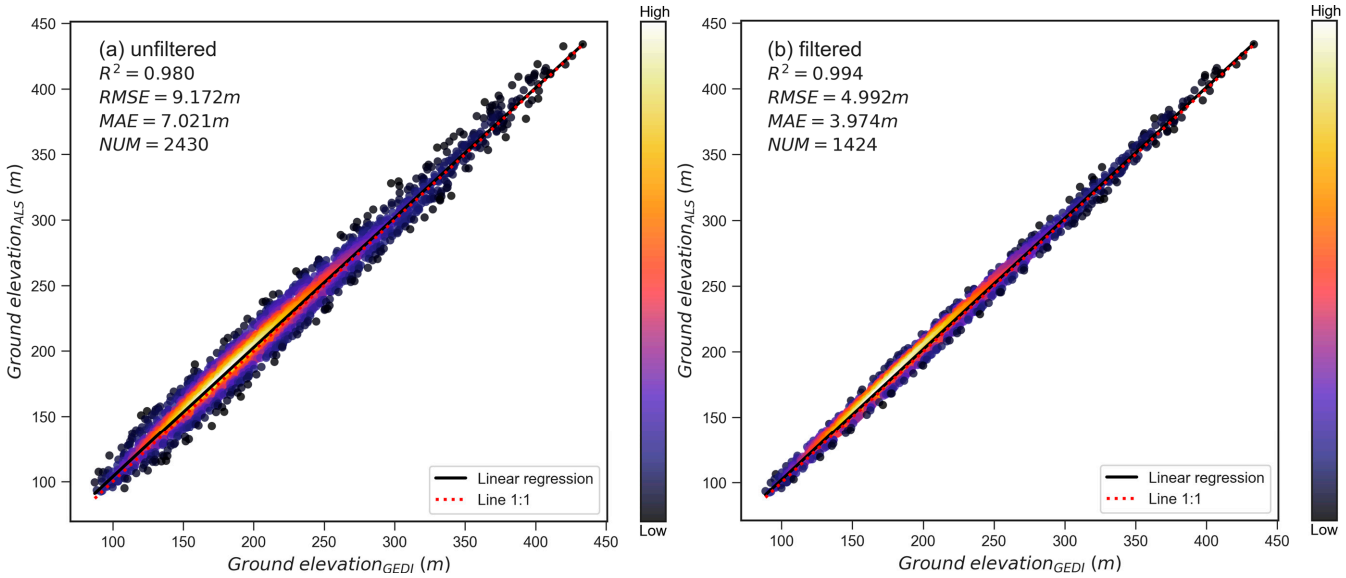


Figure 5. Scatter plot between ground elevations extracted using GEDI L2A and ground elevations extracted using airborne LiDAR. (**a**) GEDI data without parameter filtering. (**b**) GEDI L2A data filtered with parameters.

The accuracy of GEDI L2A extracted forest height information before and after filtering was evaluated by comparing the GEDI-extracted forest height information with the airborne LiDAR-extracted CHM height baseline values in the experimental area of the sub-field (Figure 6). Before filtering, the accuracy of the forest height information extracted from the GEDI footprints in the subfield was poor (R^2 = 0.11, RMSE = 7.41 m, MAE = 6.06 m (Figure 6a)). After applying the filtering conditions to screen the GEDI footprints, the accuracy of the footprints improved (R^2 = 0.42, RMSE = 5.17 m, MAE = 4.07 m (Figure 6a)). The R^2 increased by 0.31, the RMSE decreased by 30.3%, and the MAE decreased by 32.9%. These findings suggest that filtering conditions have an enhancing effect on improving the extraction of forest height information from GEDI footprints within the subfield area.

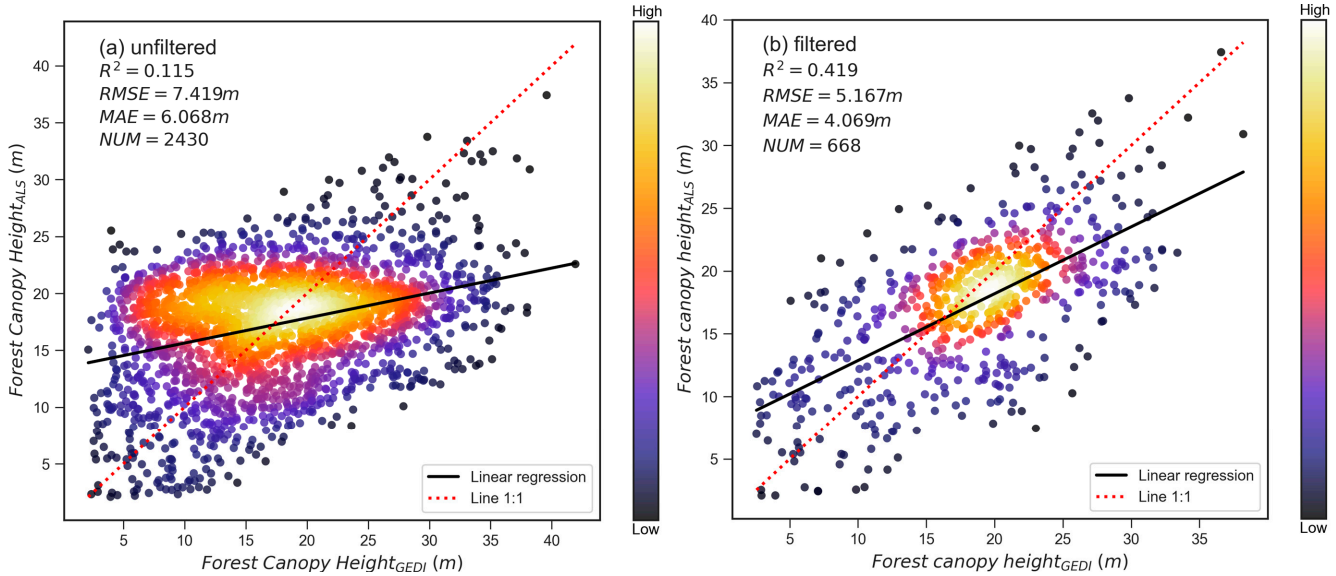


Figure 6. Scatter plot between forest height extracted using GEDIL2A and canopy height extracted using airborne LiDAR. (**a**) GEDI data without parameter filtering. (**b**) GEDIL2A data filtered with parameters.

3.2. GEDI L1B Waveform Calibration Model Accuracy

In this part of the study, the GEDI L1B waveforms within the sub-field area were used as inputs for model training, and the height values extracted from the corresponding airborne LiDAR point cloud pseudo-waveforms were used as outputs of the model. The method uses a 1D CNN to construct the model, and then utilizes the waveforms of footprints within the entire Gaofeng Forest Farm (outside the sub-field area) as input data. The trained model estimates the forest height values of these footprints and calibrates the original height parameters.

Figure 7 illustrates the accuracy results of the model on the validation dataset. Figure 7a on the left shows the forest height information extracted with L1B has a high accuracy (R^2 = 0.84, RMSE = 3.13 m, MAE = 2.43 m (Figure 7a)). Compared to the optimized L2A data, the RMSE was reduced by 2.03 m and the MAE by 1.64 m. In Figure 7b, the horizontal axis represents the number of training iterations (epochs), and the vertical axis represents the loss value of the training data. By observing the graphical trend, one can clearly determine whether the model is converging or not, and how quickly it is converging. The model reaches a steady state after 60 iterations, and demonstrates improved convergence performance.

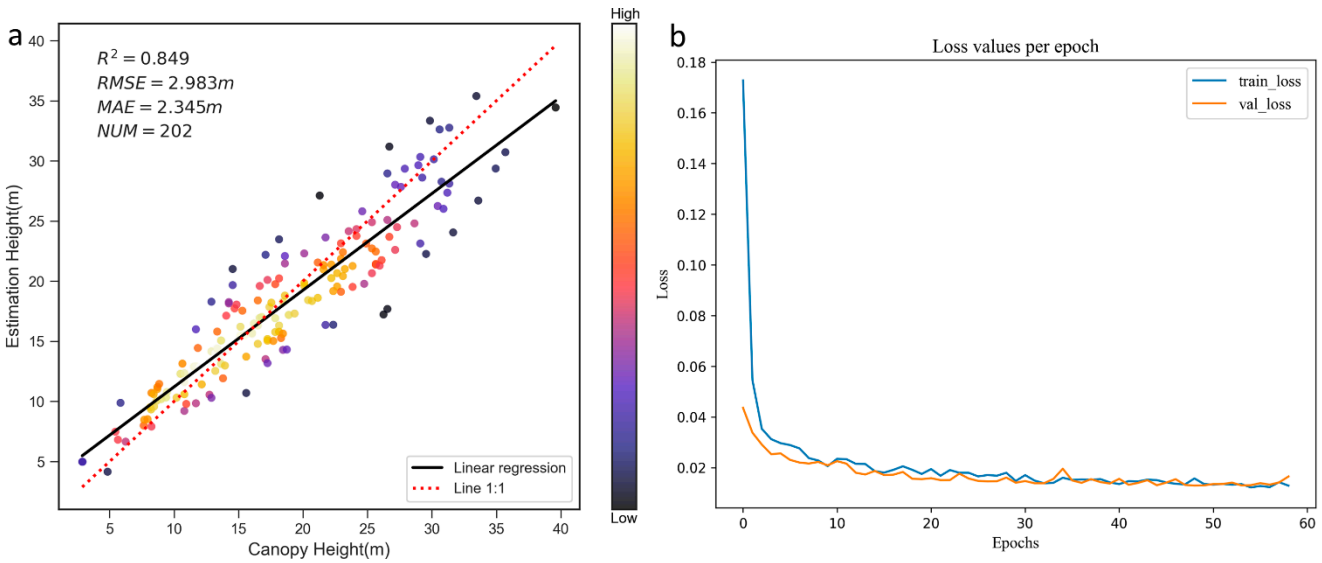


Figure 7. 1D CNN model estimation accuracy validation. (**a**) Shows a scatter plot of the accuracy between the predicted canopy height and the reference canopy height extracted after calibrating the waveforms. (**b**) Shows the training loss map for the forest height inversion model.

In the process of inverting forest height information based on GEDI L1B waveform data, factors such as slope, solar elevation angle, and beam type had an impact on the accuracy of the forest height inversion results [53], and we analyzed the impact of different factors on canopy height estimation, respectively. Slope data were extracted from the SRTM digital elevation model product, with the majority falling within the 0–40° range in the study area. To facilitate analysis, the slope was divided into four categories at 10° intervals. Evaluation metrics, including R^2 and RMSE, were employed to assess the results. Figure 8a illustrates that the forest height estimation error is minimized when the slope ranges from 0 to 10° (R^2 = 0.58, RMSE = 3.93 m, num = 269). As the slope increases, the R^2 decreases while the RMSE increases, indicating reduced accuracy. The least accurate estimations occur within the 30–40° slope range (R^2 = 0.15, RMSE = 5.64 m, num = 58). Figure 8b reflects the distribution of the differences between predicted and control values under different slope conditions, with smaller height differences in the region of slopes with slower gradients of 0–10°, and an increasing range of differences with increasing slope.

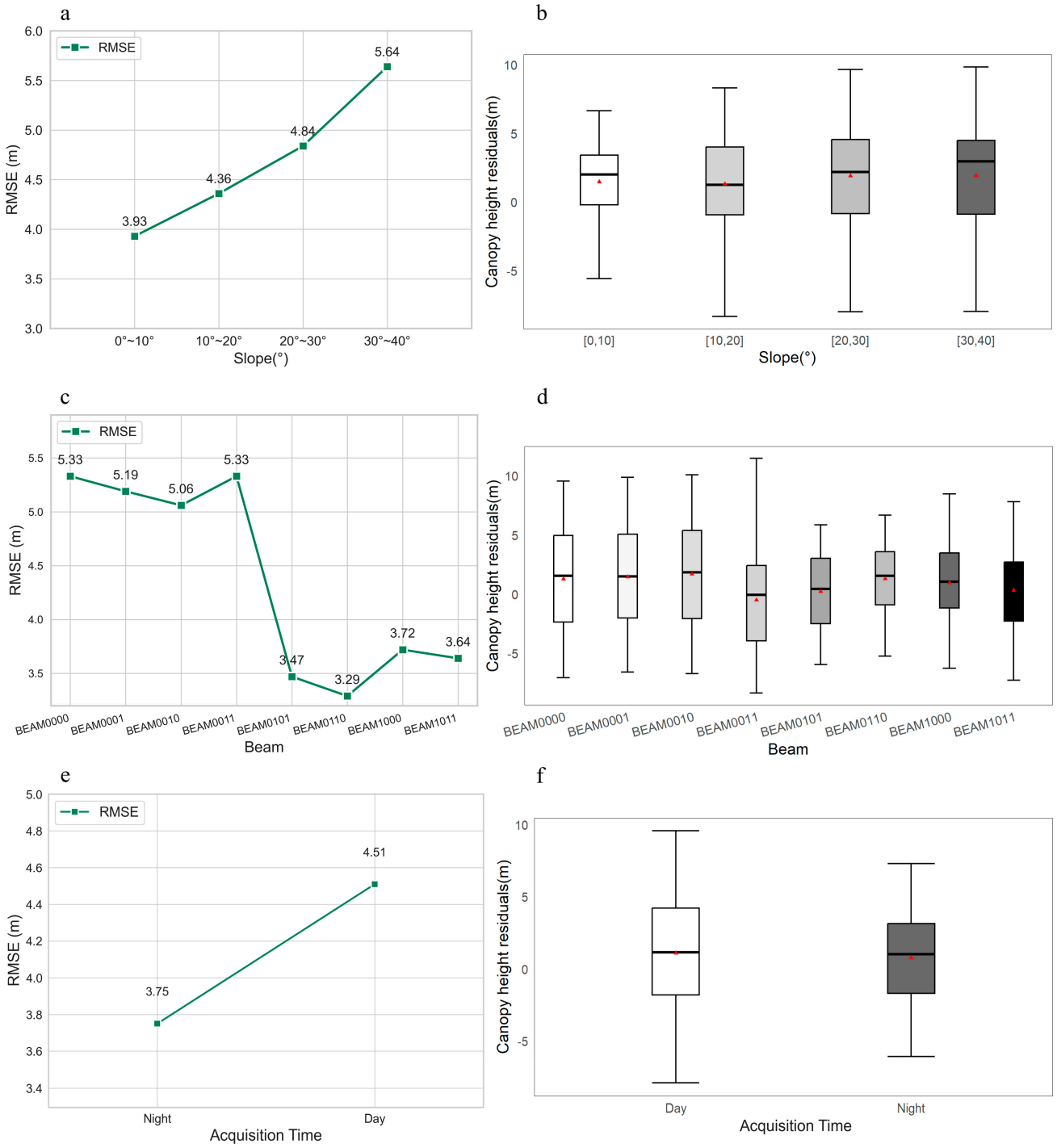


Figure 8. Plots of error trends and distribution of prediction errors for forest height inversion results for different slope, beam type, and solar elevation angle conditions. Box plots show the mean (red triangles), median, and upper and lower quartiles. Positive difference values indicate that the predicted value is greater than the reference value.

GEDI uses three lasers to transmit the laser beams and receive the return signals to complete the ground measurements, with two lasers operating at full power and one laser splitting into two coverage beams at half power, which are then passed through the beam dithering unit to form eight tracks, with BEAM0000, BEAM0001, BEAM0010, and BEAM0011 representing the average-power tracks, and BEAM0101, BEAM0110, BEAM1000, and BEAM1011 representing the tracks for the full-power beam mode, and R^2 and RMSE

are used as the evaluation metrics according to the division data of beam types. The results in Figure 8c show that for beams in strong power mode, BEAM0110 has the smallest forest height estimation error (R^2 = 0.67, RMSE = 3.29 m, num = 91). BEAM1000 has the largest forest height estimation error (R^2 = 0.56, RMSE = 3.72 m, num = 83). For beams in average power mode, BEAM0010 has the highest forest height estimation accuracy (R^2 = 0.39, RMSE = 5.06 m, num = 108). BEAM0011 has the lowest forest height estimation accuracy (R^2 = 0.32, RMSE = 5.33 m, num = 101). There is a more pronounced difference in RMSE between the predicted forest heights and the reference values for the beams under the strong power and the average power, and the distribution of the differences in the prediction of forest heights is reflected in Figure 8d. For the solar elevation angle factor, the solar elevation greater than 0° was taken as daytime and the solar elevation less than 0° was taken as nighttime, and R^2 and RMSE were used as evaluation metrics. The results show that the accuracy of forest height estimation during daytime is lower (R^2 = 0.45, RMSE = 4.51 m, num = 478 (Figure 8e)). The accuracy of forest height estimation during nighttime is higher (R^2 = 0.62, RMSE = 3.75 m, num = 333 (Figure 8e)). Figure 8f reflects the distribution of the difference between predicted and control values under daytime and nighttime conditions, with a larger range of distribution of the difference in the daytime than in the nighttime.

3.3. Inverse Mapping of Forest Height Information

The GEDI spaceborne LiDAR system provides valuable information about the vertical structure of forests. Collecting data in the form of individual footprint observations at specific intervals, the spatially discontinuous nature of GEDI data requires integration with other optical remote sensing data to achieve precise and continuous estimates of vegetation parameters, particularly forest height. Mapping forest height information is crucial for effective forest monitoring.

After removing outliers, the forest height was estimated using the RF model and the ANN model, respectively. For the RF regression model, the model parameters mtry and ntree are set to 200, the parameter mtry indicates the number of input variables randomly selected for each segmentation, and the parameter mtree indicates the number of decision trees generated, and the model is trained and evaluated for the importance of feature variables. From the results in Figure 9b, it can be concluded that the most important factors of the image RF model are mainly CIRE, NDVI, and Slope, and then the canopy height inversion model was constructed using the top five feature variables in terms of importance, and the accuracy of the RF modeling R^2 was 0.64 after training (Figure 10b). In the artificial neural network model, the input variables and output variables used to construct the model are consistent with the RF model. After training, parameter adjustment, and model construction, the R^2 of the artificial neural network model is 0.49 (Figure 10a), and then the factor importance of the constructed artificial neural network is evaluated, and the results show that the factors ARI, RVI, MSI, MNDVI, and NDVI contribute the most to the artificial neural network. Figure 10 shows the accuracy of the RF model compared to the ANN model, with RMSE ranging from 4.59 to 5.07 m and MAE ranging from 3.82 to 4.29 m, highlighting the consistency of the RF algorithm between the predicted and validated heights. The RF regression algorithm had the highest accuracy in extrapolation modeling of the study area. Therefore, the RF regression algorithm was selected for regression modeling of GEDI data and then the reflective mapping of forest height information across the forest farm.

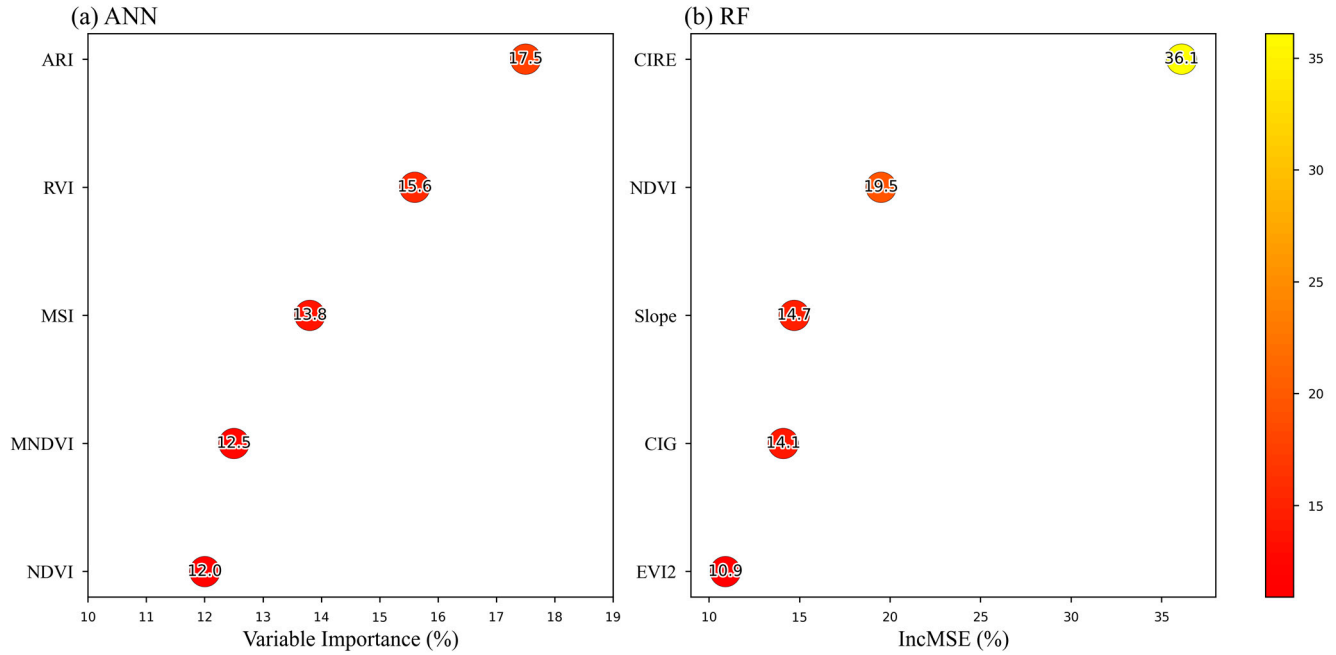


Figure 9. The graphs of feature importance analysis of the artificial neural network model and RF model. (**a**) Graph shows the result of feature importance of the ANN model, and (**b**) graph shows the result of feature importance of the RF model.

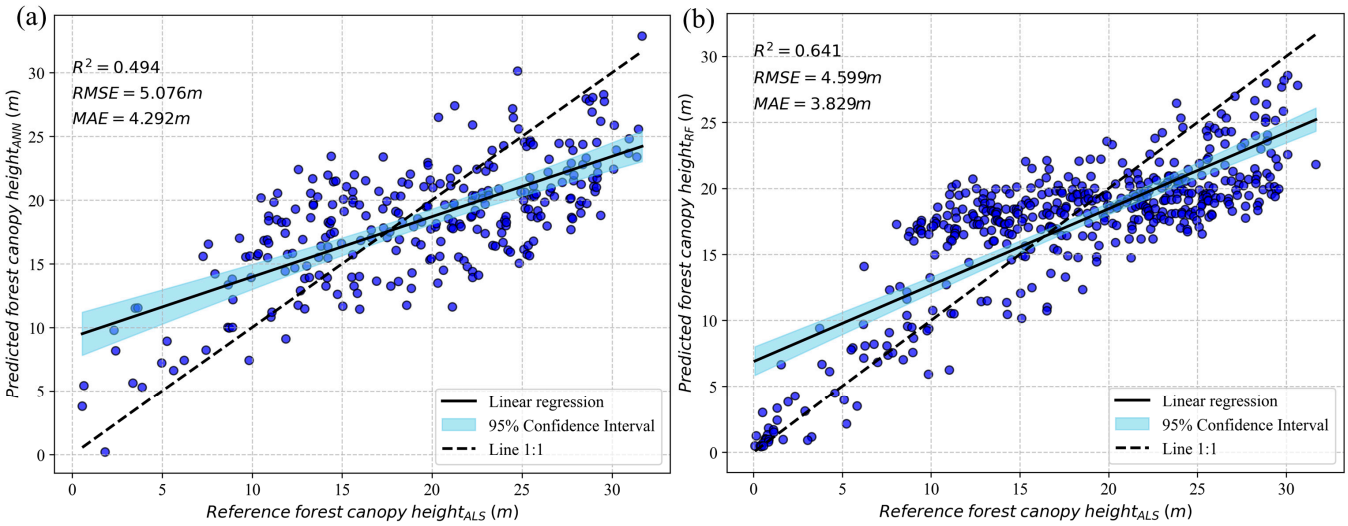


Figure 10. Accuracy scatterplot of height regression results between the artificial neural network model and the RF model.

In conclusion, the RF regression model outperformed the ANN model in the forest height inversion at the stand scale in the Gaofeng Forest Farm. Therefore, we used the optimized GEDI L1B forest height information in combination with the Sentinel-2 optical image data features and topographic and climatic variables to estimate the forest height of the complete area of the Gaofeng Forest Farm at the upscaled scale based on the RF model.

Figure 11 shows the high variability in the height distribution of the Gaofeng forest farm estimated based on the RF algorithm. A small number of predicted tree heights between 0 and 12 m may represent felling or reforestation areas dominated by grassland and low small trees. The largest proportion of tree heights falls between 14 and 22 m, displaying a generally normal distribution. Notably, tree heights ranging from 17 m to 20 m have relatively higher proportions, reaching 18.25%, 16.03%, 13.60%, and 12.55%,

respectively. This segment likely corresponds to forest plantation areas comprising fast-growing eucalyptus and cedar.

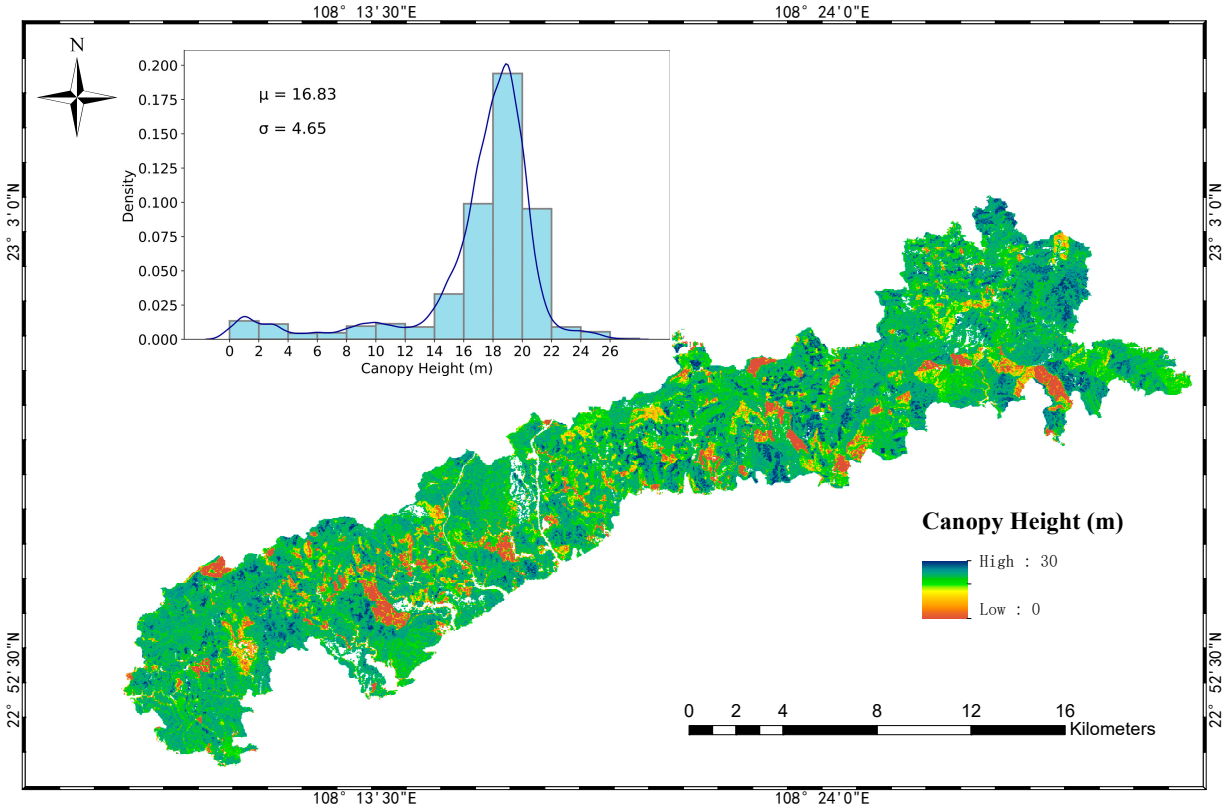


Figure 11. Spatial distribution of forest heights in the study area based on the inversion of the RF method. The Table in the Figure shows the frequency distribution of predicted heights in the study area.

4. Discussion

4.1. Analysis of Factors Affecting Understory Elevation and Forest Height for GEDI Extraction

In this paper, the accuracy of understory elevation and forest height measured with GEDI was first evaluated using GEDI's L1B and L2A footprint data and airborne LiDAR data, and the results of the evaluation showed that GEDI's ability to estimate understory elevation was relatively reliable, and the accuracy of estimating forest height was less than ideal. For the L2A data, the canopy height model extracted with the airborne LiDAR was used as the reference value for forest height information, and the accuracy was improved by filtering using sensitivity, relative height thresholds, degrade flag, and the difference between ground elevation products. The sensitivity reflects the relative minimum percentage of returns that can be detected using the ground return [24], which is able to detect whether the waveform has enough energy to penetrate the canopy to reach the ground [37], and the forest height data extracted from invalid waveforms can be effectively removed by filtering the footprints with sensitivity greater than 0.9. Filtering footprints with relative height RH95 greater than 2 m can remove non-forested areas such as logged areas and roads, and reduce the influence of non-forested areas on forest height estimation. The degrade flag indicator indicates the state information of the pointing and positioning information of the spaceborne LiDAR, and the filtering indicator of 0 can remove the footprints that have problems with the pointing and positioning information, and the error data can be removed. The difference between the ground elevation products refers to the footprints where the absolute difference between the ground elevation acquired with GEDI and the DEM height of SRTM is less than 50 m [25], and the abnormal height footprints caused by factors such as the atmosphere or clouds can be effectively removed. The results in

Figure 6 show that the R^2 is improved from 0.11 to 0.41 and the RMSE is 7.41 m reduced to 5.16 by filtering, and that the R^2 is improved by 0.31 and the RMSE is decreased by 30.3%. Therefore, reasonable selection of condition filtering is a method to improve the accuracy of GEDI understory elevation and canopy height estimation. Wang's study focused on forest height accuracy validation of GEDI L2A and airborne LiDAR data from 33 study areas with different stand conditions in the U.S.A. (R^2 = 0.96, RMSE = 2.62 m) [61], while Liu's study on GEDI in China region forest height accuracy validation resulted in R^2 and RMSE of 0.40 and 6.72 m, respectively, which proved that the estimation errors in China region were generally larger than the GEDI estimation errors in the U.S. study area, and thus the accuracy of canopy height needs to be further improved.

We investigated the influence of slope, beam type, and solar elevation on the accuracy of the results after the GEDI calibration. As the slope increases, the R^2 between the predicted and reference values of forest height decreases and the RMSE increases, with the highest accuracy in gently sloping areas (slopes less than 10°) and the lowest estimation accuracy in steeply sloping areas (slopes greater than 30°), and the range of differences increases with the slope, with the predicted values being about 2 m higher than the control values in general. The reason for this phenomenon may be that the complex terrain areas with high slopes lead to a more complex reflection of the GEDI laser on the surface, which makes the intensity distribution of the laser, as well as the reflection time, biased, resulting in changes in the reflection characteristics and intensity of the laser signals, which in turn affects the accuracy of the extraction of forest height information. Regarding the beam type, the forest height estimation in the strong beam mode at full power has a smaller error compared to the covered beam mode at average power. The reason for this phenomenon may be that the footprint energy of the strong beam mode at full power is stronger than that of the coverage beam mode at average power. On the one hand, the return signal is more difficult to be interfered by external factors, and on the other hand, the stronger emission energy represents a stronger canopy penetration ability, which allows for the measurement of more forests with high depressions [62]. Regarding the solar elevation, the estimation at night has a better accuracy compared to the estimation during the day, which may be caused by the fact that the solar returns during the daytime lead to more solar background noise in the signal, whereas the night is less affected by the solar returns and thus has a lower signal-to-noise ratio of the waveform [63].

4.2. Analysis of Forest Height Inversion for Convolutional Neural Networks

The average horizontal offset geolocalization error present in the GEDI data is around 10 m. Matching the GEDI waveforms with the corresponding pseudo-waveforms of the ALS data based on the footprint position, and then extracting the corresponding height metrics from the pseudo-waveforms, the study has shown that extracting the height metrics in this way has a higher accuracy than direct extraction [64]. The method used in this study relies on weighting the contribution of ALS footprints [44]. Its advantage lies in addressing the intensity variation across laser spots. By modeling the laser footprint intensity as a Gaussian distribution and weighting the contribution of each discrete return frame based on its proximity to the footprint center, this approach offers a more precise representation of LiDAR observation conditions. The GEDI strong-beam laser transmitter parameters are used to set the spot waveform simulation parameters, where σp is set to -1, pFWHM is set to 15, σf is set to 5.5, and the vertical resolution is set to 0.15 m to ensure that the simulation results have a vertical sufficient accuracy in the vertical direction, and the simulated maximum frame number maxBins is 1420 to ensure a consistent range of simulated waveforms. The L1B waveform and ALS pseudo-waveform extraction parameters are used as inputs and outputs of the model, and the modeling method uses a 1D CNN to calibrate the original height parameters of the footprint, and Figure 7a shows the results of the accuracy of the model (R^2 = 0.84, RMSE = 3.13 m, MAE = 2.43 m), which indicates that the model error is small and the prediction result is relatively close to the real value. The reason may be that the 1D CNN can effectively capture local features

in the signal, and identifying local patterns containing temporal or spatial information through convolutional operations can help to learn important features of the waveform. Convolutional kernel sliding over the entire waveform can learn repeated patterns or common features in the waveform. This parameter sharing reduces the complexity of the model and also reduces the risk of overfitting. In addition, 1D CNN can automatically learn the features of the input waveform data, which reduces the manual feature engineering, and stacking multiple convolutional layers can abstract the high-level features to better train the data. While our study improved the accuracy of the forest height inversion results with the help of the calibration algorithm, no further research was conducted regarding the potential impact of the calibration algorithm on the otherwise high-quality data. This will guide our future work to refine our algorithm and maintain high data quality standards.

The accuracy of forest heights derived from the GEDI L1B inversion was significantly higher than those extracted from the L2A product. This difference can be attributed to several factors. First, the applicability of different percentile height metrics in various scenarios affects the accuracy. At the beginning of the study, we compared the accuracy of different percentile height metrics of GEDI with the CHM extracted with the airborne LiDAR. We found that RH95 exhibited the best performance, with an RMSE of 5.16 m. Second, using the statistical maximum of the CHM extracted from the airborne LiDAR data as a reference value for the range of GEDI points may introduce bias. The GEDI points represent a 25 m diameter forested area. Relying solely on the statistical maximum as the true altitude is not comprehensive enough, and may result in systematic bias. The GEDI L2A product has been standardized to provide ground and forest height information that is broadly applicable. Standardized processing may compromise accuracy in specific application scenarios, like when extracting heights of plantation forests in southern China without making specific adjustments. In contrast, forest height data inverted from the L1B product using CNN algorithms can be more customized to the specific conditions of a particular region, thereby enhancing measurement accuracy. Finally, the geographic location bias of the GEDI product is also an important factor. The geographic location error in the L2A product remains uncorrected, while the L1B product has been corrected. This partially explains the higher accuracy of the inverted heights derived from the L1B product. Future research could delve deeper into these factors to enhance the accuracy performance of GEDI data in different application scenarios.

4.3. Analysis of the Inversion Results of the RF Model and ANN Model

In this study, the RF model is more appropriate compared to the ANN model. The modeling accuracy of the RF model was $R^2 = 0.64$, and that of the artificial neural network model was $R^2 = 0.49$. Then, the models were analyzed for the importance of the characteristic variables to understand the degree of contribution of the variables to the prediction model, and for the artificial neural network model. Figure 9a shows the five variables with the highest importance of the characteristic variables, and the highest importance of the characteristic variables was ARI, followed by RVI and MSI. This indicates that ARI, RVI, and MSI play a key role in estimating forest height in the ANN model, and that the neural network learns in a way that is highly sensitive to atmospheric and vegetation reflections. Figure 9b shows the five variables that contribute the most to the RF model, using %IncMSE to indicate the importance of the predictor variables [65]. The %IncMSE of CIRE is the largest, followed by NDVI and Slope, which suggests that CIRE, NDVI, and Slope are more important for the modeling of the RF model, and it is possible that when the decision tree of the RF is constructed, these features are easier to separate and more interpretable, thus having a greater impact on forest height estimation. A comparison of modeling accuracy results can be seen based on the RF regression algorithm to construct the model that has the highest accuracy (Figure 10). The use of calibrated footprint data based on the RF algorithm is used to construct a regression model for the upscaling inversion of the forest height. The RF model is a nonparametric regression method. The RF algorithm is used to construct the complex nonlinear relationship between the dependent variable and the predicted

features. The RF consists of multiple decision trees, which can synthesize the advantages of a large number of decision trees to fit the nonlinear relationship to improve the accuracy of the estimation, taking into account the fact that the feature variables come from multiple bands of the GEDI and the Sentinel-2, as well as the multidimensional information such as climate and terrain. The randomness characteristic of the RF decision tree can enable it to effectively deal with the multidimensional data, and the RF can effectively deal with the multidimensional data by random selection of features and samples, coupled with the voting mechanism of many decision trees, which can reduce the risk of overfitting and improve the generalization ability of the model. The ANN model is prone to transition learning noise when dealing with smaller datasets, and is less resistant to overfitting than the RF method. Therefore, we upscaled the inversion and mapped the distribution of forest height information in the summit forest at 30 m resolution based on the RF regression model. The accuracy and effectiveness of model upscaling in our study are influenced by the resolution of the input data (Sentinel-2 imagery), the quality of the data, preprocessing steps such as noise reduction and geolocation calibration, as well as model selection and parameter settings. Additionally, the size and diversity of the training dataset are critical for the model to generalize across various forest types and conditions. While our research focuses on enhancing upscaling accuracy, assessing model uncertainty is also crucial. Future research will incorporate uncertainty quantification techniques, such as Monte Carlo simulations, to better assess the uncertainties associated with our upscaling models. This will enhance the robustness and applicability of our findings in various forest environments.

In addition, we analyzed the accuracy in comparison with other related studies. The accuracy of the forest height map in this paper (R^2 = 0.62, RMSE = 4.60 m) was compared with the accuracy of the 2019 forest height mapping at 30 m resolution in China by Liu [66] (R^2 = 0.60, RMSE = 4.88 m), the accuracy of the 2019 global forest height mapping by Potapov [25] (R^2 = 0.61, RMSE = 9.07 m), and the accuracy of Lang's 2019 global forest height mapping (RMSE = 4.4 m) [30]. The results of this paper are close to those of Liu's, with little difference between the heights of dominant trees and major forests, and Liu's study mainly used Sentinel-2 images combined with GEDI and ICESat-2 for the inversion of forest heights at the national scale, because on the one hand, both studies used Sentinel-2 and GEDI as the data sources, and on the other hand, Liu's study was at the national scale while the other two scholars' studies were at the national scale. On the other hand, Liu's study is at the national scale, while the other two scholars' studies are at the global scale, and the scale effect also affects the accuracy. It can be seen that the applicability of forest height mapping at the global scale or even the national scale is not high in the local area, and the forest height estimation technique at the forest farm scale has its value.

5. Conclusions

In this study, we addressed the two main research questions posed at the outset. Firstly, our research focuses on enhancing forest height parameters in GEDI's L1B and L2A products through the utilization of airborne LiDAR data. By optimizing L2A footprints using environmental and acquisition parameters, and calibrating GEDI L1B waveform data with pseudo-waveforms simulated from airborne LiDAR point cloud and a 1D CNN model, we achieved notable improvements in accuracy. Specifically, the ground elevation and forest canopy height extracted from L2A data showed strong consistency with airborne LiDAR data (ground elevation: R^2 = 0.99, RMSE = 4.99 m; canopy height: R^2 = 0.42, RMSE = 5.16 m). Through optimization, we reduced the ground elevation extraction error by 45.5% (RMSE) and the canopy height extraction error by 30.3% (RMSE). Moreover, after calibration, the accuracy of forest height extraction from L1B waveform data significantly improved (R^2 = 0.84, RMSE = 3.13 m). Secondly, concerning the extrapolation mapping of forest height information based on GEDI data and satellite optical images, our results confirm the effectiveness of our approach. By combining the more accurate L1B forest height data with Sentinel-2 multispectral images, and by utilizing machine learning models for upscaling estimation, we successfully obtained a detailed distribution of forest height

on a large scale. The RF model (R^2 = 0.64, RMSE = 4.59 m) exhibited superior prediction accuracy in comparison to the ANN model (R^2 = 0.49, RMSE = 5.07 m). In conclusion, our study effectively addresses the research questions posed, providing insights into enhancing the accuracy of forest height parameters in GEDI data and realizing the extrapolation mapping of forest height information on a large scale.

Author Contributions: Conceptualization, L.C.; methodology, J.W. and X.S.; validation, J.W. and X.S.; formal analysis, J.W.; resources, L.C.; data curation, J.W. and X.S.; writing—original draft preparation, J.W.; writing—review and editing, J.W., X.S., and L.C.; visualization, J.W.; supervision, L.C.; funding acquisition, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program (2017YFD0600904), the Natural Science Foundation of Jiangsu Province (BK20220415), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Data Availability Statement: Data is contained within the article.

Acknowledgments: The authors extend their sincere appreciation to the foresters at Gaofeng Forest for their invaluable support in data collection and for generously sharing their insights into the local forests. Additionally, we would like to express our gratitude to the graduate students from the Department of Forest Management at Nanjing Forestry University for their constructive suggestions, which have enhanced the quality of this study.

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

References

1. Kramer, P.J. Carbon Dioxide Concentration, Photosynthesis, and Dry Matter Production. *BioScience* **1981**, *31*, 29–33. [CrossRef]
2. Pan, Y.; Birdsey, R.A.; Fang, J.; Houghton, R.; Kauppi, P.E.; Kurz, W.A.; Phillips, O.L.; Shvidenko, A.; Lewis, S.L.; Canadell, J.G.; et al. A Large and Persistent Carbon Sink in the World's Forests. *Science* **2011**, *333*, 988–993. [CrossRef] [PubMed]
3. Dixon, R.K.; Solomon, A.M.; Brown, S.; Houghton, R.A.; Trexier, M.C.; Wisniewski, J. Carbon Pools and Flux of Global Forest Ecosystems. *Science* **1994**, *263*, 185–190. [CrossRef] [PubMed]
4. Fang, J.; Guo, Z.; Hu, H.; Kato, T.; Muraoka, H.; Son, Y. Forest Biomass Carbon Sinks in East Asia, with Special Reference to the Relative Contributions of Forest Expansion and Forest Growth. *Glob. Chang. Biol.* **2014**, *20*, 2019–2030. [CrossRef] [PubMed]
5. Payn, T.; Carnus, J.-M.; Freer-Smith, P.; Kimberley, M.; Kollert, W.; Liu, S.; Orazio, C.; Rodriguez, L.; Silva, L.N.; Wingfield, M.J. Changes in Planted Forests and Future Global Implications. *For. Ecol. Manag.* **2015**, *352*, 57–67. [CrossRef]
6. Szulecka, J.; Monges Zalazar, E. Forest Plantations in Paraguay: Historical Developments and a Critical Diagnosis in a SWOT-AHP Framework. *Land Use Policy* **2017**, *60*, 384–394. [CrossRef]
7. Hunter, M.O.; Keller, M.; Victoria, D.; Morton, D.C. Tree Height and Tropical Forest Biomass Estimation. *Biogeosciences* **2013**, *10*, 8385–8399. [CrossRef]
8. Lefsky, M.A.; Cohen, W.B.; Harding, D.J.; Parker, G.G.; Acker, S.A.; Gower, S.T. Lidar Remote Sensing of Above-ground Biomass in Three Biomes. *Glob. Ecol. Biogeogr.* **2002**, *11*, 393–399. [CrossRef]
9. Simard, M.; Pinto, N.; Fisher, J.B.; Baccini, A. Mapping Forest Canopy Height Globally with Spaceborne Lidar. *J. Geophys. Res.* **2011**, *116*, G04021. [CrossRef]
10. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
11. Nelson, R.; Krabill, W.; MacLean, G. Determining Forest Canopy Characteristics Using Airborne Laser Data. *Remote Sens. Environ.* **1984**, *15*, 201–212. [CrossRef]
12. Duncanson, L.; Armston, J.; Disney, M.; Avitabile, V.; Barbier, N.; Calders, K.; Carter, S.; Chave, J.; Herold, M.; Crowther, T.W.; et al. The Importance of Consistent Global Forest Aboveground Biomass Product Validation. *Surv. Geophys.* **2019**, *40*, 979–999. [CrossRef] [PubMed]
13. Cui, S.; Fan, Y.; Jin, S.; Li, M. Extraction of Individual Tree Height Using Quickbird Images Based on Tree Shadow. *J. Northeast For. Univ.* **2011**, *39*, 47–50.
14. Pang, Y.; Li, Z.; Chen, E.; Sun, G. Lidar Remote Sensing Technology and Its Application in Forestry. *Sci. Silvae Sin.* **2005**, *41*, 129–136.
15. Zolkos, S.G.; Goetz, S.J.; Dubayah, R. A Meta-Analysis of Terrestrial Aboveground Biomass Estimation Using Lidar Remote Sensing. *Remote Sens. Environ.* **2013**, *128*, 289–298. [CrossRef]
16. Zengyuan, L.; Qingwang, L.; Yong, P. Review on forest parameters inversion using LiDAR. *Natl. Remote Sens. Bull.* **2016**, *20*, 1138–1150. [CrossRef]

17. Popescu, S.C.; Zhao, K.; Neuenschwander, A.; Lin, C. Satellite Lidar vs. Small Footprint Airborne Lidar: Comparing the Accuracy of Aboveground Biomass Estimates and Forest Structure Metrics at Footprint Level. *Remote Sens. Environ.* **2011**, *115*, 2786–2797. [CrossRef]
18. Zhang, W.; Chen, E.; Li, Z.; Zhao, L.; Ji, Y. Development of Forest Height Estimation Using InSAR/PolInSAR Technology. *Remote Sens. Technol. Appl.* **2017**, *32*, 983–997.
19. Wallace, L.; Lucieer, A.; Watson, C.; Turner, D. Development of a UAV-LiDAR System with Application to Forest Inventory. *Remote Sens.* **2012**, *4*, 1519–1543. [CrossRef]
20. Næsset, E.; Økland, T. Estimating Tree Height and Tree Crown Properties Using Airborne Scanning Laser in a Boreal Nature Reserve. *Remote Sens. Environ.* **2002**, *79*, 105–115. [CrossRef]
21. Adam, M.; Urbazaev, M.; Dubois, C.; Schmullius, C. Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters. *Remote Sens.* **2020**, *12*, 3948. [CrossRef]
22. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002. [CrossRef]
23. Quiros, E.; Polo, M.-E.; Fragoso-Campon, L. GEDI Elevation Accuracy Assessment: A Case Study of Southwest Spain. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5285–5299. [CrossRef]
24. Fayad, I.; Baghdadi, N.; Bailly, J.S.; Frappart, F.; Zribi, M. Analysis of GEDI Elevation Data Accuracy for Inland Waterbodies Altimetry. *Remote Sens.* **2020**, *12*, 2714. [CrossRef]
25. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping Global Forest Canopy Height through Integration of GEDI and Landsat Data. *Remote Sens. Environ.* **2021**, *253*, 112165. [CrossRef]
26. Liu, L.; Wang, C.; Nie, S.; Zhu, X.; Xi, X.; Wang, J. Analysis of the influence of different algorithms of GEDI L2A on the accuracy of ground elevation and forest canopy height. *J. Univ. Chin. Acad. Sci.* **2022**, *39*, 502–511.
27. Lin, X.; Cao, C. Remote Sensing Diagnosis of Forest Canopy Height and Forest Aboveground Biomass Based on ICESat-2 and GEDI. Ph.D. Thesis, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, 2021. [CrossRef]
28. Zhu, X.; Wang, C. Forest Height Retrieval of China with a Resolution of 30m Using ICESat-2 and GEDI Data. Ph.D. Thesis, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, 2022. [CrossRef]
29. Lahssini, K.; Baghdadi, N.; Le Maire, G.; Fayad, I. Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests. *Remote Sens.* **2022**, *14*, 6264. [CrossRef]
30. Lang, N.; Kalischek, N.; Armston, J.; Schindler, K.; Dubayah, R.; Wegner, J.D. Global Canopy Height Regression and Uncertainty Estimation from GEDI LIDAR Waveforms with Deep Ensembles. *Remote Sens. Environ.* **2022**, *268*, 112760. [CrossRef]
31. Liu, H.; Cao, F.; She, G.; Cao, L. Extrapolation Assessment for Forest Structural Parameters in Planted Forests of Southern China by UAV-LiDAR Samples and Multispectral Satellite Imagery. *Remote Sens.* **2022**, *14*, 2677. [CrossRef]
32. Crippen, R.; Buckley, S.; Agram, P.; Belz, E.; Gurrola, E.; Hensley, S.; Kobrick, M.; Lavalle, M.; Martin, J.; Neumann, M.; et al. Nasadem global elevation model: Methods and progress. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B4*, 125–128. [CrossRef]
33. Li, X.; Wessels, K.; Armston, J.; Hancock, S.; Mathieu, R.; Main, R.; Naidoo, L.; Erasmus, B.; Scholes, R. First Validation of GEDI Canopy Heights in African Savannas. *Remote Sens. Environ.* **2023**, *285*, 113402. [CrossRef]
34. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [CrossRef]
35. Poggio, L.; de Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.M.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty. *Soil* **2021**, *7*, 217–240. [CrossRef]
36. Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. Stable Classification with Limited Sample: Transferring a 30-m Resolution Sample Set Collected in 2015 to Mapping 10-m Resolution Global Land Cover in 2017. *Sci. Bull.* **2019**, *64*, 370–373. [CrossRef] [PubMed]
37. Blair, J.B.; Hofton, M.A. Modeling Laser Altimeter Return Waveforms over Complex Vegetation Using High-Resolution Elevation Data. *Geophys. Res. Lett.* **1999**, *26*, 2509–2512. [CrossRef]
38. Lecun, Y.A.; Bottou, L.; Orr, G.B.; Müller, K.R. Efficient Backprop. In *Neural Networks: Tricks of the Trade*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 1998; pp. 9–48.
39. Montavon, G.; Orr, G.B.; Müller, K.-R. *Neural Networks: Tricks of the Trade*, 2nd ed.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012. [CrossRef]
40. Zhao, X.; Guo, Q.; Su, Y.; Xue, B. Improved Progressive TIN Densification Filtering Algorithm for Airborne LiDAR Data in Forested Areas. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 79–91. [CrossRef]
41. Guo, Q.; Li, W.; Yu, H.; Alvarez, O. Effects of Topographic Variability and Lidar Sampling Density on Several DEM Interpolation Methods. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 701–712. [CrossRef]
42. Silva, C.A.; Duncanson, L.; Hancock, S.; Neuenschwander, A.; Thomas, N.; Hofton, M.; Fatoyinbo, L.; Simard, M.; Marshak, C.Z.; Armston, J.; et al. Fusing Simulated GEDI, ICESat-2 and NISAR Data for Regional Aboveground Biomass Mapping. *Remote Sens. Environ.* **2021**, *253*, 112234. [CrossRef]

43. Silva, C.A.; Saatchi, S.; Garcia, M.; Labrière, N.; Klauberg, C.; Ferraz, A.; Meyer, V.; Jeffery, K.J.; Abernethy, K.; White, L.; et al. Comparison of Small- and Large-Footprint Lidar Characterization of Tropical Forest Aboveground Structure and Biomass: A Case Study From Central Gabon. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3512–3526. [CrossRef]
44. Hancock, S.; Armston, J.; Hofton, M.; Sun, X.; Tang, H.; Duncanson, L.I.; Kellner, J.R.; Dubayah, R. The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. *Earth Space Sci.* **2019**, *6*, 294–310. [CrossRef]
45. Shen, W.; Li, M.; Huang, C.; Wei, A. Quantifying Live Aboveground Biomass and Forest Disturbance of Mountainous Natural and Plantation Forests in Northern Guangdong, China, Based on Multi-Temporal Landsat, PALSAR and Field Plot Data. *Remote Sens.* **2016**, *8*, 595. [CrossRef]
46. Gitelson, A.A.; Merzlyak, M.N.; Chivkunova, O.B. Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochem. Photobiol.* **2001**, *74*, 38–45. [CrossRef] [PubMed]
47. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [CrossRef] [PubMed]
48. Vincini, M.; Frazzi, E. Comparing Narrow and Broad-Band Vegetation Indices to Estimate Leaf Chlorophyll Content in Planophile Crop Canopies. *Precis. Agric.* **2011**, *12*, 334–344. [CrossRef]
49. Huete, A. A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. *Remote Sens. Environ.* **1997**, *59*, 440–451. [CrossRef]
50. Jiang, Z.; Huete, A.; Didan, K.; Miura, T. Development of a Two-Band Enhanced Vegetation Index without a Blue Band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845. [CrossRef]
51. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]
52. Huntjr, E.; Rock, B. Detection of Changes in Leaf Water Content Using Near- and Middle-Infrared Reflectances ☆. *Remote Sens. Environ.* **1989**, *30*, 43–54. [CrossRef]
53. Gitelson, A.; Merzlyak, M.N. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus Hippocastanum L. and Acer Platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *J. Plant Physiol.* **1994**, *143*, 286–292. [CrossRef]
54. Birth, G.S.; McVey, G.R. Measuring the Color of Growing Turf with a Reflectance Spectrophotometer [1]. *Agron. J.* **1968**, *60*, 640–643. [CrossRef]
55. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]
56. LeCun, Y.; Bengio, Y. Convolutional Networks for Images, Speech, and Time Series. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 1998; pp. 255–258. ISBN 978-0-262-51102-5.
57. Fayad, I.; Baghdadi, N.N.; Alvares, C.A.; Stape, J.L.; Bailly, J.S.; Scolforo, H.F.; Zribi, M.; Maire, G.L. Assessment of GEDI's LiDAR Data for the Estimation of Canopy Heights and Wood Volume of Eucalyptus Plantations in Brazil. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7095–7110. [CrossRef]
58. Radosavovic, I.; Kosaraju, R.P.; Girshick, R.; He, K.; Dollar, P. Designing Network Design Spaces. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10428–10436.
59. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
60. Singh, K.; Ojha, P.; Malik, A.; Jain, G. Partial Least Squares and Artificial Neural Networks Modeling for Predicting Chlorophenol Removal from Aqueous Solution. *Chemom. Intell. Lab. Syst.* **2009**, *99*, 150–160. [CrossRef]
61. Wang, C.; Elmore, A.J.; Numata, I.; Cochrane, M.A.; Shaogang, L.; Huang, J.; Zhao, Y.; Li, Y. Factors Affecting Relative Height and Ground Elevation Estimations of GEDI among Forest Types across the Conterminous USA. *GIScience Remote Sens.* **2022**, *59*, 975–999. [CrossRef]
62. Stereńczak, K.; Ciesielski, M.; Balazy, R.; Zawiła-Niedźwiecki, T. Comparison of Various Algorithms for DTM Interpolation from LIDAR Data in Dense Mountain Forests. *Eur. J. Remote Sens.* **2016**, *49*, 599–621. [CrossRef]
63. Qi, W.; Dubayah, R.O. Combining Tandem-X InSAR and Simulated GEDI Lidar Observations for Forest Structure Mapping. *Remote Sens. Environ.* **2016**, *187*, 253–266. [CrossRef]
64. Huettermann, S.; Jones, S.; Soto-Berelov, M.; Hislop, S. Intercomparison of Real and Simulated GEDI Observations across Sclerophyll Forests. *Remote Sens.* **2022**, *14*, 2096. [CrossRef]
65. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 23.
66. Liu, X.; Su, Y.; Hu, T.; Yang, Q.; Liu, B.; Deng, Y.; Tang, H.; Tang, Z.; Fang, J.; Guo, Q. Neural Network Guided Interpolation for Mapping Canopy Height of China's Forests by Integrating GEDI and ICESat-2 Data. *Remote Sens. Environ.* **2022**, *269*, 112844. [CrossRef]

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