Article

Inversion of Forest Aboveground Biomass in Regions with Complex Terrain Based on PolSAR Data and a Machine Learning Model: Radiometric Terrain Correction Assessment

Yonghui Nie [1], Rula Sa [1], Sergey Chumachenko [2], Yifan Hu [1], Youzhu Wang [1] and Wenyi Fan [1,*]

[1] Key Laboratory of Sustainable Forest Ecosystem Management (Ministry of Education), School of Forestry, Northeast Forestry University, Harbin 150040, China; nieyonghui@nefu.edu.cn (Y.N.); sarula@nefu.edu.cn (R.S.); yifan_hu2021@nefu.edu.cn (Y.H.); wangyouzhu1995@nefu.edu.cn (Y.W.)

[2] Head of Department of Forest Management, GIS of Bauman Moscow State Technical University, 2-Ya Baumanskaya Ulitsa, 5c1, Moscow 105005, Russia; chumachenko.s.i@gmail.com

[*] Correspondence: fanwy@nefu.edu.cn; Tel.: +86-139-4605-5384

Abstract: The accurate estimation of forest aboveground biomass (AGB) in areas with complex terrain is very important for quantifying the carbon sequestration capacity of forest ecosystems and studying the regional or global carbon cycle. In our previous research, we proposed the radiometric terrain correction (RTC) process for introducing normalized correction factors, which has strong effectiveness and robustness in terms of the backscattering coefficient of polarimetric synthetic aperture radar (PolSAR) data and the monadic model. However, the impact of RTC on the correctness of feature extraction and the performance of regression models requires further exploration in the retrieval of forest AGB based on a machine learning multiple regression model. In this study, based on PolSAR data provided by ALOS-2, 117 feature variables were accurately extracted using the RTC process, and then Boruta and recursive feature elimination with cross-validation (RFECV) algorithms were used to perform multi-step feature selection. Finally, 10 machine learning regression models and the Optuna algorithm were used to evaluate the effectiveness and robustness of RTC in improving the quality of the PolSAR feature set and the performance of the regression models. The results revealed that, compared with the situation without RTC treatment, RTC can effectively and robustly improve the accuracy of PolSAR features (the Pearson correlation R between the PolSAR features and measured forest AGB increased by 0.26 on average) and the performance of regression models (the coefficient of determination R^2 increased by 0.14 on average, and the rRMSE decreased by 4.20% on average), but there is a certain degree of overcorrection in the RTC process. In addition, in situations where the data exhibit linear relationships, linear models remain a powerful and practical choice due to their efficient and stable characteristics. For example, the optimal regression model in this study is the Bayesian Ridge linear regression model (R^2 = 0.82, rRMSE = 18.06%).

Keywords: forest AGB; PolSAR features; ALOS-2; radiometric terrain correction; machine learning; linear regression model

1. Introduction

Forest aboveground biomass (AGB) is a core parameter for monitoring forest ecosystems and assessing carbon sink capacity. Therefore, it plays an important role in quantifying the carbon sequestration capacity of forest ecosystems and studying regional and global carbon cycles [1,2]. However, it is challenging to accurately estimate forest AGB, especially in regions with complex and variable topography, which increases the uncertainty of assessments of forest ecosystem productivity. Accuracy is important for improving forest resource status surveys [3,4]. Remote sensing technology, including optical remote sensing, microwave remote sensing, and light detection and ranging (LiDAR), has become the most practical way to estimate forest AGB because it is long-term, efficient, and economical

for acquiring large-scale spatial distribution information about forests [5–7]. Among the potential techniques, synthetic aperture radar (SAR) has been widely used for the dynamic monitoring of forest AGB in large-scale areas due to its good performance in all weathers, sustainable observation, and cloud penetration [8–10]. Although multi-source remote sensing data can significantly improve the inversion accuracy of forest AGB [11–15], it is still crucial to explore AGB estimation methods relying only on SAR data. Especially in areas where optical remote sensing is limited due to frequent cloud and fog, SAR becomes an irreplaceable and effective tool [16].

At present, there are four common technologies for forest AGB based on SAR data: interferometric SAR (InSAR) [17], polarimetric SAR (PolSAR) [18,19], polarimetric interferometric SAR (PolInSAR) [20–22], and tomographic SAR (TomoSAR) [23,24]. PolSAR technology is widely used in large-scale inversion of forest AGB because it does not rely on interference technology and does not need to meet the strict requirements of repeated orbit observation and accurate time synchronization. It has more relaxed data acquisition conditions and simplicity of data collection. In addition, the modeling methods for estimating forest AGB based on PolSAR data can be divided into scattering mechanism methods [25], machine learning methods [26–28], and deep learning methods [29–31]. Scattering mechanism methods (such as the water cloud model (WCM)) [32,33] can be useful because with a simplified physical model it is difficult to describe the real scattering characteristics of a forest with a complex structure. Therefore, the universal applicability of this kind of model may be poor in a forest area with strong heterogeneity. For deep learning methods, a large amount of labeled data are required for the model training. However, the actual measurement of forest biomass is often costly and difficult, and so such models are relatively poor in interpretation and prone to overfitting. Therefore, this study is based on PolSAR data, using a flexible and more adaptable machine learning method to estimate forest AGB.

The process of inversion of forest AGB by the machine learning method based on PolSAR data includes three key stages: firstly, accurate extraction and derivation of PolSAR feature parameters; secondly, selection of the optimal feature subset; and, finally, selection of the most suitable regression model and algorithm. Accurate extraction of PolSAR features allows us to mine the information contained in the original data and generate new features based on physical mechanisms or statistical methods. This enhances the ability of the statistical model to reflect the complex relationship between forest AGB and PolSAR features [34]. Backscattering coefficient and polarization decomposition parameters are the most commonly used PolSAR features in estimating forest AGB, with their extraction theories having been extensively studied in literature [35–37]. However, due to the side-view imaging principle of the SAR system and the characteristics of tilt distance being inferred based on echo delay, the actual topographic relief and other factors can cause serious interference to the scattered echo information received by SAR. This can limit the accuracy and reliability of the features extracted from PolSAR data [38]. Therefore, fully considering the influence of terrain and implementing effective terrain correction is essential to improving the accuracy of PolSAR features. At present, the radiometric terrain correction (RTC) process designed for PolSAR data is a relatively complete solution for terrain correction [39], which includes polarization orientation angle correction (POAC), effective scattering area correction (ESAC), and angular variation effect correction (AVEC). We introduce normalized ESAC factors on the basis of RTC to further improve the RTC process [40]. Although the RTC method has shown excellent performance in the correction of backscattering coefficient and its effectiveness and robustness have been widely recognized [41,42], there is a lack of in-depth research on the correction effect and quantitative analysis of the polarization decomposition parameters of PolSAR data by RTC. In addition, the effect of the fusion of the RTC process with different machine learning algorithms and potential improvements need to be discussed further.

Feature selection is essential to achieve robust and high-precision estimation of forest AGB based on PolSAR data [43–45]. The data and features determine the upper limit

of machine learning, while models and algorithms only approach this upper limit [46]. The purpose of selecting the optimal feature subset is to remove invalid, redundant, and interference-producing variables from many potential feature variables, and solve the multicollinearity problem between the features synchronously. The aim is to select the feature set with the greatest information value and prediction efficiency so as to improve the efficiency and accuracy of the model construction [47–50]. At present, the methods of feature selection are mainly divided into three categories: filter, wrapper, and embedded algorithms [51,52]. In addition, a hybrid feature selection algorithm combining the advantages of multiple feature selection methods through multi-step screening has also been proposed [53–55].

The selection of appropriate statistical models can help to build models that accurately reflect the complex relationship between AGB and remote sensing data, thus improving the reliability and accuracy of carbon storage assessment. At present, according to the different construction methods of statistical models and the characteristics of parameter assumptions, regression models based on machine learning can be divided into two categories: parametric models and non-parametric models [5,56]. Non-parametric models can flexibly adapt to various complex data structures without pre-setting the specific form of the model, and thus show excellent adaptability and robustness. However, the parametric model is still a competitive regression analysis tool because of its clear explanations and high computational efficiency [57]. In addition, in the inversion of forest AGB based on PolSAR characteristics, a widely accepted view is that forest AGB has the best correlation with the backscattering coefficient of HV polarization channels, but there is no direct and simple linear relationship between the two; in fact, a logarithmic function may be the more appropriate relationship [58]. Moreover, our previous research focused on the potential of RTC processes to improve the accuracy of forest AGB inversion based on univariate models [40]. Therefore, this study will not focus on the construction of monadic linear or nonlinear models, but instead will seek multivariate and high-level nonlinear model structures that can capture this complex relationship.

This research aims to achieve three key tasks: (1) quantitatively evaluate the ability of the RTC process to enhance the correlation between forest AGB and PolSAR features (backscattering coefficient, polarization decomposition parameters, and related derived features); (2) explore the effect of RTC on polarization decomposition parameters and the underlying mechanism of action; and (3) evaluate the potential of the RTC process to improve the performance of forest AGB inversion based on machine learning models, and identify the optimal model by comparing the performance of different models.

2. Materials

2.1. Study Area

The study area is located in Saihanba Forest Farm (116°53′~117°43′E, 41°55′~42°35′N, Figure 1) in the border area between the northern mountains of Weichang County, Hebei Province, China, with an altitude of 1010 to 1940 m. The study area covers a wide range of vegetation types, including deciduous coniferous forest, evergreen coniferous forest, broad-leaved forest, scrubland, and grassland. Moreover, the topography of the study area is complex, with the forest cover being a combination of flat woodlands and steep hillsides with slopes up to 55.47°. This variety of terrain conditions provides an ideal natural laboratory for conducting comparative studies of complex terrain. In addition, the study area has a large-scale artificial forest, covering an area of about 93,461 hectares, and its large area of artificial forest is convenient for scientific researchers to collect real and rich test data on the spot.
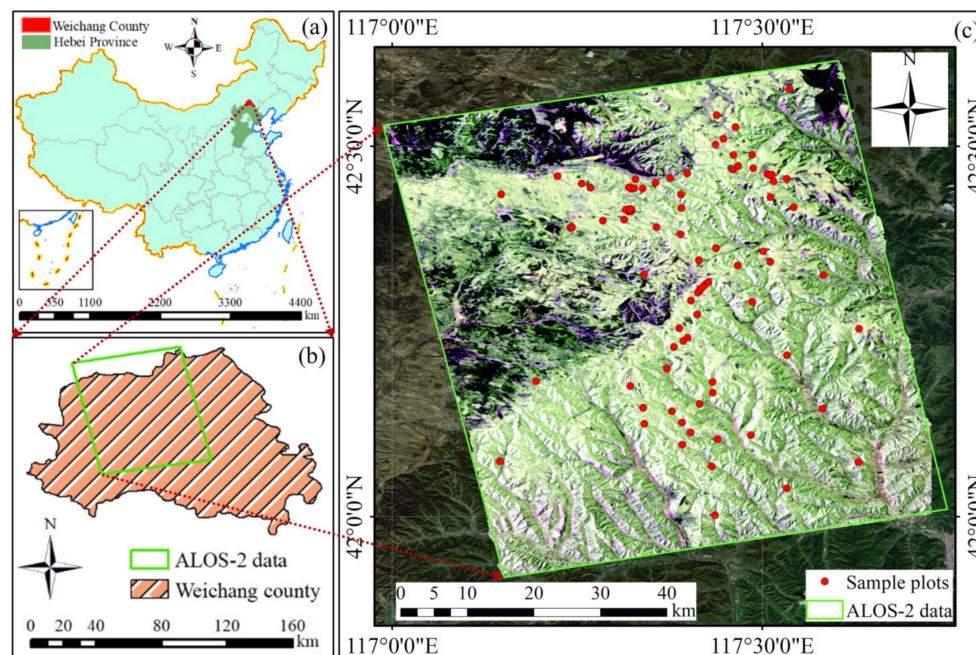
Figure 1. Overview of study sites: (**a**) the location of Saihanba Forest Farm in relation to the provinces and counties in China; (**b**) the spatial location of ALOS-2 data relative to Weichang County; (**c**) the Pauli RGB image (R: |HH-VV|, G: |HV|, B: |HH + VV|) based on PolSAR data and the location of the measured samples; the basemap is the optical image of Tianditu.

2.2. PolSAR Data and Pre-Processing

The PolSAR data used in this study are L-band full-polarization single-look complex (SLC) data (Table 1) obtained by the Advanced Land Observing Satellite-2 (ALOS-2) (developed by the Japan Aerospace Exploration Agency (JAXA)) on 11 July, 25 July, and 8 August 2020. As the only platform currently equipped with a satellite-borne L-band SAR sensor, ALOS-2's four-polarization observation mode can significantly enhance its ability to capture complex forest structures and scattering mechanisms. Moreover, compared with C-band, L-band has stronger penetration and can more accurately reveal the details of the forest internal structure, which is conducive to improving the accuracy of forest AGB.

Table 1. The main parameters of the PolSAR data.

Parameter	Value	Parameter	Value
Data formats and processing level	CEOS level 1.1	Observation mode	HBQ
Observation date of scene center	27.8054°	Radar wavelength	0.2424525 m
Length of range direction	49.8 km	Range resolution	2.860844 m
Length of azimuth direction	69.3 km	Azimuth resolution	2.642742 m
Orbit direction	Ascending	Range pixel	8392
Observation direction	Right looking	Azimuth pixel	26,105

Note: HBQ is High-sensitive mode Full (Quad.) polarimetry.

In this study, the Sentinel Application Platform (SNAP_v9.0.0) software developed by the European Space Agency completed the pre-processing steps for the PolSAR data: cross-channel SNR correction, Faraday rotation correction, and radiometric calibration for complex data (Equation (2), derived from radiometric calibration of real data; Equation (1)) [35,40], multi-look averaging of 4 × 9 (range × azimuth), refined LEE filter, and range doppler terrain correction). The ensemble average size for the covariance matrix is crucial, especially when studying natural vegetation. For natural vegetation, when the reflection symmetry condition remains unchanged, the volume scattering power induced by vegetation will improve with the increase in the average ensemble size [59]. The purpose of choosing

4×9 (range $\times$ azimuth) as the core parameter of multi-look processing is to process the pixels of the PolSAR data into square pixels as close as possible to the reference DEM data and the measured plot.

$$\sigma_{slc}^0 = 10 \cdot \log_{10}\left\langle I^2 + Q^2 \right\rangle + CF_1 - A, \tag{1}$$

where σ_{slc}^0 is the backscattering coefficient after radiation calibration; I and Q represent the real and imaginary parts of SLC 1.1 products, respectively; CF_1 is -83.0 dB; and A is 32.0 dB.

$$I_{cal} = \frac{I}{10^{5.75}}; Q_{cal} = \frac{Q}{10^{5.75}}, \tag{2}$$

where I_{cal} and Q_{cal} are the real and imaginary parts of SLC 1.1 products after radiation calibration.

2.3. Ground-Measured Forest AGB

The measured forest AGB data were collected through a field survey conducted at Saihanba Forest Farm in August 2020 for the training and validation of the forest AGB inversion model. Through screening the key stand characteristics of the plots, including tree species composition, understory environment, topographic slope, and stand thinning density, 132 rhomboid plots (Figure 1) with an area of 0.06 ha (24.49 m $\times$ 24.49 m) were established and investigated, including 28 plots distributed on a 1.0 km grid and 104 plots with a relatively uniform distribution. All plots were located using differential GPS (Unistrong RTK-G10, Unistrong, Beijing, China), and plot information such as plot type, plant number, and slope, as well as individual information of trees with a diameter at breast height (DBH) $\geq$ 5.0 cm, was recorded, including DBH, tree height, and tree species. Finally, the biomass per plant was calculated separately according to the allometric equation of different tree species (Table 2); then the total forest AGB of the plot was obtained by summing (Equation (3)) [40] all the biomass per plant. Table 3 provides statistical details of the field data for this study.

$$AGB_j = \sum_{i=1}^{m} W_i / 1000 \times S, \tag{3}$$

where AGB_j is the *AGB* (t/ha) of the *j*th sample plot, *m* is the number of trees in the plot, W_i is the *AGB* of the *i*th tree (kg), and *S* is the area of the sample plot (*S* = 0.06 ha).

Table 2. Allometric growth equation of different tree species.

Tree Species	Allometric Equation	Ref.
Larix principis-rupprechtii	$W = 0.1431 \cdot DBH^{2.2193}$	[60]
Betula platyphylla	$W = 0.0330 \cdot DBH^{2.9314}$	[61]
Pinus sylvestris var. mongolica	$W = 0.0930 \cdot DBH^{2.3429}$	[62]
Pinus tabuliformis	$W = 0.0520 \cdot DBH^{2.5830}$	[63]
Acer truncatum	$W = 0.1260 \cdot DBH^{2.3830}$	[64]

Note: *W* is the AGB (kg) per plant, and *DBH* is the diameter at breast height (cm).

Table 3. Sample plot of AGB statistics.

Count	Min (t/ha)	Max (t/ha)	Mean (t/ha)	SD (t/ha)
132	36.39	166.58	103.88	32.31

Note: SD is standard deviation; Count is the number of sample plots.

2.4. SRTM DEM

Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) data with a resolution of 30 m were used to complement the PolSAR data to implement geocoding and calculate RTC-related factors.

3. Methods

The processing framework of this study (Figure 2) includes the following steps: (1) radiometric terrain correction (RTC) of PolSAR data, including polarization orientation angle correction (POAC), effective scattering area correction (ESAC), and angular variation effect correction (AVEC); (2) feature extraction and feature derivation based on PolSAR data, including backscattering coefficient, polarization decomposition parameters, and their related derived feature variables; and (3) feature selection and model training. All machine learning algorithms were executed in the Python 3.10 development environment. The programs that involved partitioning the dataset used the default parameter (test_size = 0.25), which was 75% for the training set and 25% for the test set. The following sections present the necessary theories and related formulas for each stage in the flowchart in more detail.
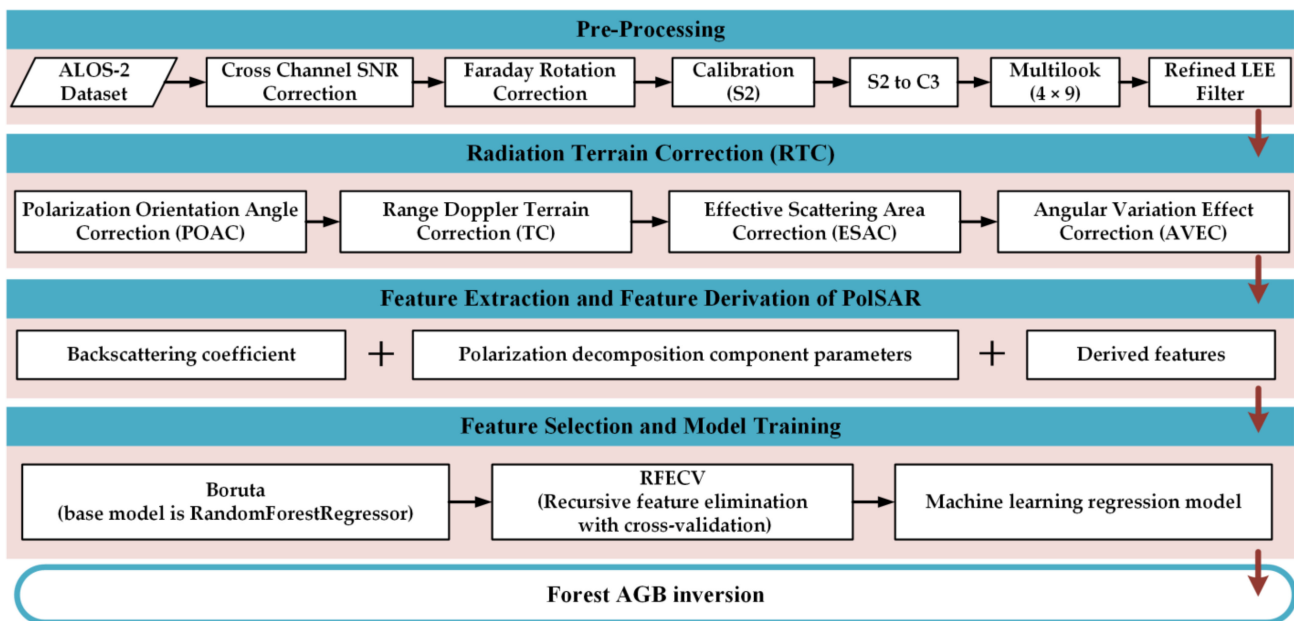

Figure 2. A flowchart of the proposed forest AGB mapping scheme.

3.1. Radiometric Terrain Correction for PolSAR

3.1.1. Polarization Orientation Angle Correction

When there is topographic relief in the target area, the polarization orientation angle (POA)—that is, the angle between the long axis of the polarization ellipse and the horizontal axis of the ground plane—will be biased due to the slope (mainly the azimuth slope). In order to compensate for this deviation of POA caused by terrain, the POA extracted by the circular polarization method [39,65] is used first, and then the three-dimensional polarization covariance (C3) matrix is compensated (Equation (4)).

$$
\begin{aligned}
C_{3\_POAC} &= U_{3(\eta)} C_3 U_{3(\eta)}^{-1} \\
U_{3(\eta)} &= \frac{1}{2} \begin{bmatrix} 1+\cos(2\eta) & \sqrt{2}\sin(2\eta) & 1-\cos(2\eta) \\ -\sqrt{2}\sin(2\eta) & 2\cos(2\eta) & \sqrt{2}\sin(2\eta) \\ 1-\cos(2\eta) & -\sqrt{2}\sin(2\eta) & 1+\cos(2\eta) \end{bmatrix}, \\
\eta &= \frac{1}{4}\left[-1 \times Arg\left(\left\langle \frac{-4\mathrm{Re}\left\langle (S_{HH}-S_{VV})S_{HV}^* \right\rangle}{-\left\langle |S_{HH}-S_{VV}|^2 \right\rangle + 4\left\langle |S_{HV}|^2 \right\rangle} \right\rangle \right) + \pi \right]
\end{aligned}
\tag{4}
$$

where C_3 and C_{3_POAC} are the C_3 matrix before and after polarization orientation angle correction (POAC), respectively; η is the polarization orientation angle (POA); $Arg(\cdot)$ is a phase function; and when $\eta > \pi/4$, $\eta = \eta - \pi/2$.

3.1.2. Effective Scattering Area Correction

The effective scattering area is the projected area of the actual ground element on the isophase surface. However, due to the difference in terrain slope, ground units of the same SAR data in the geographical coordinate system correspond to different effective scattering areas. Therefore, effective scattering area correction (ESAC) takes flat land as a reference to correct the area effect of SAR data under different slope conditions so as to eliminate the influence of terrain slope on scattering information [39,40]. The C_3 matrix can be corrected using Equation (5):

$$C_{3\_ESAC} = C_3 \cdot \frac{\sin(\theta_{loc})}{\sin(\theta_{ref})}, \tag{5}$$

where C_{3_ESAC} is the C_3 matrix after ESAC; C_3 refers to the C_3 matrix without ESAC; θ_{loc} is the local incidence angle; and θ_{ref} is the radar incidence angle.

3.1.3. Angular Variation Effect Correction

The angular variation effect (AVE) refers to the phenomenon whereby the local incidence angle changes due to a change in terrain, which leads to a change in the scattering mechanism of the target object (such as forest). The angular variation effect correction (AVEC) is typically based on the cosine model (Equation (6)) to correct forest-covered areas. In this study, the forest cover area was first extracted by Freeman3 decomposition and Wishart unsupervised classification [66], and then the optimal n value matrix of the C_3 matrix (different elements correspond to different optimal n values) was obtained by calculating the optimal n value corresponding to the main diagonal elements of the C_3 matrix. For each element of the three principal diagonal elements, the optimal n value was first determined by iterating the minimum value of the absolute value of $f(n)$ when n takes different values [39].

$$C_{3\_AVEC} = C_3 \cdot \left(\frac{\cos\theta_{ref}}{\cos\theta_{loc}}\right)^{n_{pq}} = C_3 \odot \begin{bmatrix} k(n_{HH}) & k\left(\frac{n_{HH}+n_{HV}}{2}\right) & k\left(\frac{n_{HH}+n_{VV}}{2}\right) \\ k\left(\frac{n_{HH}+n_{HV}}{2}\right) & k(n_{HV}) & k\left(\frac{n_{HV}+n_{VV}}{2}\right) \\ k\left(\frac{n_{HH}+n_{VV}}{2}\right) & k\left(\frac{n_{HV}+n_{VV}}{2}\right) & k(n_{VV}) \end{bmatrix}, \tag{6}$$

$$n_{pq} = argmin\left\{f_{(n)}\right\} = argmin\left\{\left|\rho\left(\theta_{loc}, C_{3\_AVEC\_pq}\right)\right|\right\}$$

where $(p, q) = (H, V)$ is the polarization mode of incidence and scattering of electromagnetic waves; n_{pq} is the optimal n value of different polarization channels; $\odot$ is the Hadamard product; $argmin\{\cdot\}$ is the operation that takes the minimum value; $f(n)$ is the absolute value of the correlation coefficient between the local incidence angle θ_{loc} and $C_{3\_AVEC\_pq}$; $C_{3\_AVEC\_pq}$ is the decibel state of the intensity value after AVEC is applied to one of the principal diagonal elements of the C_3 matrix at a certain value of n; and $\rho(\cdot)$ is the correlation coefficient function.

3.2. Feature Extraction and Feature Derivation of PolSAR

All 118 PolSAR features extracted in this study were based on the Polarimetric SAR Data Processing and Educational Toolbox (PolSARpro_v5.1.1) and the Sentinel Application Platform (SNAP_v9.0.0).

3.2.1. Backscattering Coefficient Features of PolSAR

This study extracted the intensity values (σ_{HH}^0, σ_{HV}^0, and σ_{VV}^0) and decibel state values (σ_{HH}^0_db, σ_{HV}^0_db, and σ_{VV}^0_db) of the three main diagonal elements of the C_3 matrix as the six original backscatter coefficient features, and then calculated 20 classical derived features (Table 4).

Table 4. The features derived from the backscattering coefficients extracted from the PolSAR data.

Features	Symbol	Equation	Symbol (dB)	Source
Span	Span	$\sigma^0_{HH} + 2\sigma^0_{HV} + \sigma^0_{VV}$	Span_db	[67]
Co-Pol HH/VV Ratio	R_HH/VV	$\sigma^0_{HH}/\sigma^0_{VV}$	R_HH/VV_db	[67]
Cross-Pol HH/HV Ratio	R_HH/HV	$\sigma^0_{HH}/\sigma^0_{HV}$	R_HH/HV_db	[67]
Cross-Pol VV/VH Ratio	R_VV/VH	$\sigma^0_{VV}/\sigma^0_{VH}$	R_VV/VH_db	[68]
Biomass Index	BMI	$(\sigma^0_{HH} + \sigma^0_{VV})/2$	BMI_db	[69]
Volume Scattering Index	VSI	$\sigma^0_{HV}/(\sigma^0_{HV} + BMI)$	VSI_db	[69]
Canopy Structure Index	CSI	$\sigma^0_{VV}/(\sigma^0_{HH} + \sigma^0_{VV})$	CSI_db	[69]
Radar Vegetation Index	RVI	$8\sigma^0_{HV}/(\sigma^0_{HH} + 2\sigma^0_{HV} + \sigma^0_{VV})$	RVI_db	[70]
Radar Forest Degradation Index	RFDI	$(\sigma^0_{HH} - \sigma^0_{HV})/(\sigma^0_{HH} + \sigma^0_{HV})$	RFDI_db	[71]
modified RFDI	mRFDI	$(\sigma^0_{VV} - \sigma^0_{HV})/(\sigma^0_{VV} + \sigma^0_{HV})$	mRFDI_db	[72]

Note: The backscattering coefficient σ^0_{pq} in the table is the backscattering intensity in linear units.

3.2.2. Polarization Decomposition Features of PolSAR

In this study, 72 original polarization decomposition-related features in linear units and decibel units were extracted based on 12 polarization decomposition methods (Table 5), and then we calculated 20 derived features (Table 6). The Volume scattering component (Vol) of the Pauli three-component (PAU3) in this study is exactly the same as σ^0_{HV}, so only the Surface scattering component (Odd) and the Double-bounce scattering component (Dbl) of the PAU3 decomposition method were used.

Table 5. The polarization decomposition features extracted from the PolSAR data.

Decomposition Methods	Abbreviation	Symbol	Source
Freeman two-component	FRE2	*F2V, F2O*	[73]
Freeman three-component	FRE3	*F3V, F3D, F3O*	[74]
Yamaguchi three-component	YAM3	*Y3V, Y3D, Y3O*	[75]
Yamaguchi four-component	YAM4	*Y4V, Y4D, Y4O, Y4H*	[76]
Van Zyl three-component	VAZ3	*V3V, V3D, V3O*	[77]
An and Yang three-component	ANY3	*A3V, A3D, A3O*	[78]
Model-free three-component	MF3CF	*M3V, M3D, M3O*	[79]
Model-free four-component	MF4CF	*M4V, M4D, M4O, M4H*	[80]
Cloude three-component	CLD3	*C3V, C3D, C3O*	[81]
Krogager three-component	KRO3	*K3S, K3D, K3H*	[82]
H-A-alpha	HAα	*H, A, α*	[83]
Pauli three-component	PAU3	*P3D, P3O*	[84]

Note: In this table, the symbol of each component of the different polarization decomposition methods (except H-A-alpha) has a consistent marking method: the first character (letter) is the first letter of the abbreviation of the polarization decomposition method, the second character (number) is the number of its components, and the third character (letter) is the first letter of the abbreviation of the different component types, including Volume scattering component (Vol), Surface scattering component (Odd), Double-bounce scattering component (Dbl), Helix scattering component (Hel), and Sphere scattering component (Sph). In addition, the decibel symbol of each polarization decomposition component adds a suffix ("_db") to the symbol of its intensity unit, which is not listed in this table. For example, the decibel symbol of *F3V* is F3V_db.

Table 6. The features derived from the polarization decomposition extracted from the PolSAR data.

Symbol	Equation	Symbol (dB)	Source
F2R_1	*F2V/(F2O)*	F2R_1_db	[85]
F3R_1	*F3V/(F3D + F3O)*	F3R_1_db	[85]
A3R_1	*A3V/(A3D + A3O)*	A3R_1_db	[85]
V3R_1	*V3V/(V3D + V3O)*	V3R_1_db	[85]
Y3R_1	*Y3V/(Y3D + Y3O)*	Y3R_1_db	[85]
F2R_2	*F2V/(F2V + F2O)*	F2R_2_db	[86]
F3R_2	*F3V/(F3V + F3D + F3O)*	F3R_2_db	[86]
A3R_2	*A3V/(A3V + A3D + A3O)*	A3R_2_db	[86]
V3R_2	*V3V/(V3V + V3D + V3O)*	V3R_2_db	[86]
Y3R_2	*Y3V/(Y3V + Y3D + Y3O)*	Y3R_2_db	[86]

Note: The meanings of the symbols in the equations in this table are shown in Table 5.

3.3. Forest AGB Regression Modeling Algorithms and Model Evaluation

In this study, 10 multiple regression models (Table 7) were selected to evaluate the effectiveness and robustness of RTC on the performance of different machine learning regression models. Multiple regression models are statistical and machine learning methods used to analyze the relationship between a continuous dependent variable (response variable) and two or more independent variables (feature variables), which can be divided into multiple linear regression models and non-parametric regression models. Non-parametric regression models do not depend on the form of the model, while the multiple linear regression model has an explicit parametric form (Equation (7)) [67]:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n + \varepsilon, \tag{7}$$

where Y is the target variable of the prediction (decibels of forest AGB); X_1, X_2, ..., X_n are the predictor variables; α_0 is a constant; α_1, α_2, ..., α_{2n} are the regression coefficients associated with the corresponding variables; n is the number of the predictor variables; and ε is the error term.

Table 7. The 10 regression models used in this study.

Model	Python Package	Module	Estimator	Category
Ridge	scikit-learn (v1.4.2)	sklearn.linear_model	Ridge	linear
Lasso	scikit-learn (v1.4.2)	sklearn.linear_model	Lasso	linear
ElasticNet	scikit-learn (v1.4.2)	sklearn.linear_model	ElasticNet	linear
BysRidge	scikit-learn (v1.4.2)	sklearn.linear_model	BayesianRidge	linear
ARD	scikit-learn (v1.4.2)	sklearn.linear_model	ARDRegression	linear
RF	scikit-learn (v1.4.2)	sklearn.ensemble	RandomForestRegressor	non-parametric
ET	scikit-learn (v1.4.2)	sklearn.ensemble	ExtraTreesRegressor	non-parametric
AdaBoost	scikit-learn (v1.4.2)	sklearn.ensemble	AdaBoostRegressor	non-parametric
XGBoost	xgboost (v2.0.3)	xgboost	XGBRegressor	non-parametric
CatBoost	catboost (v1.2.3)	catboost	CatBoostRegressor	non-parametric

The machine learning process in this study included the following steps (implementation details shown in Table 8): (1) The random forest regression model (RF) based on the default parameters served as the base model for Boruta's algorithm [87] and performed the initial feature filtering on the entire dataset instead of the training set. The goal is to identify and remove features that are irrelevant or weakly related to the target variable. (2) First, the Optuna algorithm (n_trials = 500, including 10-fold cross-validation) was used to obtain the optimal hyperparameters of 10 regression models on the feature set after the initial feature selection. Then, the 10 regression models based on the optimal hyperparameters corresponding to different models were used as the base model for recursive feature elimination with cross-validation (RFECV) feature selection, and the feature selection of the second step was completed. The aim is to iteratively search the optimal feature subset of the base model. (3) Again, the Optuna algorithm was used to obtain and output the corresponding evaluation indicators (based on the test set, i.e., 25% of all sample plots) of optimal performance of 10 regression models on the feature set after two-step feature selection. These evaluation indicators included the coefficient of determination (R^2, Equation (8)), the root mean square error (RMSE, Equation (9)), and the relative root mean square error (rRMSE, Equation (10)) [18,27,42]. This was repeated 10 times and we calculated the average of the above indicators. This is because the performance of the model largely depends on the selected hyperparameters; however, the values of these hyperparameters are often sensitive to the dataset, so the hyperparameters of Optuna need to be tuned separately for different feature sets. In addition, since random states cannot be completely controlled, they were not fixed in this study. This means that the initial state of model training and possible random factors in each experiment followed a natural (random) process, which was more in line with the real situation. However, through cross-validation and repeated

tests, randomness can be reduced and a more stable and representative evaluation index can be obtained.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{9}$$

$$rRMSE = (RMSE/\overline{y}) \times 100\% \tag{10}$$

where N is the sample size, y_i represents the true observed value, $\overline{y}$ represents the average of the true observed value, and $\hat{y}_i$ represents the predicted value.

Table 8. The Python packages on which the implementation of two-step feature selection depends.

Algorithm	Boruta	RFECV	RF	Optuna
package	Boruta (v0.3)	scikit-learn (v1.4.2)	scikit-learn (v1.4.2)	Optuna (v3.6.0)

4. Results

4.1. The Impact of Radiometric Terrain Correction on PolSAR Features

In order to quantitatively analyze the impact of the radiometric terrain correction (RTC) process on the correlation between forest AGB and PolSAR features, we first set the non-RTC (NRTC) data as the blank control and the RTC data as the test data. The core correction factors of the RTC process are shown in our previous study [40]. Secondly, we extracted a total of 117 effective PolSAR features (as mRFDI_db is invalid) based on NRTC and RTC data. Subsequently, Pearson correlation coefficients (R) were calculated separately between forest AGB and each of the PolSAR features (Table A1, Appendix B). Finally, in order to compare and visually present the changes in the R-values of the NRTC data and RTC data and, at the same time, reveal the relationship between various PolSAR features and forest AGB correlation strength after RTC, we sorted and plotted them according to the absolute R-values of the data (25 July 2020) after RTC (Figure 3).

The R-values of the RTC data increased to varying degrees in all (Figure 3a,b) but three original features (α, α_db, and A_db). The R-values of the 75 original features (with increased R-value) increased by an average of 0.26, and the largest R-value increase was 0.48 (σ^0_{HV}), indicating that RTC processing can effectively improve the accuracy of most PolSAR features and the sensitivity of AGB inversion. However, among the original features, the three (H, H_db, and A) that belong to the same H-A-alpha (HAα) decomposition method as α, α_db, and A_db were relatively less affected by RTC. This shows that HAα decomposition can alleviate terrain disturbance. This is because the three parameters of the HAα decomposition are defined by a function of the ratio class of the three eigenvalues (the relative scattering power of different scattering mechanisms). Moreover, the R-values of the RTC data in all but four derived features (Span, Span_db, BMI, and BMI_db) (Figure 3c) were reduced to varying degrees. This suggests that the ratio operation (feature derived algorithm) of features can eliminate some multiplicative interference (including terrain interference). It also indicates that there is an overcorrection problem in the RTC process. Although the overcorrection problem of RTC is small, it causes an error of 0.05 in the mean value of the absolute value of the Pearson correlation coefficient between forest AGB and these 35 derived features.

In addition, the feature R_HH/HV (R = 0.76) has the largest R-value among all the features derived through ratio operations, but it is still 0.1 lower than σ^0_{HV}_db (R = 0.86), which has the highest R-value of all 117 PolSAR features. This suggests that, based only on the backscattering coefficient, the HV polarization mode carries more polarization information related to forest AGB. Although the ratio algorithm of features can mitigate the topographic effects, it will also weaken some AGB-related information. The second-largest R-value of all 117 PolSAR features is Y4V_db (R = 0.84), which indicates that

the YAM4 decomposition method can decompose the scattering characteristics of the target into different scattering components relatively more accurately. Moreover, among the different components of the same polarization decomposition method, the Volume scattering component (Vol) usually has the largest R-value, indicating that the relationship between Vol and the scattering information in vegetation is relatively close. Furthermore, regardless of the decomposition method, the R-value of volume scattering in decibels is higher than that in linear units, which is consistent with the trend of previous studies [58].
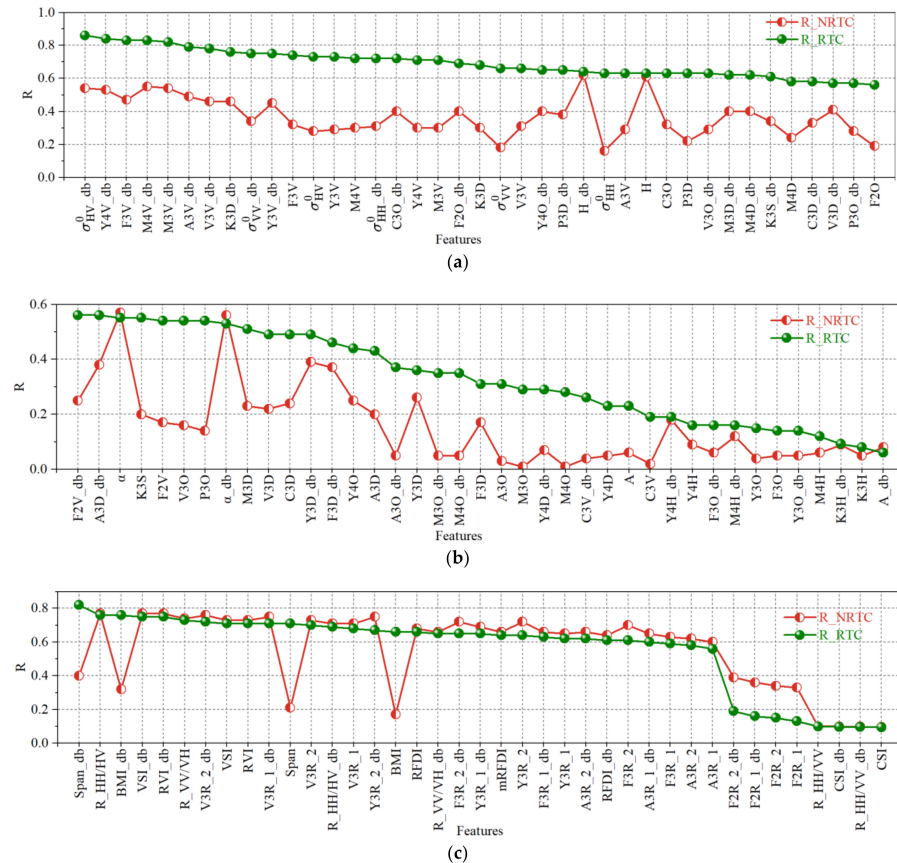

Figure 3. Absolute value of Pearson correlation coefficient (R) between forest AGB and the PolSAR features based on the data (25 July 2020) with radiometric terrain correction (RTC, olive) and non-RTC data (NRT, red). Sorted based on R_RTC (i.e., absolute value of R value between forest AGB and SAR features extracted based on RTC data). (**a**) The first set of the extracted original PolSAR features; (**b**) the second set of the extracted original PolSAR features (39 in total); (**c**) derived features based on PolSAR original features (39 in total).

4.2. The Impact of Radiometric Terrain Correction on Polarization Decomposition Component

In order to analyze the influence of different stages of RTC on the removal of topographic factors in the polarization decomposition component, we took the PolSAR data from 25 July 2020 as an example and plotted a scatter density plot between the decibel values of the three components of the Freeman three-decomposition method and the local incidence angle. As can be seen from Figure 4, there is a relatively obvious linear relationship between each component and the local incidence angle in the NRTC and POAC stages; but with the implementation of ESAC and AVEC, this linear relationship becomes significantly weaker, which is consistent with the results of our previous research regarding the backscattering coefficient [40]. The results show that RTC is also effective at removing the effect of terrain on the polarization decomposition.
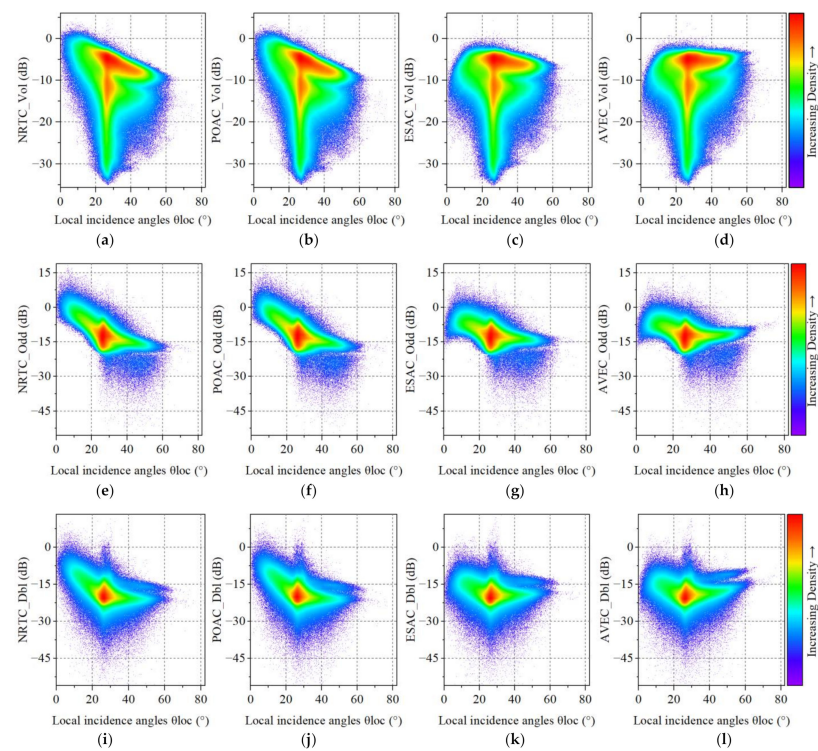
Figure 4. Taking PolSAR data from 25 July 2020 as an example, we created scatter density plots between the decibel values of the three components (Volume scattering component (Vol), Surface scattering component (Odd), and Double-bounce scattering component (Dbl)) of the Freeman three-decomposition in different topographic correction stages (non-radiometric terrain correction (NRTC), polarization orientation angle correction (POAC), effective scattering area correction (ESAC), and angular variation effect correction (AVEC)) and the local incidence angle θ_{loc}. (**a**) NRTC_Vol; (**b**) POAC_Vol; (**c**) ESAC_Vol; (**d**) AVEC_Vol; (**e**) NRTC_Odd; (**f**) POAC_Odd; (**g**) ESAC_Odd; (**h**) AVEC_Odd; (**i**) NRTC_Dbl; (**j**) POAC_Dbl; (**k**) ESAC_Dbl; (**l**) AVEC_Dbl.

In order to further analyze the specific mechanism of influence of different stages of RTC (POAC, ESAC, and AVEC) on each component of polarization decomposition, taking PolSAR data from 25 July 2020 as an example, we plotted the scatter density (Figure 5) of each component of Freeman3 decomposition at each stage of different RTC relative to the previous stage.

In the comparison stage between POAC and NRTC (Figure 5a,e,i), POAC caused the intensity values of almost all pixels of the Vol component (Figure 5a) to decrease, while the intensity values of almost all pixels of the Odd (Figure 5e) and Dbl components (Figure 5i) showed an increasing trend. This suggests that the absence of POAC will lead to an overestimation of Vol and an underestimation of Odd and Dbl. In addition, the correction of the ESAC and AVEC stages did not cause the pixel intensity value of the polarization decomposition component to undergo an overall increase or decrease, but there was an increase or decrease in the pixel value in each component. This is because the correction factors of ESAC and AVEC stages depend on both the value of the local incidence angle of each pixel and its relationship with the radar incidence angle, which also indicates that the correction effects of ESAC and AVEC stages are closely related to the ground conditions in the study area. However, in the ESAC phase, most of the pixels that cause a large change in the intensity value of the pixel (away from the 1:1 line) were decreasing, while in the AVEC phase, the situation was the opposite. This indicates that going without ESAC will lead to a more serious overestimation of some pixels, while going without AVEC will lead to a more serious underestimation of some pixels. Moreover, the effect of RTC on the three polarization decomposition components cannot be described by a simple law of increase

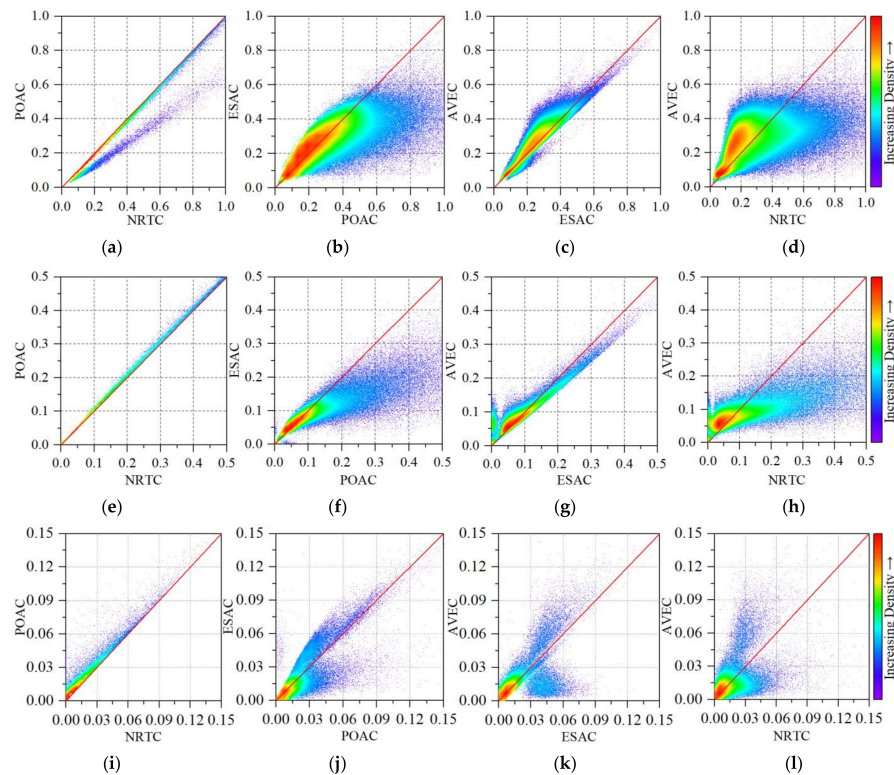and decrease because the correction effect of RTC is the superposition of the three stages (Figure 5d,h,l).

Figure 5. Taking PolSAR data from 25 July 2020 as an example, we created a scatter density plot for each component of Freeman three-decomposition (FRE3) at different radiometric terrain correction (RTC) stages (Y-axis) relative to the previous stage (X-axis), and in AVEC stages (that is, after all processing of the RTC was completed) with respect to non-RTC (NRTC). The three components of FRE3 are the Volume scattering component (Vol), Surface scattering component (Odd), and Double-bounce scattering component (Dbl). The three stages of RTC are polarization orientation angle correction (POAC), effective scattering area correction (ESAC), and angular variation effect correction (AVEC). The red line is a 1:1 line. (**a**) NRTC vs. POAC of Vol; (**b**) POAC vs. ESAC of Vol; (**c**) ESAC vs. AVEC of Vol; (**d**) NRTC vs. AVEC of Vol; (**e**) NRTC vs. POAC of Odd; (**f**) POAC vs. ESAC of Odd; (**g**) ESAC vs. AVEC of Odd; (**h**) NRTC vs. AVEC of Odd; (**i**) NRTC vs. POAC of Dbl; (**j**) POAC vs. ESAC of Dbl; (**k**) ESAC vs. AVEC of Dbl; (**l**) NRTC vs. AVEC of Dbl.

In addition, in order to comprehensively evaluate the universality and robustness of the effect of RTC on polarization decomposition, we used the Yamaguchi three-component (YAM3) method for analysis and mapping (Figure A1, Appendix A). On the whole, the trend of the scatter density plots of the two polarization decomposition methods was the same in each stage except for slight differences caused by the different decomposition algorithms. This shows that the effect of RTC on the components of polarization decomposition is relatively robust in different polarization decomposition methods.

4.3. The Impact of Radiometric Terrain Correction on Regression Model Performance

In order to analyze the impact of RTC on the performance of machine learning models and select the optimal regression model, we conducted feature selection (Figure A2, Appendix A) and model training (Table A2, Appendix B) for 10 regression models based on the NRTC data and RTC data (25 July 2020), respectively. Then, we conducted a descending order analysis and graphed the results (Figure 6) based on the R^2 value of the training results of the RTC data.
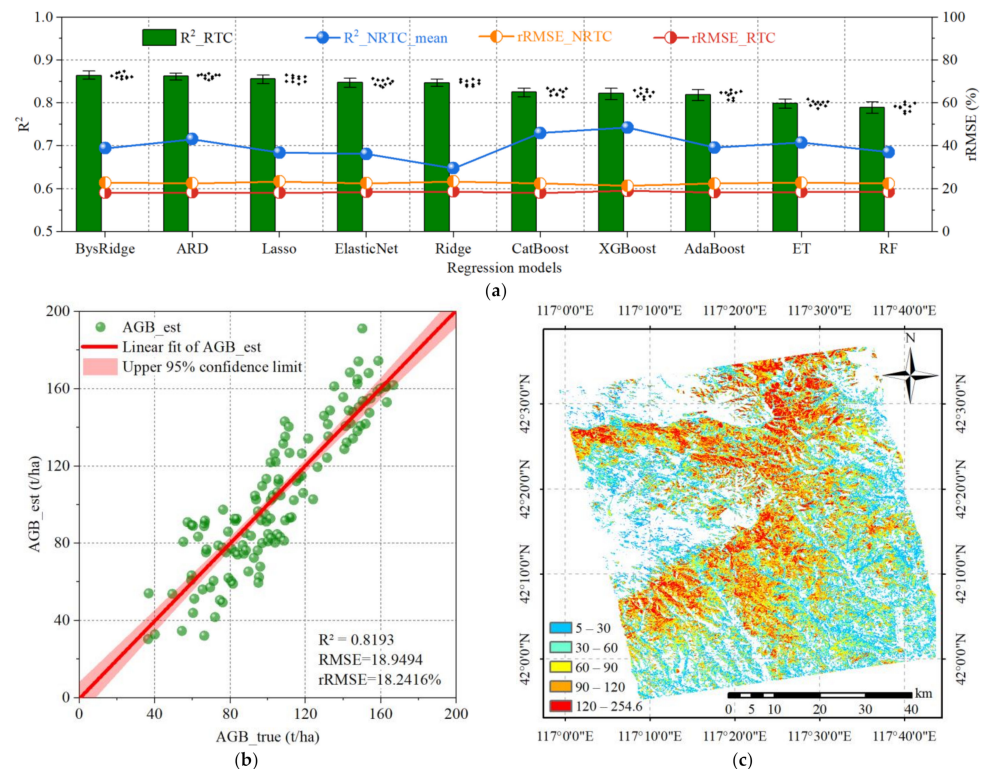
Figure 6. Analysis of the effectiveness of RTC and the optimal regression model of this study, taking the SAR data from 25 July 2020 as an example. (**a**) The training results of the NRTC and RTC data, where the black dots are the results of the corresponding single training; (**b**) scatter plot of the measured forest AGB and the AGB predicted by the optimal regression model (BysRidge); (**c**) spatial distribution map of forest AGB in the study area based on optimal model prediction.

Compared with training using the RTC data (Figure 6a), R^2 values obtained by different models based on training using the NRTC data are generally lower (0.14 lower on average), while rRMSE values are generally higher (4.2% higher on average), indicating that not implementing the RTC process will cause greater estimation error in forest AGB inversion. In addition, from the R^2 values of different models trained on the NRTC data and RTC data, it can be found that the RTC process generally improves the performance of linear models more than non-parametric models, indicating that the existence of terrain factors will cause the real relationship between SAR information and forest AGB to become more complex and difficult to describe by a simple linear relationship. Meanwhile, RTC can effectively reduce the signal distortion caused by terrain factors and make SAR data more closely reflect the real situation of forest biomass, thus reducing the complexity of data.

Figure 6b is a scatter plot of the measured forest AGB of 132 plots and the predicted value obtained using the BysRidge regression model (the largest R^2 value in 10 repetitions of Optuna hyperparameter optimizations, Equation (11)), which had the best performance on the data of this study, indicating that the model has a strong ability to explain the data trend (R^2 = 0.82). Its prediction error was relatively small (rRMSE = 18.06%), but there were still some problems of overestimation and underestimation. In addition, the multivariate model has a higher R^2 value (0.82) compared to the R^2 value of the univariate model (0.73 for the data dated 25 July 2020 in our previous study [40]), indicating that the multivariate model has a stronger advantage in retrieving forest AGB. Moreover, we evaluated the generalization ability of the model (Equation (11)) with respect to the PolSAR data from different dates (11 July and 8 August 2020), and created a scatter plot of the measured values and the predicted values of forest AGB calculated based on the model (Figure A3).

Finally, the forest AGB in the whole range of the study area was predicted and mapped (Figure 6c) based on the trained model.

$$
\begin{aligned}
AGB(db) = {}& 56.4598 + 3.5369 \times \sigma^0_{HV}\_db + 1.5128 \times BMI\_db + 3.5913 \times V3R\_2\_db \\
& + 23.2417 \times H + 2.0974 \times mRFDI + 2.7412 \times F3V\_db - 6.1481 \times RVI \\
& - 2.9493 \times Span\_db - 0.5524 \times R\_VV/HV\_db - 26.397 \times A3V
\end{aligned} \tag{11}
$$

5. Discussion

5.1. The Significance of Radiometric Terrain Correction

As an important part of the terrestrial ecosystem carbon pool [88], the accurate estimation of forest AGB is of great significance for quantifying the value of forest carbon storage and assessing the dynamics of the global carbon cycle, but the interference of topographic factors cannot be ignored [89]. This study found that the effective implementation of the RTC process can significantly suppress the interference of terrain factors with PolSAR data, thereby improving the quality and application potential of PolSAR data. This is specifically reflected in the following points. On the one hand, compared to the R-value between forest AGB and SAR features extracted from the NRTC data, the average R-value of 75 original features extracted from the RTC data increased by 0.26, and the maximum improvement was 0.48 (σ^0_{HV}). On the other hand, RTC is highly reasonable and robust at correcting components obtained by different polarization decomposition methods (FRE3 and YAM3 are examples). Moreover, results consistent with those of other researchers [90–92] were obtained during the POAC stage: the Vol component decreased, while the Odd and Dbl components increased. However, the correction effect on each ground unit during the ESAC and AVEC stages was closely related to the ground conditions in the study area and did not exhibit a monotonic trend. Finally, the fitting performance of 10 regression models based on RTC data was generally superior to that based on NRTC data (R^2 increased by 0.14 and rRMSE decreased by 4.2% on average). These results indicate that RTC is effective and robust for improving the accuracy of PolSAR features and the reliability of forest AGB inversion. This is because RTC is a set of effective step-by-step terrain correction processes designed by in-depth analysis of the SAR physical mechanism, which accurately corrects the interference caused by terrain factors in terms of the scattering mechanism (POAC and AVEC) and intensity information (ESAC). The original PolSAR features after RTC processing carry more accurate polarization scattering information, thus improving the sensitivity of AGB inversion. In addition, if a suitable RTC process is created in the application of other SAR technologies to eliminate the interference of terrain factors, the application effect will be improved to a certain extent.

5.2. The Overcorrection of Radiometric Terrain Correction

This study also found that, without RTC processing, the R-value of the derived features extracted based on ratio arithmetic is significantly higher than that of the original features [86], but the original features showed significant improvement after RTC. This indicates that ratio-derived features can be prioritized for SAR data application when RTC is not feasible [93,94]. However, compared with the derived features without RTC, the average R-value of the 35 ratio-related PolSAR-derived features after RTC decreased by 0.05, indicating that there are some overcorrection issues in the RTC process that need to be improved. In addition, since the ratio operation can eliminate multiplicative noise, there is a greater likelihood of overcorrection due to POAC and AVEC. The extraction of POA needs to distinguish between the offset caused by forest targets and terrain factors, and only correct the offset caused by terrain, while the correction factor in the AVEC stage only considers the local incidence angle and radar incidence angle, without introducing factors such as surface roughness, terrain slope, and complex dielectric coefficient that are related to the backscatter coefficient in the AVEC correction process. Moreover, there is potential space for further improvement of the ESAC method. Therefore, exploring how

to effectively solve the overcorrection problems caused by the above factors in the RTC process is a valuable future research direction.

5.3. Performance Comparison of Different Regression Models

Improving dataset quality (such as RTC) has a greater impact on improving AGB estimation performance than selecting regression algorithms [95]. However, after the significant improvement in data quality, the selection of machine learning regression models becomes particularly crucial. Selecting an appropriate regression model can ensure that the algorithm can fully utilize the features in the data, capture the relationships between variables, and avoid overfitting or underfitting, thereby further improving the prediction accuracy and generalization ability of the model. This study compared 10 different regression models and found that the linear model has superior complexity (feature dimension) and interpretability compared to the non-parametric model. This is because non-parametric models do not preset the specific form of the model but adjust flexibly based on the data. However, it is precisely because of their higher flexibility that these models become too complex. In the feature selection stage, the variance inflation factor (VIF) values of the features of different linear regression models are all less than 10 (Figure A2d), indicating that regularization methods (L1 or L2) can effectively alleviate the feature collinearity problem in multivariate linear models. However, the VIF value of many features is greater than 5, indicating that there is still a moderate degree of collinearity problem in the feature set. In addition, it is to be expected that there is some degree of collinearity in the features selected by the non-parametric model (Figure A2e), but the collinearity problem usually has less effect on the non-parametric model.

In addition, the R^2 and rRMSE indicators of the non-parametric models based on the RTC data in this study are usually worse than those of the linear models. This may be because the target variable of this study is the forest AGB value in decibels, which has a better linear relationship with PolSAR features (higher accuracy) [7,58,96]. However, non-parametric models based on the NRTC data in this study typically have better R^2 and rRMSE metrics than linear models. This is because terrain factors cause PolSAR features to have a more obviously nonlinear relationship with forest AGB. Moreover, the setting of hyperparameter spaces for different models directly affects the performance of the model, while the setting of hyperparameter spaces for non-parametric models is more complex, which may be one reason why non-parametric models performed slightly worse on the data in this study. Furthermore, the comparison between parametric and non-parametric models may be unfair because there is not a large amount of data to train the non-parametric model, resulting in a feature selection method that is not ideal. Therefore, k-Nearest Neighbor is a regression algorithm worth trying because it is more suitable for small datasets than random forest. In addition, deep learning models (such as Convolutional Neural Networks (CNN) and Deep Neural Networks (DNN)) have certain limitations (such as a large amount of data being required for training, the consumption of computing resources being large, and the model interpretation being poor) compared to traditional machine learning models. However, because of their strong feature learning ability, high accuracy, and excellent generalization ability, they have shown great potential for solving complex problems, and so are worth putting into use in scenarios with sufficient resources.

5.4. Potential Limitations

In this study, only PolSAR features were applied in AGB inversion. However, it is necessary to include more actual parameters that reflect the geographical environment and physical conditions when constructing a more comprehensive and accurate AGB model [44]. For example, climatic factors (temperature, rainfall, humidity, etc.) may directly affect the state of the land surface and vegetation, thereby altering the reflection characteristics of SAR signals [97]. Moreover, although RTC is carried out in the pre-processing of PolSAR data, it may also be worth taking slope, aspect, and local incidence angle into account in model training. In addition, more polarization parameters also contribute to the application of

SAR data [67,98,99]. Through the comprehensive analysis of these multivariate parameters, we can expect a significant improvement in the predictive power of the model and a deeper understanding of the object of study.

In this study, data from only 132 plots were used to build and train machine learning models due to human constraints and real-world conditions. The relatively small forest AGB range (36.39–166.58 t/ha) of the real test sample may not be suitable for the forest in the study area (the maximum value is 254.6 t/ha). However, adding more plot data through unmanned aerial vehicle-light detecting and ranging (UAV-LiDAR) systems can significantly improve the generalization and reliability of the model, as larger datasets more accurately reflect the ecological diversity and complexity of the study area, helping the model capture more patterns and trends [5]. Furthermore, the control of data quality is crucial. It is necessary to strictly control field measurement errors, AGB calculation errors, and matching errors between sample location and remote sensing image data to reduce the uncertainty caused by human factors, so as to comprehensively improve the overall quality of data and the accuracy of the model prediction [100].

In addition, this study did not conduct differentiated modeling for different forest types. However, the AGB modeling method based on forest type is a forest AGB modeling strategy that has attracted much attention in recent years. Because it is based on the basic characteristics of forest ecosystem diversity and heterogeneity, which is helpful to reveal the biomass distribution law and the fine modeling of forest AGB, so it has significant advantages in terms of the accuracy and practicability of the AGB inversion model [101,102].

6. Conclusions

Considering that the impact of RTC on the correctness of feature extraction and regression model performance needed to be further evaluated, this study took PolSAR data provided by ALOS-2 as the data source, and used 10 machine learning regression models and the Optuna hyperparameter optimization algorithm to verify the effectiveness and robustness of RTC in improving the correctness of the PolSAR feature set and the accuracy of forest AGB inversion. The detailed results are as follows: The RTC process effectively and robustly improved the correlation between the PolSAR features and the measured forest AGB (with an average increase of 0.26 in R-values) and the performance of the regression models (with an average increase of 0.14 in R^2 values and an average decrease of 4.2% in rRMSE). The RTC process also showed prime robustness and rationality for the correction of different polarization decomposition components. The ratio operation of PolSAR features offset some multiplicative noises (including terrain), which indicates that there was a certain degree of overcorrection in the RTC process. Therefore, the ratio class-derived features were given priority when there was no condition for RTC; after RTC, $\sigma^0_{HV\_db}$ still had the highest R-value of all PolSAR features and forest AGB decibels. In addition, the PolSAR features had a closer linear relationship with the forest AGB of decibel states. In this case, the linear model remained a powerful and practical choice due to its efficiency and stability. For example, the optimal regression model in this study was BysRidge (R^2 = 0.82, rRMSE = 18.06%).

However, this study treated all forest cover areas as the same type without differentiated modeling of different forest types, and confirmed that there is an overcorrection problem in the RTC process. More PolSAR features and actual parameters, as well as more accurate sample data, should be applied in machine learning inversion of forest parameters. Although this study had these limitations, it also provides insight for future research. It is expected to significantly improve the inversion accuracy of forest AGB by overcoming the above limitations, which can provide more reliable data support for the scientific management and protection of forest resources.

In summary, this study validated the effectiveness and robustness of RTC in improving the correct extraction of PolSAR features and accurately inverting forest AGB, and analyzed the potential mechanism of RTC on polarization decomposition parameters. The method used in this study has good robustness and universality. It can therefore provide assistance

and reliable reference values for the inversion of forest AGB and related SAR applications based on machine learning models using PolSAR data in complex terrain areas.

Author Contributions: Conceptualization, Y.N. and W.F.; methodology, Y.N.; software, Y.N. and Y.W.; validation, Y.N.; formal analysis, Y.N.; investigation, Y.N. and W.F.; resources, W.F.; data curation, Y.N. and R.S.; writing—original draft preparation, Y.N.; writing—review and editing, Y.N., W.F. and S.C.; visualization, Y.N. and Y.H.; project administration, Y.N. and W.F.; funding acquisition, W.F. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy considerations.

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

Appendix A
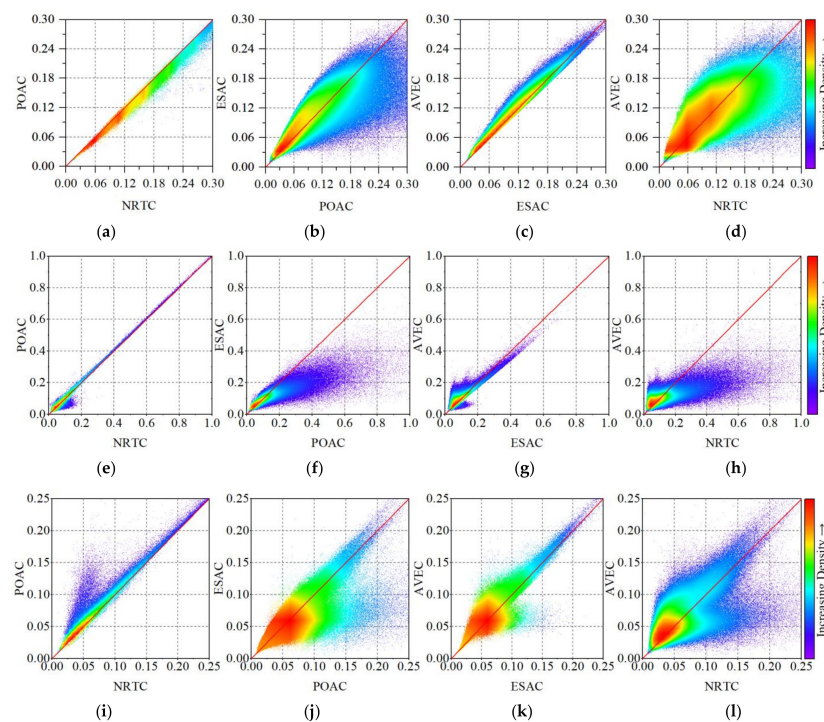
Figure A1. The scatter density plot of each component of Yamaguchi three-component (YAM3) at different radiometric terrain correction (RTC) stages (Y-axis) relative to the previous stage (X-axis), and in AVEC stages (that is, after all processing of the RTC) with respect to non-RTC (NRTC). The three components of YAM3 are the Volume scattering component (Vol), Surface scattering component (Odd), and Double-bounce scattering component (Dbl). The three stages of RTC are polarization orientation angle correction (POAC), effective scattering area correction (ESAC), and angular variation effect correction (AVEC). The red line is a 1:1 line. (**a**) NRTC vs. POAC of Vol; (**b**) POAC vs. ESAC of Vol; (**c**) ESAC vs. AVEC of Vol; (**d**) NRTC vs. AVEC of Vol; (**e**) NRTC vs. POAC of Odd; (**f**) POAC vs. ESAC of Odd; (**g**) ESAC vs. AVEC of Odd; (**h**) NRTC vs. AVEC of Odd; (**i**) NRTC vs. POAC of Dbl; (**j**) POAC vs. ESAC of Dbl; (**k**) ESAC vs. AVEC of Dbl; (**l**) NRTC vs. AVEC of Dbl.

Figure A2. The result of feature selection: (**a**) the 32 features selected in preliminary feature selection (Boruta algorithm) based on radiative terrain correction (RTC) data, including the importance score given by the RF of the selected features, and absolute values of Pearson correlation coefficients (R) between the selected features and measured forest AGB; (**b**) the 21 features selected in preliminary feature selection (Boruta algorithm) based on non-RTC (NRTC) data, including the importance score given by the RF of the selected features, and absolute values of Pearson correlation coefficients (R) between the selected features and measured forest AGB; (**c**) the number of features selected in the second step feature selection (RFECV algorithm) based on RTC and NRTC data; (**d**) the features selected in different multivariate linear models and the variance inflation factor (VIF) value corresponding to each feature; (**e**) the features selected in different non-parametric models.
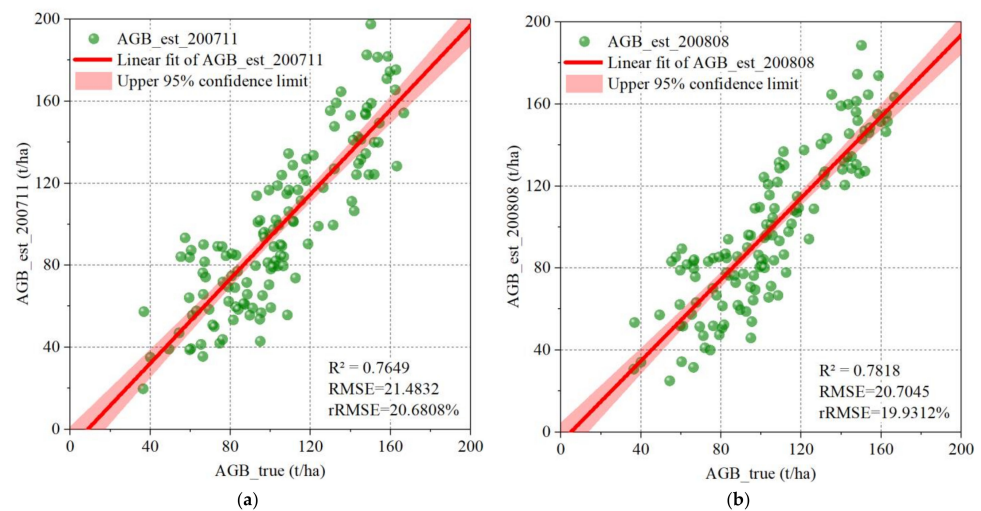
Figure A3. Scatter plot of measured forest AGB and predicted forest AGB. The prediction model is an optimal regression model based on the PolSAR data processed by radiometric terrain correction (RTC) from 25 July 2020. (**a**) The independent variable of the prediction model was derived from the PolSAR data (after RTC processing) from 11 July 2020. (**b**) The independent variable of the prediction model was derived from the PolSAR data (after RTC processing) from 8 August 2020.

Appendix B

Table A1. Pearson correlation coefficient (R) between forest AGB and the 118 PolSAR features based on the data with radiometric terrain correction (RTC) and non-RTC data (NRTC), obtained from 11 July, 25 July, and 8 August 2020.

ID	Features	NRTC/RTC			Significance
		R_200711	R_200725	R_200808	
1	σ_{HH}^0	0.12/0.58	0.16/0.63	0.14/0.60	-/**
2	σ_{HV}^0	0.25/0.70	0.28/0.73	0.27/0.71	*/**
3	σ_{VV}^0	0.15/0.61	0.18/0.66	0.17/0.62	-/**
4	Span	0.18/0.66	0.21/0.71	0.20/0.69	-/**
5	R_HH/VV	−0.09/−0.09	−0.10/−0.10	−0.10/−0.10	-/-
6	R_HH/HV	−0.74/−0.73	−0.77/−0.76	−0.76/−0.75	**/**
7	R_VV/VH	−0.70/−0.69	−0.74/−0.73	−0.72/−0.70	**/**
8	BMI	0.14/0.62	0.17/0.66	0.15/0.65	-/**
9	VSI	0.70/0.67	0.73/0.71	0.71/0.68	**/**
10	RVI	0.70/0.67	0.73/0.71	0.71/0.68	**/**
11	CSI	0.091/0.089	0.096/0.094	0.096/0.095	-/-
12	RFDI	−0.65/−0.64	−0.68/−0.66	−0.67/−0.65	**/**
13	mRFDI	−0.63/−0.61	−0.66/−0.64	−0.65/−0.62	**/**
14	F2O	0.14/0.51	0.19/0.56	0.16/0.53	-/**
15	F2V	0.14/0.51	0.17/0.54	0.14/0.52	-/**
16	F3D	0.15/0.30	0.17/0.31	0.19/0.34	-/*
17	F3O	0.04/0.11	0.05/0.14	0.03/0.10	-/-
18	F3V	0.30/0.72	0.32/0.74	0.31/0.74	*/**
19	Y3D	0.22/0.34	0.26/0.36	0.24/0.35	-/*
20	Y3O	−0.03/0.10	−0.04/0.15	−0.04/0.11	-/-
21	Y3V	0.25/0.69	0.29/0.73	0.28/0.70	-/**
22	Y4D	−0.04/0.21	−0.05/0.23	−0.04/0.21	-/-
23	Y4O	−0.24/−0.44	−0.25/−0.44	−0.27/−0.45	-/*
24	Y4V	0.31/0.71	0.30/0.71	0.30/0.70	*/**
25	Y4H	0.08/0.15	0.09/0.16	0.09/0.18	-/-
26	M3D	0.21/0.49	0.23/0.51	0.20/0.51	-/**

Table A1. *Cont.*

ID	Features	NRTC/RTC			Significance
		R_200711	R_200725	R_200808	
27	M3O	0.01/0.25	0.01/0.29	0.01/0.28	-/-
28	M3V	0.32/0.72	0.30/0.71	0.35/0.74	*/**
29	M4D	0.25/0.60	0.24/0.58	0.22/0.57	-/**
30	M4O	0.01/0.28	0.01/0.28	0.01/0.30	-/-
31	M4V	0.29/0.71	0.30/0.72	0.32/0.74	*/**
32	M4H	0.06/0.13	0.06/0.12	0.09/0.15	-/-
33	A3D	0.22/0.45	0.20/0.43	0.18/0.40	-/*
34	A3O	0.03/0.32	0.03/0.31	0.03/0.30	-/-
35	A3V	0.31/0.64	0.29/0.63	0.30/0.63	-/**
36	V3D	0.20/0.46	0.22/0.49	0.22/0.51	-/*
37	V3O	0.15/0.54	0.16/0.54	0.18/0.55	-/**
38	V3V	0.32/0.68	0.31/0.66	0.31/0.65	*/**
39	H	0.61/0.62	0.61/0.63	0.60/0.63	**/**
40	A	$-0.07/-0.23$	$-0.06/-0.23$	$-0.07/-0.25$	-/-
41	α	0.55/0.51	0.57/0.55	0.57/0.56	**/**
42	C3D	0.24/0.48	0.24/0.49	0.26/0.50	-/*
43	C3O	0.30/0.60	0.32/0.63	0.31/0.60	-/**
44	C3V	$-0.01/0.15$	$-0.02/0.19$	$-0.02/0.20$	-/-
45	K3S	0.18/0.52	0.20/0.55	0.21/0.55	-/**
46	K3D	0.31/0.70	0.30/0.68	0.30/0.67	*/**
47	K3H	0.05/0.08	0.05/0.08	0.05/0.07	-/-
48	P3D	0.20/0.60	0.22/0.63	0.22/0.65	-/**
49	P3O	0.13/0.52	0.14/0.54	0.15/0.55	-/**
50	A3R_1	0.60/0.55	0.60/0.56	0.61/0.58	**/**
51	A3R_2	0.62/0.59	0.62/0.58	0.63/0.60	**/**
52	F2R_1	$-0.31/-0.12$	$-0.33/-0.13$	$-0.37/-0.15$	-/-
53	F2R_2	$-0.32/-0.15$	$-0.34/-0.15$	$-0.38/-0.17$	-/-
54	F3R_1	0.61/0.58	0.63/0.59	0.59/0.55	**/**
55	F3R_2	0.69/0.60	0.70/0.61	0.71/0.61	**/**
56	V3R_1	0.70/0.66	0.71/0.68	0.69/0.66	**/**
57	V3R_2	0.72/0.70	0.73/0.70	0.71/0.69	**/**
58	Y3R_1	0.65/0.61	0.65/0.62	0.64/0.60	**/**
59	Y3R_2	0.70/0.64	0.72/0.64	0.70/0.63	**/**
60	$\sigma^0_{HH}\_db$	0.30/0.70	0.31/0.72	0.31/0.71	*/**
61	$\sigma^0_{HV}\_db$	0.50/0.81	0.54/0.86	0.52/0.82	**/**
62	$\sigma^0_{VV}\_db$	0.32/0.72	0.34/0.75	0.34/0.73	*/**
63	Span_db	0.38/0.79	0.40/0.82	0.40/0.80	*/**
64	R_HH/VV_db	$-0.01/-0.01$	$-0.098/-0.096$	$-0.10/-0.10$	-/-
65	R_HH/HV_db	$-0.57/-0.65$	$-0.71/-0.69$	$-0.70/-0.67$	**/**
66	R_VV/VH_db	$-0.62/-0.60$	$-0.66/-0.65$	$-0.64/-0.63$	**/**
67	BMI_db	0.30/0.73	0.32/0.76	0.30/0.75	*/**
68	VSI_db	0.75/0.74	0.77/0.75	0.78/0.76	**/**
69	RVI_db	0.75/0.74	0.77/0.75	0.78/0.76	**/**
70	CSI_db	0.10/0.09	0.099/0.097	0.10/0.10	-/-
71	RFDI_db	$-0.62/-0.60$	$-0.64/-0.61$	$-0.65/-0.62$	**/**
72	mRFDI_db	NaN/NaN	NaN/NaN	NaN/NaN	-/-
73	F2O_db	0.36/0.66	0.40/0.69	0.38/0.67	*/**
74	F2V_db	0.23/0.53	0.25/0.56	0.25/0.54	-/**
75	F3D_db	0.35/0.45	0.37/0.46	0.38/0.48	*/*
76	F3O_db	$-0.05/0.14$	$-0.06/0.16$	$-0.04/0.11$	-/-
77	F3V_db	0.45/0.81	0.47/0.83	0.46/0.83	*/**
78	Y3D_db	0.36/0.47	0.39/0.49	0.37/0.47	*/*
79	Y3O_db	$-0.03/0.10$	$-0.05/0.14$	$-0.05/0.12$	-/-
80	Y3V_db	0.42/0.72	0.45/0.75	0.43/0.72	*/**
81	Y4D_db	$-0.05/0.27$	$-0.07/0.29$	$-0.05/0.27$	-/-
82	Y4O_db	$-0.39/-0.65$	$-0.40/-0.65$	$-0.42/-0.67$	*/**

Table A1. *Cont.*

ID	Features	NRTC/RTC			Significance
		R_200711	**R_200725**	**R_200808**	
83	Y4V_db	0.54/0.84	0.53/0.84	0.53/0.83	**/**
84	Y4H_db	0.16/0.17	0.18/0.19	0.18/0.20	-/-
85	M3D_db	0.39/0.60	0.40/0.62	0.38/0.62	*/**
86	M3O_db	0.04/0.32	0.05/0.35	0.05/0.34	-/*
87	M3V_db	0.55/0.83	0.54/0.82	0.58/0.85	**/**
88	M4D_db	0.41/0.63	0.40/0.62	0.39/0.60	*/**
89	M4O_db	0.05/0.35	0.05/0.35	0.05/0.37	-/*
90	M4V_db	0.55/0.82	0.55/0.83	0.56/0.85	**/**
91	M4H_db	0.12/0.17	0.12/0.16	0.15/0.18	-/-
92	A3D_db	0.39/0.57	0.38/0.56	0.35/0.52	*/**
93	A3O_db	0.05/0.38	0.05/0.37	0.04/0.35	-/*
94	A3V_db	0.50/0.79	0.49/0.79	0.49/0.80	**/**
95	V3D_db	0.38/0.55	0.41/0.57	0.41/0.59	*/**
96	V3O_db	0.28/0.63	0.29/0.63	0.30/0.64	-/**
97	V3V_db	0.46/0.79	0.46/0.78	0.45/0.76	**/**
98	H_db	0.62/0.63	0.62/0.64	0.61/0.64	**/**
99	A_db	−0.09/−0.06	0.08/−0.06	0.09/−0.08	-/-
100	α_db	0.54/0.50	0.56/0.53	0.56/0.54	**/**
101	C3D_db	0.33/0.57	0.33/0.58	0.35/0.59	*/**
102	C3O_db	0.37/0.70	0.40/0.72	0.38/0.70	*/**
103	C3V_db	0.02/0.23	0.04/0.26	0.04/0.27	-/-
104	K3S_db	0.31/0.58	0.34/0.61	0.35/0.61	*/**
105	K3D_db	0.47/0.78	0.46/0.76	0.46/0.75	*/**
106	K3H_db	0.09/0.09	0.089/0.092	0.09/0.09	-/-
107	P3D_db	0.35/0.63	0.38/0.65	0.38/0.67	*/**
108	P3O_db	0.27/0.55	0.28/0.57	0.28/0.58	-/**
109	A3R_1_db	0.65/0.59	0.65/0.60	0.66/0.63	**/**
110	A3R_2_db	0.66/0.63	0.66/0.62	0.67/0.64	**/**
111	F2R_1_db	−0.32/−0.15	−0.36/−0.16	−0.39/−0.18	*/-
112	F2R_2_db	−0.39/−0.19	−0.39/−0.19	−0.39/−0.19	*/-
113	F3R_1_db	0.64/0.62	0.66/0.63	0.61/0.60	**/**
114	F3R_2_db	0.71/0.64	0.72/0.65	0.73/0.65	**/**
115	V3R_1_db	0.73/0.69	0.75/0.71	0.73/0.70	**/**
116	V3R_2_db	0.75/0.72	0.76/0.72	0.75/0.71	**/**
117	Y3R_1_db	0.69/0.64	0.69/0.65	0.67/0.62	**/**
118	Y3R_2_db	0.73/0.67	0.75/0.67	0.73/0.66	**/**

Note: ** means extremely significant ($p \leq 0.01$), * means significant ($p \leq 0.05$), - means did not pass the significance test ($p > 0.05$). NaN means Not a Number, because mRFDI has a partial negative value, so it is impossible to calculate the feature of its decibel unit (mRFDI_db).

Table A2. The average value of the corresponding evaluation index of the optimal performance obtained by repeating 10 times: two-step feature selection was carried out using the radiometric terrain correction (RTC) and non-RTC (NRTC) datasets, and then 10 regression models and the Optuna algorithm were used to obtain the corresponding evaluation indicators of optimal performance.

Model	R^2_RTC	R^2_NRTC	rRMSE_RTC (%)	rRMSE_NRTC (%)
BysRidge	0.86334	0.69453	18.11756	22.89453
ARD	0.86213	0.71534	18.31867	22.57860
Lasso	0.85586	0.68453	18.13603	23.37506
ElasticNet	0.84783	0.68124	18.51316	22.56761
Ridge	0.84704	0.64753	18.74611	23.45273
CatBoost	0.82521	0.73013	18.14501	22.43760
XGBoost	0.82202	0.74273	19.01550	21.37638
AdaBoost	0.81934	0.69574	18.37307	22.45734
ET	0.79859	0.70750	18.56786	22.89376
RF	0.78988	0.68545	18.53375	22.45703

References

1. Xiao, J.; Chevallier, F.; Gomez, C.; Guanter, L.; Hicke, J.A.; Huete, A.R.; Ichii, K.; Ni, W.; Pang, Y.; Rahman, A.F.; et al. Remote Sensing of the Terrestrial Carbon Cycle: A Review of Advances over 50 Years. *Remote Sens. Environ.* **2019**, *233*, 111383. [CrossRef]
2. Sugden, A.; Smith, J.; Pennisi, E. The Future of Forests. *Science* **2008**, *320*, 1435. [CrossRef]
3. Lu, D. The Potential and Challenge of Remote Sensing-based Biomass Estimation. *Int. J. Remote Sens.* **2006**, *27*, 1297–1328. [CrossRef]
4. Le Toan, T.; Quegan, S.; Davidson, M.W.J.; Balzter, H.; Paillou, P.; Papathanassiou, K.; Plummer, S.; Rocca, F.; Saatchi, S.; Shugart, H.; et al. The BIOMASS Mission: Mapping Global Forest Biomass to Better Understand the Terrestrial Carbon Cycle. *Remote Sens. Environ.* **2011**, *115*, 2850–2860. [CrossRef]
5. Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Li, G.; Moran, E. A Survey of Remote Sensing-Based Aboveground Biomass Estimation Methods in Forest Ecosystems. *Int. J. Digit. Earth* **2016**, *9*, 63–105. [CrossRef]
6. Migolet, P.; Goïta, K.; Pambo, A.F.K.; Mambimba, A.N. Estimation of the Total Dry Aboveground Biomass in the Tropical Forests of Congo Basin Using Optical, LiDAR, and Radar Data. *Gisci. Remote Sens.* **2022**, *59*, 431–460. [CrossRef]
7. Chen, Q.; McRoberts, R.E.; Wang, C.; Radtke, P.J. Forest Aboveground Biomass Mapping and Estimation across Multiple Spatial Scales Using Model-Based Inference. *Remote Sens. Environ.* **2016**, *184*, 350–360. [CrossRef]
8. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A Tutorial on Synthetic Aperture Radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [CrossRef]
9. Soja, M.J.; Quegan, S.; d'Alessandro, M.M.; Banda, F.; Scipal, K.; Tebaldini, S.; Ulander, L.M.H. Mapping Above-Ground Biomass in Tropical Forests with Ground-Cancelled P-Band SAR and Limited Reference Data. *Remote Sens. Environ.* **2021**, *253*, 112153. [CrossRef]
10. Tsokas, A.; Rysz, M.; Pardalos, P.M.; Dipple, K. SAR Data Applications in Earth Observation: An Overview. *Expert Syst. Appl.* **2022**, *205*, 117342. [CrossRef]
11. Yan, X.; Li, J.; Smith, A.R.; Yang, D.; Ma, T.; Su, Y.; Shao, J. Evaluation of Machine Learning Methods and Multi-Source Remote Sensing Data Combinations to Construct Forest above-Ground Biomass Models. *Int. J. Digit. Earth* **2023**, *16*, 4471–4491. [CrossRef]
12. Zhang, R.; Zhou, X.; Ouyang, Z.; Avitabile, V.; Qi, J.; Chen, J.; Giannico, V. Estimating Aboveground Biomass in Subtropical Forests of China by Integrating Multisource Remote Sensing and Ground Data. *Remote Sens. Environ.* **2019**, *232*, 111341. [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. Feng, Y.; Lu, D.; Chen, Q.; Keller, M.; Moran, E.; dos-Santos, M.N.; Bolfe, E.L.; Batistella, M. Examining Effective Use of Data Sources and Modeling Algorithms for Improving Biomass Estimation in a Moist Tropical Forest of the Brazilian Amazon. *Int. J. Digit. Earth* **2017**, *10*, 996–1016. [CrossRef]
15. Shao, Z.; Zhang, L.; Wang, L. Stacked Sparse Autoencoder Modeling Using the Synergy of Airborne LiDAR and Satellite Optical and SAR Data to Map Forest Above-Ground Biomass. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2017**, *10*, 5569–5582. [CrossRef]
16. Liao, Z.; He, B.; Quan, X.; Van Dijk, A.I.J.M.; Qiu, S.; Yin, C. Biomass Estimation in Dense Tropical Forest Using Multiple Information from Single-Baseline P-Band PolInSAR Data. *Remote Sens. Environ.* **2019**, *221*, 489–507. [CrossRef]
17. Solberg, S.; Hansen, E.H.; Gobakken, T.; Næssset, E.; Zahabu, E. Biomass and InSAR Height Relationship in a Dense Tropical Forest. *Remote Sens. Environ.* **2017**, *192*, 166–175. [CrossRef]
18. Zeng, P.; Zhang, W.; Li, Y.; Shi, J.; Wang, Z. Forest Total and Component Above-Ground Biomass (AGB) Estimation through C- and L-Band Polarimetric SAR Data. *Forests* **2022**, *13*, 442. [CrossRef]
19. Luo, H.; Qin, S.; Li, J.; Lu, C.; Yue, C.; Ou, G. High-Density Forest AGB Estimation in Tropical Forest Integrated with PolInSAR Multidimensional Features and Optimized Machine Learning Algorithms. *Ecol. Indic.* **2024**, *160*, 111878. [CrossRef]
20. Chowdhury, T.A.; Thiel, C.; Schmullius, C. Growing Stock Volume Estimation from L-Band ALOS PALSAR Polarimetric Coherence in Siberian Forest. *Remote Sens. Environ.* **2014**, *155*, 129–144. [CrossRef]
21. Thiel, C.; Schmullius, C. The Potential of ALOS PALSAR Backscatter and InSAR Coherence for Forest Growing Stock Volume Estimation in Central Siberia. *Remote Sens. Environ.* **2016**, *173*, 258–273. [CrossRef]
22. Eini-Zinab, S.; Maghsoudi, Y.; Sayedain, S.A. Assessing the Performance of Indicators Resulting from Three-Component Freeman–Durden Polarimetric SAR Interferometry Decomposition at P-and L-Band in Estimating Tropical Forest Aboveground Biomass. *Int. J. Remote Sens.* **2020**, *41*, 433–454. [CrossRef]
23. Tebaldini, S. Single and Multipolarimetric SAR Tomography of Forested Areas: A Parametric Approach. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2375–2387. [CrossRef]
24. Li, W.; Zhang, Y.; Zhang, J.; Chen, H.; Chen, E.; Zhao, L.; Zhao, D. Tropical Forest AGB Estimation Based on Structure Parameters Extracted by TomoSAR. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *121*, 103369. [CrossRef]
25. Zhang, H.; Wang, C.; Zhu, J.; Fu, H.; Han, W.; Xie, H. Forest Aboveground Biomass Estimation in Subtropical Mountain Areas Based on Improved Water Cloud Model and PolSAR Decomposition Using L-Band PolSAR Data. *Forests* **2023**, *14*, 2303. [CrossRef]
26. Thapa, B.; Lovell, S.; Wilson, J. Remote Sensing and Machine Learning Applications for Aboveground Biomass Estimation in Agroforestry Systems: A Review. *Agroforest. Syst.* **2023**, *97*, 1097–1111. [CrossRef]
27. Zhang, Y.; Ma, J.; Liang, S.; Li, X.; Li, M. An Evaluation of Eight Machine Learning Regression Algorithms for Forest Aboveground Biomass Estimation from Multiple Satellite Data Products. *Remote Sens.* **2020**, *12*, 4015. [CrossRef]

28. Gleason, C.J.; Im, J. Forest Biomass Estimation from Airborne LiDAR Data Using Machine Learning Approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91. [CrossRef]

29. Zhang, J.; Zhou, C.; Zhang, G.; Yang, Z.; Pang, Z.; Luo, Y. A Novel Framework for Forest Above-Ground Biomass Inversion Using Multi-Source Remote Sensing and Deep Learning. *Forests* **2024**, *15*, 456. [CrossRef]

30. Zhang, L.; Shao, Z.; Liu, J.; Cheng, Q. Deep Learning Based Retrieval of Forest Aboveground Biomass from Combined LiDAR and Landsat 8 Data. *Remote Sens.* **2019**, *11*, 1459. [CrossRef]

31. Oehmcke, S.; Li, L.; Trepekli, K.; Revenga, J.C.; Nord-Larsen, T.; Gieseke, F.; Igel, C. Deep Point Cloud Regression for Above-Ground Forest Biomass Estimation from Airborne LiDAR. *Remote Sens. Environ.* **2024**, *302*, 113968. [CrossRef]

32. Cartus, O.; Santoro, M.; Kellndorfer, J. Mapping Forest Aboveground Biomass in the Northeastern United States with ALOS PALSAR Dual-Polarization L-Band. *Remote Sens. Environ.* **2012**, *124*, 466–478. [CrossRef]

33. Peregon, A.; Yamagata, Y. The Use of ALOS/PALSAR Backscatter to Estimate above-Ground Forest Biomass: A Case Study in Western Siberia. *Remote Sens. Environ.* **2013**, *137*, 139–146. [CrossRef]

34. Kumar, B.; Dikshit, O.; Gupta, A.; Singh, M.K. Feature Extraction for Hyperspectral Image Classification: A Review. *Int. J. Remote Sens.* **2020**, *41*, 6248–6287. [CrossRef]

35. Nesha, M.K.; Hussin, Y.A.; Van Leeuwen, L.M.; Sulistioadi, Y.B. Modeling and Mapping Aboveground Biomass of the Restored Mangroves Using ALOS-2 PALSAR-2 in East Kalimantan, Indonesia. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *91*, 102158. [CrossRef]

36. Bouvet, A.; Mermoz, S.; Le Toan, T.; Villard, L.; Mathieu, R.; Naidoo, L.; Asner, G.P. An Above-Ground Biomass Map of African Savannahs and Woodlands at 25 m Resolution Derived from ALOS PALSAR. *Remote Sens. Environ.* **2018**, *206*, 156–173. [CrossRef]

37. Englhart, S.; Keuck, V.; Siegert, F. Aboveground Biomass Retrieval in Tropical Forests—The Potential of Combined X- and L-Band SAR Data Use. *Remote Sens. Environ.* **2011**, *115*, 1260–1271. [CrossRef]

38. Loew, A.; Mauser, W. Generation of Geometrically and Radiometrically Terrain Corrected SAR Image Products. *Remote Sens. Environ.* **2007**, *106*, 337–349. [CrossRef]

39. Zhao, L.; Chen, E.; Li, Z.; Zhang, W.; Gu, X. Three-Step Semi-Empirical Radiometric Terrain Correction Approach for PolSAR Data Applied to Forested Areas. *Remote Sens.* **2017**, *9*, 269. [CrossRef]

40. Nie, Y.; Hu, Y.; Sa, R.; Fan, W. Inversion of Forest above Ground Biomass in Mountainous Region Based on PolSAR Data after Terrain Correction: A Case Study from Saihanba, China. *Remote Sens.* **2024**, *16*, 846. [CrossRef]

41. Zhang, H.; Zhu, J.; Wang, C.; Lin, H.; Long, J.; Zhao, L.; Fu, H.; Liu, Z. Forest Growing Stock Volume Estimation in Subtropical Mountain Areas Using PALSAR-2 L-Band PolSAR Data. *Forests* **2019**, *10*, 276. [CrossRef]

42. Shi, J.; Zhang, W.; Marino, A.; Zeng, P.; Ji, Y.; Zhao, H.; Huang, G.; Wang, M. Forest Total and Component Biomass Retrieval via GA-SVR Algorithm and Quad-Polarimetric SAR Data. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103275. [CrossRef]

43. Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests* **2021**, *12*, 216. [CrossRef]

44. Zhang, Y.; Liu, J.; Li, W.; Liang, S. A Proposed Ensemble Feature Selection Method for Estimating Forest Aboveground Biomass from Multiple Satellite Data. *Remote Sens.* **2023**, *15*, 1096. [CrossRef]

45. Karachristos, K.; Koukiou, G.; Anastassopoulos, V. A Review on PolSAR Decompositions for Feature Extraction. *J. Imaging* **2024**, *10*, 75. [CrossRef]

46. Shang, Z.; Lian, Z.; Li, M.; Han, K.; Zheng, H. Machine-Learning-Assisted Multi-Objective Optimization in Vertical Zone Refining of Ultra-High Purity Indium. *Sep. Purif. Technol.* **2023**, *305*, 122430. [CrossRef]

47. Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature Selection in Machine Learning: A New Perspective. *Neurocomputing* **2018**, *300*, 70–79. [CrossRef]

48. Xu, L.; Saatchi, S.S.; Yang, Y.; Myneni, R.B.; Frankenberg, C.; Chowdhury, D.; Bi, J. Satellite Observation of Tropical Forest Seasonality: Spatial Patterns of Carbon Exchange in Amazonia. *Environ. Res. Lett.* **2015**, *10*, 084005. [CrossRef]

49. Sheikhpour, R.; Sarram, M.A.; Gharaghani, S.; Chahooki, M.A.Z. A Survey on Semi-Supervised Feature Selection Methods. *Pattern Recognit.* **2017**, *64*, 141–158. [CrossRef]

50. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature Selection. *ACM Comput. Surv.* **2018**, *50*, 1–45. [CrossRef]

51. Chandrashekar, G.; Sahin, F. A Survey on Feature Selection Methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]

52. Islam, M.R.; Lima, A.A.; Das, S.C.; Mridha, M.F.; Prodeep, A.R.; Watanobe, Y. A Comprehensive Survey on the Process, Methods, Evaluation, and Challenges of Feature Selection. *IEEE Access* **2022**, *10*, 99595–99632. [CrossRef]

53. Zhang, W.; Zhao, L.; Li, Y.; Shi, J.; Yan, M.; Ji, Y. Forest Above-Ground Biomass Inversion Using Optical and SAR Images Based on a Multi-Step Feature Optimized Inversion Model. *Remote Sens.* **2022**, *14*, 1608. [CrossRef]

54. Hsu, H.-H.; Hsieh, C.-W.; Lu, M.-D. Hybrid Feature Selection by Combining Filters and Wrappers. *Expert Syst. Appl.* **2011**, *38*, 8144–8150. [CrossRef]

55. Li, X.; Ren, J. MICQ-IPSO: An Effective Two-Stage Hybrid Feature Selection Algorithm for High-Dimensional Data. *Neurocomputing* **2022**, *501*, 328–342. [CrossRef]

56. Vafaei, S.; Soosani, J.; Adeli, K.; Fadaei, H.; Naghavi, H.; Pham, T.; Tien Bui, D. Improving Accuracy Estimation of Forest Aboveground Biomass Based on Incorporation of ALOS-2 PALSAR-2 and Sentinel-2A Imagery and Machine Learning: A Case Study of the Hyrcanian Forest Area (Iran). *Remote Sens.* **2018**, *10*, 172. [CrossRef]

57. Fassnacht, F.E.; Hartig, F.; Latifi, H.; Berger, C.; Hernández, J.; Corvalán, P.; Koch, B. Importance of Sample Size, Data Type and Prediction Method for Remote Sensing-Based Estimations of Aboveground Forest Biomass. *Remote Sens. Environ.* **2014**, *154*, 102–114. [CrossRef]

58. Schlund, M.; Scipal, K.; Quegan, S. Assessment of a Power Law Relationship Between *P*-Band SAR Backscatter and Aboveground Biomass and Its Implications for BIOMASS Mission Performance. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2018**, *11*, 3538–3547. [CrossRef]

59. Sugimoto, R.; Kato, S.; Nakamura, R.; Tsutsumi, C.; Yamaguchi, Y. Deforestation Detection Using Scattering Power Decomposition and Optimal Averaging of Volume Scattering Power in Tropical Rainforest Regions. *Remote Sens. Environ.* **2022**, *275*, 113018. [CrossRef]

60. Fu, X.; Jiang, D.; Wang, L.; Gao, X.; Qian, D. Study on Biomass of Larixprincipis Rupprechtiiin Saihanba Mechanized Forestry Centre. *For. Ecol. Sci.* **2015**, *30*, 113–116. [CrossRef]

61. Jin, A. Study on the Biomass of the Main Trees at Bayingzhuang Forest Land in Hebei Province. Master's Thesis, Beijing Forestry University, Beijing, China, 2012.

62. Chu, C. Study on the Growth of Scotch Pine Plantations and the Variation of Soil Nutrients in Saihanba Aera. Master's Thesis, Hebei Agricultural University, Baoding, China, 2016.

63. Cao, J. Above- and Belowground Carbon Pools in Different Ages of Chinese Pine and Oriental Arborvitae Plantation Forests in Northern Mountain Areas of Beijing. Ph.D. Thesis, Beijing Forestry University, Beijing, China, 2011.

64. Yang, L. The Build Study of Acer Truncatum Bunge Biomass Calculation Model in Western Hills of Beijing. Master's Thesis, Beijing Forestry University, Beijing, China, 2012.

65. Lee, J.-S.; Schuler, D.L.; Ainsworth, T.L.; Krogager, E.; Kasilingam, D.; Boerner, W.-M. On the Estimation of Radar Polarization Orientation Shifts Induced by Terrain Slopes. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 30–41. [CrossRef]

66. Lee, J.-S.; Grunes, M.R.; Pottier, E.; Ferro-Famil, L. Unsupervised Terrain Classification Preserving Polarimetric Scattering Characteristics. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 722–731. [CrossRef]

67. Gonçalves, F.G.; Santos, J.R.; Treuhaft, R.N. Stem Volume of Tropical Forests from Polarimetric Radar. *Int. J. Remote Sens.* **2011**, *32*, 503–522. [CrossRef]

68. Mandal, D.; Kumar, V.; Ratha, D.; Dey, S.; Bhattacharya, A.; Lopez-Sanchez, J.M.; McNairn, H.; Rao, Y.S. Dual Polarimetric Radar Vegetation Index for Crop Growth Monitoring Using Sentinel-1 SAR Data. *Remote Sens. Environ.* **2020**, *247*, 111954. [CrossRef]

69. Pope, K.O.; Rey-Benayas, J.M.; Paris, J.F. Radar Remote Sensing of Forest and Wetland Ecosystems in the Central American Tropics. *Remote Sens. Environ.* **1994**, *48*, 205–219. [CrossRef]

70. Kim, Y.; Van Zyl, J.J. A Time-Series Approach to Estimate Soil Moisture Using Polarimetric Radar Data. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2519–2527. [CrossRef]

71. Mitchard, E.T.A.; Saatchi, S.S.; White, L.J.T.; Abernethy, K.A.; Jeffery, K.J.; Lewis, S.L.; Collins, M.; Lefsky, M.A.; Leal, M.E.; Woodhouse, I.H.; et al. Mapping Tropical Forest Biomass with Radar and Spaceborne LiDAR in Lopé National Park, Gabon: Overcoming Problems of High Biomass and Persistent Cloud. *Biogeosciences* **2012**, *9*, 179–191. [CrossRef]

72. Chhabra, A.; Rüdiger, C.; Yebra, M.; Jagdhuber, T.; Hilton, J. RADAR-Vegetation Structural Perpendicular Index (R-VSPI) for the Quantification of Wildfire Impact and Post-Fire Vegetation Recovery. *Remote Sens.* **2022**, *14*, 3132. [CrossRef]

73. Freeman, A. Fitting a Two-Component Scattering Model to Polarimetric SAR Data From Forests. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 2583–2592. [CrossRef]

74. Freeman, A.; Durden, S.L. A Three-Component Scattering Model for Polarimetric SAR Data. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 963–973. [CrossRef]

75. Cui, Y.; Yamaguchi, Y.; Yang, J.; Park, S.-E.; Kobayashi, H.; Singh, G. Three-Component Power Decomposition for Polarimetric SAR Data Based on Adaptive Volume Scatter Modeling. *Remote Sens.* **2012**, *4*, 1559–1572. [CrossRef]

76. Yamaguchi, Y.; Moriyama, T.; Ishido, M.; Yamada, H. Four-Component Scattering Model for Polarimetric SAR Image Decomposition. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1699–1706. [CrossRef]

77. Van Zyl, J.J.; Arii, M.; Kim, Y. Model-Based Decomposition of Polarimetric SAR Covariance Matrices Constrained for Nonnegative Eigenvalues. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3452–3459. [CrossRef]

78. Yang, J.; Peng, Y.-N.; Yamaguchi, Y.; Yamada, H. On Huynen's Decomposition of a Kennaugh Matrix. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 369–372. [CrossRef]

79. Dey, S.; Bhattacharya, A.; Ratha, D.; Mandal, D.; Frery, A.C. Target Characterization and Scattering Power Decomposition for Full and Compact Polarimetric SAR Data. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3981–3998. [CrossRef]

80. Dey, S.; Bhattacharya, A.; Frery, A.C.; Lopez-Martinez, C.; Rao, Y.S. A Model-Free Four Component Scattering Power Decomposition for Polarimetric SAR Data. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2021**, *14*, 3887–3902. [CrossRef]

81. Cloude, S.R. Target Decomposition Theorems in Radar Scattering. *Electron. Lett.* **1985**, *21*, 22–24. [CrossRef]

82. Krogager, E. New Decomposition of the Radar Target Scattering Matrix. *Electron. Lett.* **1990**, *26*, 1525. [CrossRef]

83. Cloude, S.R.; Pottier, E. An Entropy Based Classification Scheme for Land Applications of Polarimetric SAR. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 68–78. [CrossRef]

84. Demirci, S.; Kirik, O.; Ozdemir, C. Interpretation and Analysis of Target Scattering from Fully-Polarized ISAR Images Using Pauli Decomposition Scheme for Target Recognition. *IEEE Access* **2020**, *8*, 155926–155938. [CrossRef]

85. Liu, Z.; Michel, O.O.; Wu, G.; Mao, Y.; Hu, Y.; Fan, W. The Potential of Fully Polarized ALOS-2 Data for Estimating Forest Above-Ground Biomass. *Remote Sens.* **2022**, *14*, 669. [CrossRef]

86. Chang, Q.; Zwieback, S.; DeVries, B.; Berg, A. Application of L-Band SAR for Mapping Tundra Shrub Biomass, Leaf Area Index, and Rainfall Interception. *Remote Sens. Environ.* **2022**, *268*, 112747. [CrossRef]

87. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [CrossRef]

88. 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]

89. Park, S.-E. The Effect of Topography on Target Decomposition of Polarimetric SAR Data. *Remote Sens.* **2015**, *7*, 4997–5011. [CrossRef]

90. Lee, J.-S.; Ainsworth, T.L. The Effect of Orientation Angle Compensation on Coherency Matrix and Polarimetric Target Decompositions. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 53–64. [CrossRef]

91. Li, H.; Li, Q.; Wu, G.; Chen, J.; Liang, S. The Impacts of Building Orientation on Polarimetric Orientation Angle Estimation and Model-Based Decomposition for Multilook Polarimetric SAR Data in Urban Areas. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5520–5532. [CrossRef]

92. Chowdhury, T.; Thiel, C.; Schmullius, C.; Stelmaszczuk-Górska, M. Polarimetric Parameters for Growing Stock Volume Estimation Using ALOS PALSAR L-Band Data over Siberian Forests. *Remote Sens.* **2013**, *5*, 5725–5756. [CrossRef]

93. Soja, M.J.; Sandberg, G.; Ulander, L.M.H. Regression-Based Retrieval of Boreal Forest Biomass in Sloping Terrain Using P-Band SAR Backscatter Intensity Data. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2646–2665. [CrossRef]

94. Sandberg, G.; Ulander, L.M.H.; Wallerman, J.; Fransson, J.E.S. Measurements of Forest Biomass Change Using P-Band Synthetic Aperture Radar Backscatter. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6047–6061. [CrossRef]

95. De Almeida, C.T.; Galvão, L.S.; Aragão, L.E.d.O.C.e.; Ometto, J.P.H.B.; Jacon, A.D.; Pereira, F.R.d.S.; Sato, L.Y.; Lopes, A.P.; Graça, P.M.L.d.A.; Silva, C.V.d.J.; et al. Combining LiDAR and Hyperspectral Data for Aboveground Biomass Modeling in the Brazilian Amazon Using Different Regression Algorithms. *Remote Sens. Environ.* **2019**, *232*, 111323. [CrossRef]

96. Luckman, A. A Study of the Relationship between Radar Backscatter and Regenerating Tropical Forest Biomass for Spaceborne SAR Instruments. *Remote Sens. Environ.* **1997**, *60*, 1–13. [CrossRef]

97. Kc, Y.B.; Liu, Q.; Saud, P.; Xu, C.; Gaire, D.; Adhikari, H. Driving Factors and Spatial Distribution of Aboveground Biomass in the Managed Forest in the Terai Region of Nepal. *Forests* **2024**, *15*, 663. [CrossRef]

98. Narvaes, I.d.S.; dos Santos, J.R.; Bispo, P.d.C.; Graça, P.M.d.A.; Guimarães, U.S.; Gama, F.F. Estimating Forest Above-Ground Biomass in Central Amazonia Using Polarimetric Attributes of ALOS/PALSAR Images. *Forests* **2023**, *14*, 941. [CrossRef]

99. Waqar, M.; Sukmawati, R.; Ji, Y.; Sri Sumantyo, J. Tropical PeatLand Forest Biomass Estimation Using Polarimetric Parameters Extracted from RadarSAT-2 Images. *Land* **2020**, *9*, 193. [CrossRef]

100. Tian, L.; Wu, X.; Tao, Y.; Li, M.; Qian, C.; Liao, L.; Fu, W. Review of Remote Sensing-Based Methods for Forest Aboveground Biomass Estimation: Progress, Challenges, and Prospects. *Forests* **2023**, *14*, 1086. [CrossRef]

101. Li, Y.; Li, C.; Li, M.; Liu, Z. Influence of Variable Selection and Forest Type on Forest Aboveground Biomass Estimation Using Machine Learning Algorithms. *Forests* **2019**, *10*, 1073. [CrossRef]

102. Gao, Y.; Lu, D.; Li, G.; Wang, G.; Chen, Q.; Liu, L.; Li, D. Comparative Analysis of Modeling Algorithms for Forest Aboveground Biomass Estimation in a Subtropical Region. *Remote Sens.* **2018**, *10*, 627. [CrossRef]