New Metrics and the Combinations for Estimating Forest Biomass From GLAS Data

Yuzhen Zhang, Wenhao Li, and Shunlin Liang, *Fellow, IEEE*

***Abstract*—Geoscience laser altimeter system (GLAS) data have been widely used for forest aboveground biomass (AGB) estimation, but there is no consensus on the optimal metrics. To explore whether a few optimal GLAS metrics could generate accurate AGB estimates, we proposed five metrics and explored their combinations with ten existing ones. The importance of these metrics was measured according to their contributions to changes in the cross-validated mean-squared error. The two to eight most important metrics were then selected to develop AGB models, and their performances were evaluated using field AGB. The optimal combination of GLAS metrics was finally used for AGB estimation at GLAS footprints from 2004 to 2007 within a $2^\circ \times 2^\circ$ spatial extent in Tahe and Changbai Mountain, China. The results showed that four GLAS metrics, including our proposed metric CRH25 (25th percentile of canopy reflection heights) combined with Lead, quadratic mean canopy height, and H75, yield the best AGB estimates, with an R^2 of 0.61±0.15 and RMSE of 52.20±23.50 Mg/ha, and the inclusion of more GLAS metrics did not improve the results. The estimated forest AGB in Tahe was 89.03±19.16 Mg/ha and 103.07±23.42 Mg/ha in Changbai Mountain. In both regions, the annual average forest AGB estimates for 2005 were higher than the AGB estimates for 2004, 2006, and 2007. The results of this study suggested that a few waveform parameters could enable the accurate estimation of forest AGB. Moreover, this study indicated that GLAS data might be able to monitor forest AGB changes, which require further investigation.**

***Index Terms*—Forest biomass, geoscience laser altimeter system (GLAS) data, waveform parameters.**

I. Introduction

FOREST aboveground biomass (AGB) plays an important role in the global carbon cycle and climate change studies, but its magnitude, patterns, and uncertainties remain poorly quantified [1]–[3]. Over the past decades, the science community has paid much attention to forest AGB estimates from multiple remote sensing datasets, including optical images, synthetic aperture radar (SAR), and light detection and ranging (LiDAR) data, or a combination of them [4]–[11]. LiDAR data are capable of retrieving forest AGB, particularly in forests with high AGB values where optical remote sensing and SAR techniques suffer from signal saturation problems [12]–[15].

LiDAR remote sensing for forestry applications can be categorized into ground-based, airborne, and spaceborne LiDAR, according to the platform used [16]–[18]. Currently, ground-based and airborne LiDAR has been successfully applied for individual tree species classification [19], deriving leaf area index [20], [21], analyzing forest canopies [22], [23], and estimating forest structure and biomass [24]–[26] but is only feasible at local scales at effective costs. Spaceborne LiDAR provides the solution for large-area or even global forest AGB estimates [27]. The first spaceborne LiDAR for continuous global observations of the earth was the geoscience laser altimeter system (GLAS), carried on the ice, cloud, and land elevation satellite (ICESat) [28], [29]. Due to its availability on a global scale from 2003 to 2009, GLAS has been widely used to estimate tree heights and forest biomass across wide areas [5], [7], [30]–[32]. In 2018, the ICESat-2 satellite and global ecosystem dynamics investigation (GEDI) mission were launched. A few studies have demonstrated that AGB mapping accuracy could be improved through the fusion of ICESat-2 and GEDI [33], [34]. Until now, ICESat-2 and GEDI have covered a relatively shorter period compared with GLAS data.

Previous studies estimated forest AGB with GLAS data mainly by initially extracting waveform metrics and then using them as predictor variables [9], [35]. The canopy height, which is the distance between the signal beginning to ground return [36]; the maximum, mean, median, and quadratic mean canopy height (QMCH) proposed by Lefsky *et al*. [37]; the height of median energy (HOME) [38], [39]; the waveform extent, leading edge extent, and trailing edge extent [40]; the energy quantiles H25, H50, H75, and H100 from Sun *et al*. [35]; the canopy closure [36], [41]; the mean canopy transmittance (MCT) and 50th percentile of the canopy transmittance calculated from the canopy transmittance profile [42]; and the amplitude, heights, width, and area of Gaussian peaks in the GLA14 product [9] have been extensively used. However, there is no consensus among researchers regarding promising metrics for AGB prediction. For example, Hu *et al*. [43] used waveform extent, leading edge extent, and trailing edge extent, in combination with optical data, to produce a wall-to-wall global forest AGB map. Fayad *et al*. [44] extracted the waveform extent, Gaussian peaks from the decomposition of each GLAS waveform, canopy height percentiles (10 through 90%), leading edge, and trailing edge and selected the best variables with stepwise regression based on the Bayesian information criterion to build the AGB estimation

Manuscript received May 23, 2021; revised July 6, 2021; accepted July 27, 2021. Date of publication August 2, 2021; date of current version August 18, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 41801347, in part by the National Key Research and Development Program of China under Grant 2016YFA0600103, and in part by the Fundamental Research Funds for the Central Universities under Grant FRF-TP-19-041A2. *(Corresponding author: Yuzhen Zhang.)*

Yuzhen Zhang and Wenhao Li are with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: yzhang@ustb.edu.cn; lwh2019@xs.ustb.edu.cn).

Shunlin Liang is with the Department of Geographical Sciences, University of Maryland, College Park, MD 20742 USA (e-mail: sliang@umd.edu).

Digital Object Identifier 10.1109/JSTARS.2021.3101285

model in French Guiana. Chi *et al.* [45] calculated the mean and median canopy heights, waveform length, top tree height, quantile heights, decile heights, leading edge extent, training edge extent, terrain index, and vegetation to surface ratio from waveform data for estimating forest AGB at GLAS footprints. The main limitations of these published studies are evident. The role of different GLAS metrics in deriving forest AGB has not been carefully examined, hindering the usage of GLAS metrics across the different forests or regions. Moreover, a large number of GLAS metrics were included in AGB modeling, which could lead to overfitting and, thus, poor accuracy of AGB predictions, particularly when a small number of field samples were used.

Therefore, we aim for the following.

1) Explore whether fewer waveform metrics could generate the accurate AGB predictions.
2) Find a few promising GLAS metrics for forest AGB estimates.
3) Assess the impacts of GLAS metrics on biomass estimates.

In this article, we proposed some new metrics extracted from GLAS waveform data and evaluated the AGB models developed with a combination of GLAS metrics to explore whether fewer metrics and which combination of metrics yield the best biomass estimates. In addition, we examined the effects of GLAS metrics on the accuracy of AGB estimates compared with the training samples, which had become an important factor affecting the accuracy of model prediction due to limited field biomass data [46]. Finally, a model developed with an optimal combination of GLAS metrics was used to estimate forest AGB at GLAS footprints within a 2°×2° spatial extent around the field measurements.

II. Data and Methods

A. Field AGB Data

Field measurements were used to build biomass models together with GLAS metrics. They were geographically located at GLAS footprints in the Tahe and Changbai Mountain forest regions of Northeast China (see Fig. 1), where the main forest types are conifer mixed forests, conifer forests, and broadleaf mixed forests [9]. A total of 86 GLAS footprints were measured in 2006 and 2007. For each footprint, four sampling plots with a radius of 7.5 m were located using differential GPS. Three sampling plots were distributed along the radial direction of the center sampling plots at 120° angles and 22.5 m away from the center sampling plot. For all trees with diameters at breast height (DBH) larger than 5 cm in the four sampling plots, the DBH and tree height were measured, and the tree biomass was computed using the allometric equation [47], [48]. Plot-level AGB was obtained by summing all the tree biomasses in the four sampling plots and then dividing by the total area of these plots.

B. ICESat-1 GLAS Data

The National Snow and Ice Data Centre provided 15 standard GLAS data products.[1] We used the L1A global altimetry data

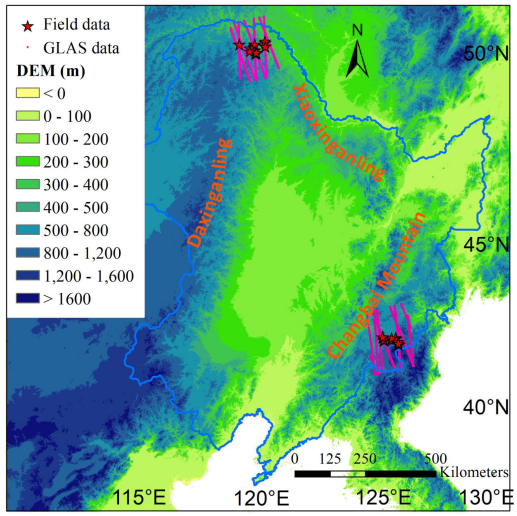

Fig. 1. Location of the study area. Red stars represent the field measurements in Tahe County and the Changbai Mountain region, and the pink points represent the GLAS footprints within a 2° × 2° spatial extent surrounding the field measurements in Tahe and Changbai Mountain. The background digital elevation data were provided by the NASA shuttle radar topographic mission.

product (GLAH01) and L2 global land surface altimetry data (GLAH14). The GLAH01 data included the transmitted and received waveforms from the altimeter and were the source data for extracting key waveform parameters closely correlated with forest AGB. The GLAH14 data provide the geolocation of GLAS footprints and are linked with GLAH01 by the record index and shot number [35]. The GLAS shots corresponding to field measurements and within a 2°×2° spatial extent surrounding the field measurements in Tahe County and Changbai Mountain are shown in Fig. 1. All GLAS shots from 2004 to 2007 were selected, and their waveforms were processed as follows.

We first removed the background noise contained in the raw data of full waveforms according to the noise threshold that was determined using the following:

$$\text{Threshold} = \mu + n \times \sigma \tag{1}$$

where the mean background noise level (μ) and standard deviation (σ) are provided in GLAH01, and n was 4.5. The signal start and end locations were determined where three consecutive bins were higher than the noise threshold [40], [49], [50].

Gaussian filtering was carried out for waveform denoising to identify the initial amplitude, width, and sigma of the Gaussian decomposition components [51], [52]. We then implemented the Gaussian decomposition method to fit the raw waveform from the signal start and end locations using the nonlinear Levenberg–Marquardt algorithm [52]–[55]. By searching forward from the signal endpoint, the center of a Gaussian component with higher amplitude between the last two components was considered the ground position. We determine the ground position using the last two Gaussian components instead of the last Gaussian component, which could reduce the influences of complex terrain [56], [57].

The GLAS shots were excluded when the signal start location was more than 60 m above the ground position, suggesting

[1][Online]. Available: https://nsidc.org/data/icesat/data.html, accessed on 28 July 2020

TABLE I
GLAS METRICS DERIVED FROM WAVEFORM DATA

GLAS metrics	Definition	Reference
H25, H50, H75, H100	Quantile heights, calculated by subtracting the ground elevation from the elevation at which 25%, 50%, 75%, and 100% of the accumulated energy starting from signal beginning to signal end occur	[35]
Lead, Trail	Leading edge extent and trailing edge extent calculated from the waveform	[40]
MeanH, MedH, QMCH	Mean canopy height, median canopy height, and quadratic mean canopy height calculated from the canopy height profile	[37]
MCT	Mean canopy transmittance, calculated from the canopy transmittance profile	[42]
CRH25, CRH50, CRH75	Quantile canopy reflection heights calculated by subtracting the lowest canopy elevation from the elevation at which 25%, 50%, and 75% of the accumulated canopy energy occur	This study
MCR	Mean canopy reflection height calculated from canopy reflection profiles	This study
QMCR	Quadratic mean canopy reflection height	This study

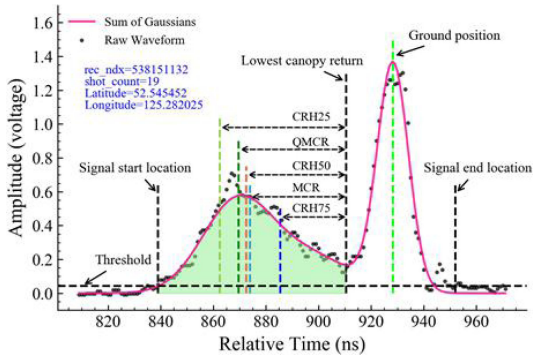

Fig. 2. GLAS waveform in the study area. The light gray dots represent the raw waveform; the pink curve indicates the sum of Gaussian decomposed components; and the green filled area represents the canopy returns. The new proposed waveform parameters, CRH25, QMCR, CRH50, MCR, and CRH75, represent, respectively, the location where 25% of the accumulated canopy energy occur, QMCR height, where 50% of the accumulated canopy energy occur, MCR height, and where 75% of the accumulated canopy energy occur, respectively, and are shown from left to right.

possible cloud cover or noisy return [58]. Noisy waveforms, where the maximum return value was lower than twice the mean background noise level or less than 20 times the background noise standard deviation, were also eliminated. To ensure that the GLAS shots were more likely reflected from the forest canopy, we assumed that the waveforms characterized by only one Gaussian component were due to bare ground reflection [59], [60] and excluded them.

C. Extracting GLAS Metrics for Estimating Forest AGB

Ten widely used waveform parameters, including four quantile heights (H25, H50, H75, and H100), leading edge and trailing edge extents, mean canopy height (MeanH), median canopy height (MedH), QMCH, and MCT, were extracted from GLAH01 waveform data. Methods to obtain these metrics can be found in the literature, as given in Table I. In addition, we propose five metrics as a supplement for AGB estimation. The five proposed metrics in this study are all based on the canopy reflectance profile (CRP), which is defined as the ratio of energy from canopy returns to the sum of canopy energy and calculated using the following:

$$\mathrm{CRP}(h) = \frac{\mathrm{CR}(h)}{\sum_{h=0}^{\max \mathrm{TH}} \mathrm{CR}(h)} \tag{2}$$

where CR is the energy of the waveform from canopy returns, h is the distance between the canopy returns and the lowest canopy return, and $\sum_{h=0}^{\max \mathrm{TH}} \mathrm{CR}(h)$ is the sum of canopy reflection returns, referred to as the canopy reflection sum [36]. $h = 0$ and maxTH in (2) represent the lowest canopy return and the distance between the canopy top and the lowest canopy return, respectively.

To define the CRP, it is essential to separate the ground returns and canopy returns of a waveform. The portion of the waveform below the detected ground peak was mirrored above it, which was the ground return [42]. The canopy returns were obtained by subtracting the ground return from the filtered waveform and are shown in Fig. 2 with green-filled areas. Based on the CRP, the proposed metrics CRH25, CRH50, and CRH75 were derived by subtracting the lowest canopy elevation from the elevation at which canopy energy accumulation reached 25, 50, and 75% of the canopy reflection sum that occurred from the beginning of the signal. The mean canopy reflection (MCR) height and the quadratic mean canopy reflection (QMCR) height were calculated from the CRP as follows:

$$\mathrm{MCR} = \sum_{h=0}^{\max \mathrm{TH}} \mathrm{CRP}(h) \times h \tag{3}$$

$$\mathrm{QMCR} = \sqrt{\sum_{h=0}^{\max \mathrm{TH}} \mathrm{CRP}(h) \times h^2}. \tag{4}$$

D. Modeling Forest AGB With GLAS Metrics

Since published studies have suggested that the random forest (RF) and support vector regression (SVR) algorithms exhibit good performances in estimating forest AGB with GLAS data [9], [61], we modeled forest AGB with GLAS metrics using the RF model, the SVR with the linear kernel (SVR-linear) algorithm, and SVR with the radial basis function kernel (SVR-RBF) algorithm in this study. For each algorithm, three AGB models were developed based on the different combinations of GLAS metrics. The first model was built with 10 GLAS metrics defined in previous studies, including H25, H50, H75, H100, Lead, Trail, MeanH, MedH, QMCH, and MCT (Metrics I), the second model was built with 5 proposed GLAS metrics, including CRH25, CRH50, CRH75, MCR, and QMCR (Metrics II), and the last model was built with all 15 metrics, including both Metrics I and Metrics II (Metrics III). A total of nine AGB models were developed, including RF, SVR-linear, and SVR-RBF trained with Metrics I, Metrics II, and Metrics III, respectively.

All data were randomly split into training data and test data. The training data were used to train these models, and the test data were used to evaluate their performances. To explore whether random splitting or the number of test data affects the

performance diagnostics of AGB models, we split the whole datasets using the following four strategies:

1) 90% for training and 10% for testing;
2) 80% for training and 20% for testing;
3) fivefold cross validation;
4) tenfold cross validation.

For the first and second scenarios, 100 runs were implemented to randomly split the data into training data and test data. The procedures of fivefold cross validation and tenfold cross validation were repeated 20 and 10 times, respectively, to ensure that all four strategies had 100 evaluation results.

Moreover, to investigate whether only a few GLAS metrics could enable the accurate estimation of forest AGB, AGB models were developed with different numbers of promising predictors ranging from 2 to 8 using the optimal AGB modeling algorithm among RF, SVR-linear, and SVR-RBF, and their performances were assessed in terms of R^2 and root-mean-square error (RMSE). The GLAS metrics were selected according to their importance. We first built an AGB model with all 15 predictors, ranked their importance, applied the backward elimination method, and progressively eliminated the least promising variable from the GLAS metrics [62]. The above steps were repeated until only one predictor remained as input for the AGB model. The importance of a variable was measured by the change of the mean-squared error (MSE) when the variable was excluded for AGB modeling, and the removal of a more important variable corresponded to a larger increase in MSE. We ranked the GLAS metrics according to the changes in MSE and progressively eliminated the metric with the smallest MSE. The linear Borda count was used to score a variable, and a higher score corresponded to a more important variable.

In agreement with the AGB modeling with different combinations of waveform parameters, we used fivefold cross validation for splitting the training and test datasets and repeated 20 times, generating 100 rankings of GLAS metrics according to their importance measured by changes in MSE. For a visual display of variable importance, the 100 rankings of GLAS metrics were linearly aggregated through weighted voting [63].

Finally, to fully address the performances of five proposed metrics CRH25, CRH50, CRH75, MCR, and QMCR (Metrics II) for AGB estimates, comparisons with their similar metrics (H25, H50, H75, MeanH, and QMCH, Metrics IV) were carried out. For each of the newly proposed metrics (Metrics II) and their similar metrics (Metrics IV), the relationship with field AGB was explored, respectively. Furthermore, we evaluated the performances of Metrics II, Metrics IV, five importance metrics among the ten metrics (CRH25, CRH50, CRH75, MCR, QMCR, H25, H50, H75, MeanH, and QMCH), and the optimal combination of GLAS parameters from Metrics III in estimating forest AGB, using the optimal AGB modeling algorithm among RF, SVR-linear, and SVR-RBF. Consistent with the exploration of optimal combinations of waveform parameters, fivefold cross validation was used to split the whole datasets and 20 times were repeated.

E. Estimating Forest AGB at GLAS Footprints

Based on the optimal AGB modeling algorithms and optimal combination of GLAS metrics, we built 100 AGB models with bootstrap samples of field AGB and estimated forest AGB at GLAS footprints from 2004 to 2007 within a $2°\times2°$ spatial extent in Tahe and Changbai Mountain. Forest AGB estimate results from 100 bootstrap samples were averaged, and the coefficient of variation (CV) of AGB estimates was calculated to quantify the impacts of training samples on estimated AGB.

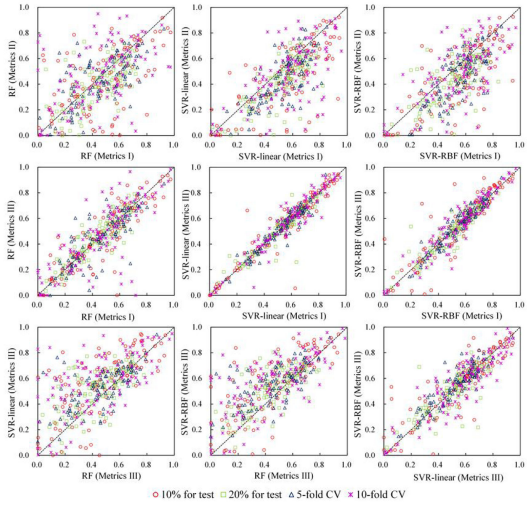

Fig. 3. Performances of nine forest AGB models in terms of R^2. Metrics I, Metrics II, and Metrics III represent that the AGB models were developed with the GLAS metrics in previous studies (H25, H50, H75, H100, Lead, Trail, MeanH, MedH, QMCH, and MCT), GLAS metrics proposed in this study (CRH25, CRH50, CRH75, MCR, QMCR), and all the metrics included in Metrics I and Metrics II, respectively. The 10% for the test and 20% for the test indicate that the whole dataset was randomly split into training data (90%) and test data (10%) and training data (80%) and test data (20%), and the test data were used to evaluate the estimated AGB. Fivefold CV and tenfold CV represent that the whole data were split, and the results were evaluated by fivefold cross validation (repeating 20 times) and tenfold cross validation (repeating 10 times), respectively.

III. Results

A. Performances of AGB Models

The performances of the nine AGB models for 100 runs in terms of R^2 and RMSE are shown in Figs. 3 and 4, respectively. The comparison results show large discrepancies in the accuracies of estimated forest AGB between the RF model with Metrics I and that with Metrics II, which also holds true for the SVR-linear and SVR-RBF algorithms (see Fig. 3). However, when Metrics II was added to Metrics I, no substantial changes in the accuracy of estimated AGB by RF, SVR-linear, and SVR-RBF algorithms are found for most of the 100 points, as shown in Figs. 3 and 4, indicating that it might be essential to perform feature selection or select optimal variables when GLAS parameters are used for predicting forest AGB.

Compared with the SVR-linear and SVR-RBF, RF models developed with Metrics I and those with Metrics III are more different, particularly when tenfold cross validation was used to evaluate the models. This could be partly because RF models are more sensitive to the training samples when the sample sizes are small, while the SVR algorithms are excellent for prediction in the cases of small sample sizes [9].

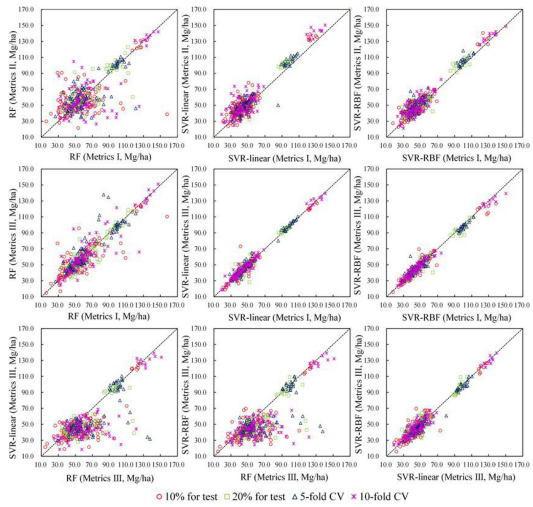

Fig. 4. Performances of nine forest AGB models in terms of RMSE (in units of Mg/ha). Metrics I, Metrics II, and Metrics III represent that the AGB models were developed with the GLAS metrics in previous studies (H25, H50, H75, H100, Lead, Trail, MeanH, MedH, QMCH, and MCT), GLAS metrics proposed in this study (CRH25, CRH50, CRH75, MCR, and QMCR), and all the metrics included in Metrics I and Metrics II, respectively. The 10% percent for the test and 20% for the test indicate that the whole dataset was randomly split into training data (90%) and test data (10%) and training data (80%) and test data (20%), and the test data were used to evaluate the estimated AGB. Fivefold CV and tenfold CV represented that the whole data were split, and the results were evaluated by fivefold cross validation (repeating 20 times) and tenfold cross validation (repeating 10 times), respectively.

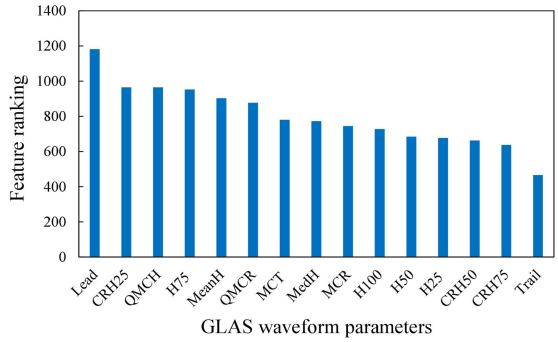

Fig. 5. Importance ranking of GLAS waveform parameters for AGB estimates.

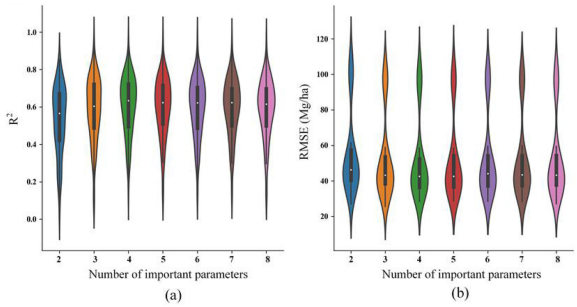

Fig. 6. Impacts of the number of GLAS key metrics used on the accuracy of estimated forest AGB.

In addition, the performances of nine AGB models for 100 runs were quite diverse, with R^2 ranging from approximately zero to 0.99 provided by the SVR-RBF algorithm with Metrics III and RMSE ranging from 15.90 to 157.83 Mg/ha provided by the RF algorithm with Metrics I (see Figs. 3 and 4), which indicates the large impacts of the training datasets on the estimated results. This phenomenon is more evident in terms of RMSE, showing that the 100 RMSE values are generally grouped into two separate parts for each of the nine AGB models due to the use of different datasets for training and testing of the AGB model.

The validation methods also affected the performances of the AGB models. Fig. 4 shows that the results of AGB estimates are more diverse with the changes in training datasets and predictor variables when 10% of the whole datasets are used for testing, either through random splitting or through tenfold cross validation. In contrast, the results are more stable when 20% of the data, in particular fivefold cross validation, are selected for evaluation of the estimated AGB.

SVR-linear and SVR-RBF significantly outperform RF and provide more accurate AGB estimates (see Figs. 3 and 4). The SVR-RBF models perform better than the SVR-linear model when fivefold cross validation is used to evaluate the accuracy of the AGB estimates. Therefore, the following analysis results are based on the SVR-RBF algorithm and evaluated by fivefold cross validation.

B. Importance of GLAS Metrics in Forest AGB Estimation

Since more variables could not lead to improved AGB estimates, we selected key GLAS parameters for retrieving forest AGB using the feature selection method, as mentioned in Section II-D. The results showed that Lead had the largest contributions to the estimation of forest AGB, followed by CRH25, QMCH, H75, MeanH, QMCR, MCT, and MedH (see Fig. 5).

Based on the SVR-RBF algorithm, we explored the optimal number of GLAS metrics that could provide the accurate AGB estimates. The fivefold cross-validated R^2 and RMSE of the SVR-RBF models developed with the two to eight most important variables for AGB estimates are shown in Fig. 6. The results suggest that the SVR-RBF model with four variables, including Lead, CRH25, QMCH, and H75, provides AGB estimates that are more accurate, with the highest cross-validated R^2 of 0.61 ± 0.15 and the lowest cross-validated RMSE of 52.20 ± 23.50 Mg/ha. SVR-RBF built with the four most important predictors significantly outperforms the SVR-RBF built with all the GLAS metrics, as shown in Fig. 7(a) and (b).

C. Performances of Newly Proposed Metrics and Previous Metrics for Estimating Forest AGB

The relationships of five newly proposed metrics and their similar metrics with field AGB were shown in Fig. 8. The results showed that in comparison with previous metrics H25, MeanH, and QMCH, the proposed metrics CRH25, MCR, and QMCR had better relationships with field AGB, while H50 and CRH50 had comparable correlations with forest AGB. However, H75 had a closer relationship with AGB than CRH75.

Although newly proposed Metrics II generally had better relationships with field AGB than Metrics IV, Metrics IV outperformed Metrics II and provided more accurate estimation

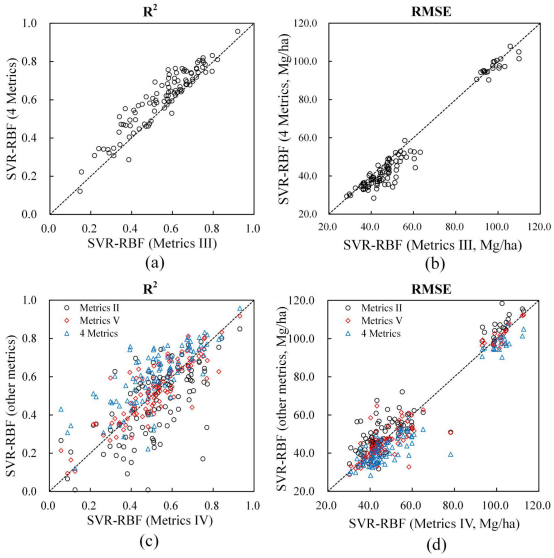

Fig. 7. Comparison of the performance of different combinations of GLAS metrics for estimating forest AGB using the SVR-RBF model. Metrics II, Metrics III, Metrics IV, Metrics V, and four metrics correspond to the performances of AGB models developed with GLAS metrics proposed in this study (CRH25, CRH50, CRH75, MCR, and QMCR), all the 15 GLAS parameters listed (including both Metrics I and Metrics II), five previous metrics (H25, H50, H75, MeanH, and QMCH), five important metrics among Metrics II and Metrics IV (CRH25, QMCH, H75, MeanH, and QMCR), and four most important metrics among Metrics III (Lead, CRH25, QMCH, and H75), respectively.

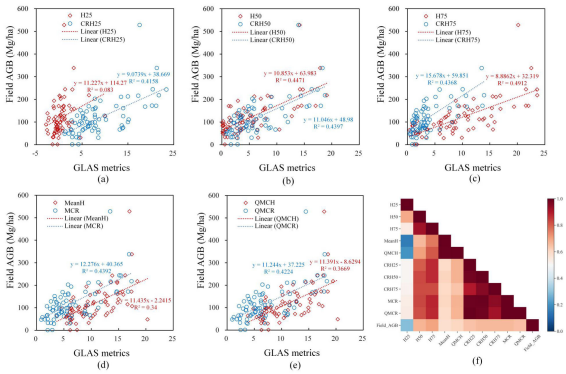

Fig. 8. Relationships of newly proposed metrics (CRH25, CRH50, CRH75, MCR, and QMCR) and their similar metrics (H25, H50, H75, MeanH, and QMCH) with field AGB.

of forest AGB (see Fig. 8). This could be attributed to the strong correlations among Metrics II [(see Fig. 8(f)]. When we selected five most important metrics among Metrics II and Metrics IV (see Fig. 5), which were CRH 25, QMCH, H75, MeanH, and QMCR (Metrics V), the accuracy of estimated AGB was improved in comparison with that obtained using Metrics IV. Four important metrics Lead, CRH25, QMCH, and H75 led to further improved estimation of forest AGB (see Fig. 8). Averaging the results for 20 runs obtained using Lead, CRH25, QMCH, and H75, we found that the estimated AGB was closer to field AGB than those obtained using Metrics IV (see Fig. 9). These results suggested that the newly proposed metrics could make an effective complement to published GLAS metrics for estimating forest AGB.

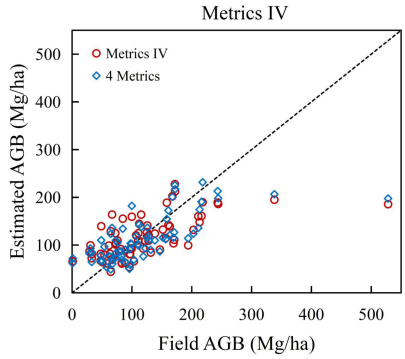

Fig. 9. Comparison of estimated AGB using Metrics IV (H25, H50, H75, MeanH, and QMCH) and four important metrics (Lead, CRH25, QMCH, and H75) with field AGB.

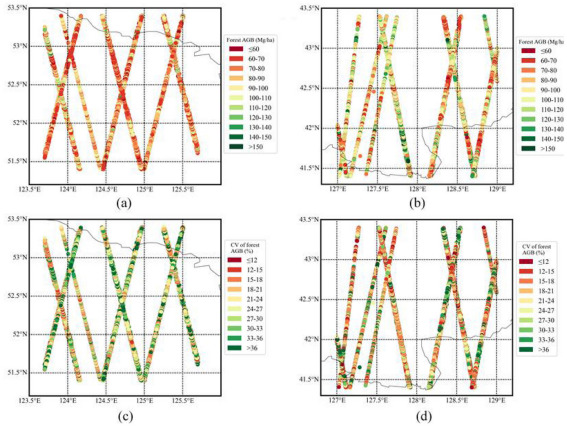

Fig. 10. Forest AGB estimates at GLAS footprints in the study area. (a) and (b) represent the mean AGB estimates in the Tahe and Changbai Mountain regions, respectively. (c) and (d) represent the CV of forest AGB in the Tahe and Changbai Mountain regions, respectively.

D. Estimated Forest AGB at GLAS Footprints

At the GLAS footprints from 2004 to 2007 in the Tahe and Changbai Mountain regions, forest AGB was estimated with Lead, CRH25, QMCH, and H75 using the SVR-RBF algorithm. The results of the mean value and CV of AGB estimates from 100 AGB models built with bootstrap samples are shown in Fig. 10.

The forest AGB in Tahe County is generally lower than that in Changbai Mountain [(see Fig. 10(a) and (b)]. According to the previous studies, the lower AGB in Tahe is due to the severe fire disturbance that occurred in 1987 [64]. In Tahe County, 53.80% of the GLAS footprints have a forest AGB of no larger than 90 Mg/ha, while only at 6.12% of the GLAS footprints is the estimated forest AGB more than 120 Mg/ha [see Fig. 10(a)]. Forest AGB estimates in the Changbai Mountain region vary in a wider range, with more than half of the GLAS footprints having AGB values ranging from 90 to 120 Mg/ha. In addition, at 20.00% of the GLAS footprints, the estimated forest AGB is larger than 120 Mg/ha. Averaging the estimated AGB at all GLAS footprints, we found that the forest AGB was 89.03 $\pm$ 19.16 Mg/ha within a $2^\circ \times 2^\circ$ spatial extent in Tahe and 103.07 $\pm$ 23.42 Mg/ha in Changbai Mountain. The estimated results are in agreement with published studies that used GLAS data to

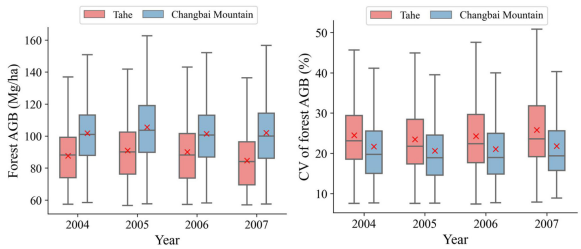

Fig. 11. Annual forest AGB estimates and the associated CV from 2004 to 2007 in the Tahe and Changbai mountain regions.

estimate AGB [9] and slightly higher than the AGB estimates with optical data [65].

The CV results show that the training samples greatly affect the estimated forest AGB, particularly in the Tahe region [see Fig. 10(c) and (d)]. The CV values are larger than 30% at 23.59% of GLAS footprints in Tahe and 14.88% in Changbai Mountain. Only 11.59% of footprints in Tahe and 25.67% of footprints in Changbai Mountain have a CV of no more than 15%. These results highlight the importance of training samples on the AGB models with GLAS data, especially in forests with lower AGB values.

We separated the annual GLAS footprints from 2004 to 2007 and found that most forests in the Tahe and Changbai Mountain regions had forest AGB values of approximately 85 Mg/ha and 100 Mg/ha, respectively, and the annual forest AGB estimates in Tahe had a more concentrated distribution (see Fig. 11). The annual mean forest AGB ranges from 84.68 Mg/ha in 2007 to 91.01 Mg/ha in 2005 in Tahe, and the corresponding CV ranges from 23.51% in 2005 to 25.82% in 2007. In the Changbai Mountain region, the annual averages of forest AGB are between 101.46 Mg/ha in 2006 to 105.59 Mg/ha in 2005, with CV values ranging from 20.64% in 2005 to 21.80% in 2007. Forest AGB estimates for 2005 are higher than forest AGB estimates at GLAS footprints for 2004, 2006, and 2007, with an average AGB of 91.01$\pm$19.78 Mg/ha in Tahe and 105.59$\pm$24.91 Mg/ha in Changbai Mountain. Meanwhile, the uncertainty of AGB estimates associated with the resampling of training data was the lowest in 2005.

IV. Discussion

A. Performances of AGB Modeling Algorithms

The RF, SVR-linear, and SVR-RBF algorithms were selected to predict forest AGB with GLAS data in this study. RF is a tree-based ensemble algorithm and is regarded as one of the best machine learning algorithms to estimate forest AGB due to its high predictive accuracy and high computational speed [66], [67]. However, RF tends to overfit noisy regression problems [68]. SVR, due to its excellent performance even with limited training samples, is also widely used in remote sensing fields [69]. In most studies, SVR is mainly referred to as SVR-RBF [67], [70]. The RBF kernel was chosen for the SVR algorithm because it has been shown to be effective for forest parameter retrieval [24]. Until now, few studies have focused on the SVR model with a linear kernel, although the linear SVR gave competitive accuracy for some applications and had a much faster speed of training and testing.

In this study, we explored the performances of RF, SVR-linear, and SVR-RBF in estimating forest AGB with GLAS data. Our results suggested that SVR-RBF and SVR-linear outperformed RF for forest AGB estimates with GLAS data, and SVR-RBF provided slightly better accuracy of estimates than the SVR-linear models. Some published results have also suggested that SVR-RBF could produce more accurate estimates of forest AGB than other algorithms, including RF [24]. However, some studies have indicated that the SVR-RBF model is inferior to the RF model. With GLAS data as well as Thematic Mapper data, Liu *et al*. [61] found that the RF model had better estimates of forest AGB than the SVR-RBF and stepwise regression. Wu *et al*. [70] compared five regression approaches, including stepwise regression, K-nearest neighbor, SVR, RF, and stochastic gradient boosting, and found that the RF algorithm provided the best estimates of AGB with Landsat data. The inconsistent results on the accuracy of AGB estimates using different algorithms could be due to the predictors (e.g., LiDAR metrics, derived from optical data, or a combination of multiple sources), sample sizes, and forest conditions.

B. Importance of GLAS Metrics for AGB Estimates

Since there is no consensus among studies regarding which GLAS metric was suitable for forest AGB estimates, researchers generally extract a large number of variables from GLAS data in combination with metrics from other remote sensing data, and then select some predictors for modeling forest AGB mainly based on stepwise regression. Fayad *et al*. [44] suggested that GLAS metrics, including the waveform extent, the leading edge, and canopy height percentiles h20 and h80 and the top of canopy heights, were the best variables to estimate forest AGB in French Guiana, which were correlated with field AGB estimation with an R^2 of 0.54 and RMSE of 48.3 Mg/ha. Liu *et al*. [61] selected eight GLAS metrics (quantile height H25, HOME, QMCH, top tree height with correction Treeht2, area under the waveform from vegetation, intensity of the first waveform from Gaussian decomposition, LEE, and TEE) and five metrics extracted from Landsat 5 data for AGB modeling and achieved a prediction accuracy with an R^2 of 0.76 and RMSE of 39.60 Mg/ha. Guo *et al*. [71] implemented stepwise regression to select GLAS parameters from a set of metrics, including the MeanH and MedH calculated from the waveform, top tree height from GLA14 products, top tree heights with various slope corrections, quantile heights, ratio between energy from vegetation and land surface, and locations of the six Gaussian peaks in GLA14 for forest AGB estimation, and generated optimal AGB prediction models based on seven GLAS metrics. The R^2 and RMSE of the AGB model for the conifer forest were 0.68 and 15.93 Mg/ha, respectively. For broadleaf forests, the R^2 was 0.71, and the RMSE was 17.02 Mg/ha. In this study, we ranked the importance of GLAS metrics for AGB estimates and selected key GLAS variables to predict forest AGB. The results showed that the four most important waveform parameters, Lead, CRH25, QMCH, and H75, generated the best predictive accuracy, with an R^2 of

0.61±0.15 and RMSE of 52.20±23.50 Mg/ha, and the inclusion of more GLAS variables did not result in improved estimates. Our results indicated that a few GLAS metrics might provide sufficient information for AGB estimates, which highlighted the importance of finding optimal GLAS parameters for forest AGB estimates in future studies and could facilitate the use of GLAS data for estimating forest AGB across large regions or the globe in a simple and efficient way.

It was observed that forest AGB was underestimated at high AGB values, which could be partly attributed to the limited number of samples at high AGB values for AGB modeling. More samples should be used for AGB modeling with GLAS metrics in future studies.

We did not find that the HOME metric or the 50% energy quantiles (H50) was extremely important for AGB estimates [35], [38], [39]. In our study, the proposed GLAS metric CRH25 was demonstrated to be a key predictor of forest AGB. Compared with previous height metrics (e.g., quantile heights), CRH25 and other GLAS metrics proposed in this study were calculated from the canopy return and, thus, might be less sensitive to the slope. To further confirm that CRH25 is an important predictor of forest AGB, more extensive field data need to be obtained, and the analysis should be carried out in other regions.

With GLAS data, we only achieved an accuracy with an R^2 of 0.61 and an RMSE value of 52.20 Mg/ha, lower than the published studies that included LiDAR data as well as optical imagery or SAR data [44], [61]. Optical and radar remote sensing data could be considered to improve the accuracy of the AGB estimates at GLAS footprints. When other types of remote sensing data were included in forest AGB estimates, the optimal number and combination of predictor variables to estimate forest AGB need to be further investigated.

C. Impacts of Sample Sizes on AGB Estimates

Due to the difficulties in collecting plot-level AGB data, previous studies typically included a few tens of field measurements for estimating forest AGB [46]. The sample size is often small relative to the predictor variables we could retrieve from remote sensing data, which had notable effects on the accuracy assessments [72]. In this study, a total of 86 field measurements were used for the calibration of remote sensing data and evaluation of estimated forest AGB. The results showed considerable variances of AGB estimates associated with random splitting of training and test samples, consistent with previous studies that found that when a small number of sample units were applied, subsampling in the algorithm had great effects on estimates [46], [73]. The use of more reference data may improve the results [44]. However, it might not be a superior way to increase the accuracy of AGB estimates since the prediction algorithm or data type used could have more effects than the sample size [46]. Moreover, the accuracy of forest AGB estimates may be subject to the size of inventory plots. Previous studies have proven that larger field plots could improve the performance of forest AGB estimates since they could reduce the likelihood of coregistration error between GLAS data and field plots and increase the chances of capturing the trees located at the boundary [11], [74], [75].

D. Potential of GLAS Data for Monitoring Forest AGB Dynamics

Recent studies have paid more attention to the capability of LiDAR data in monitoring forest AGB changes but have mainly focused on airborne LiDAR [76], [77]. Due to space and time limitations, LiDAR data alone could not be applied for the large-scale estimation of forest AGB dynamics [78].

For GLAS data, the footprints at different periods did not overlap in space, limiting the direct use of GLAS data from different years for estimating changes in forest AGB. Therefore, optical data or SAR data have been included in mapping forest AGB changes [79]. With the release of ICESat-2 and GEDI data, LiDAR data with a high laser repetition rate could be obtained. The combination of ICESat-1, ICESat-2, and GEDI data might provide promising results in studying monitoring forest AGB changes.

V. Conclusion

We have proposed five GLAS metrics and combined them with ten existing waveform parameters to estimate forest AGB. The results suggested that the proposed metric CRH25 was an important predictor of forest AGB, and the use of four metrics, Lead, CRH25, QMCH, and H75, provided the best prediction accuracy of AGB estimates, with an R^2 of 0.61±0.15 and RMSE of 52.20±23.50 Mg/ha. The inclusion of more GLAS metrics did not lead to improved AGB estimates, suggesting that a few waveform parameters could enable the accurate prediction of forest AGB, which also provided a simple and efficient way to estimate forest AGB with waveform LiDAR across large areas or globally. In addition, training samples tend to greatly affect the estimated results, even more than the GLAS metrics used, highlighting the importance of training samples in AGB estimates.

Based on the four most important GLAS metrics, forest AGB at GLAS footprints within a 2°×2° spatial extent in the Tahe and Changbai Mountain regions was predicted using the SVR-RBF algorithm. The results showed that the average forest AGB in Tahe was 89.03±19.16 Mg/ha from 2004 to 2007 and 103.07±23.42 Mg/ha in Changbai Mountain. In both regions, forest AGB estimates in 2005 were higher than those in other years. GLAS data alone or together with ICESat-2 or GEDI may have the potential to monitor forest AGB changes, which should be further investigated in future studies.

References

[1] A. A. Bloom, J.-F. Exbrayat, I. R. van der Velde, L. Feng, and M. Williams, "The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 5, pp. 1285–1290, Feb. 2016.

[2] K.-H. Erb *et al.*, "Unexpectedly large impact of forest management and grazing on global vegetation biomass," *Nature*, vol. 553, no. 7686, pp. 73–76, Jan. 2018.

[3] R. A. Houghton, F. Hall, and S. J. Goetz, "Importance of biomass in the global carbon cycle," *J. Geophys. Res., Biogeosci.*, vol. 114, 2009, Art. no. G00E03.

[4] R. Ahmed, P. Siqueira, and S. Hensley, "Analyzing the uncertainty of biomass estimates from L-band radar backscatter over the Harvard and Howland forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3568–3586, Jun. 2014.

[5] A. Baccini *et al.*, "Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps," *Nature Climate Change*, vol. 2, no. 3, pp. 182–185, 2012.

[6] D. Lu, Q. Chen, G. Wang, L. Liu, G. Li, and E. Moran, "A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems," *Int. J. Digit. Earth*, vol. 9, no. 1, pp. 63–105, 2016.

[7] S. S. Saatchi *et al.*, "Benchmark map of forest carbon stocks in tropical regions across three continents," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 24, pp. 9899–9904, Jun. 2011.

[8] S. Sinha, C. Jeganathan, L. K. Sharma, and M. S. Nathawat, "A review of radar remote sensing for biomass estimation," *Int. J. Environ. Sci. Technol.*, vol. 12, no. 5, pp. 1779–1792, 2015.

[9] Y. Zhang, S. Liang, and G. Sun, "Forest biomass mapping of northeastern China using GLAS and MODIS data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 140–152, Jan. 2014.

[10] Y. Zhang, S. Liang, and L. Yang, "A review of regional and global gridded forest biomass datasets," *Remote Sens.*, vol. 11, no. 23, 2019, Art. no. 2744.

[11] S. G. Zolkos, S. J. Goetz, and R. Dubayah, "A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing," *Remote Sens. Environ.*, vol. 128, pp. 289–298, 2013.

[12] M. L. Clark, D. A. Roberts, J. J. Ewel, and D. B. Clark, "Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors," *Remote Sens. Environ.*, vol. 115, no. 11, pp. 2931–2942, 2011.

[13] P. Hyde, R. Nelson, D. Kimes, and E. Levine, "Exploring LiDAR–RaDAR synergy—Predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR and InSAR," *Remote Sens. Environ.*, vol. 106, no. 1, pp. 28–38, 2007.

[14] M. A. Lefsky, W. B. Cohen, G. G. Parker, and D. J. Harding, "Lidar remote sensing for ecosystem studies," *BioScience*, vol. 52, pp. 19–30, Jan. 2002.

[15] P. Rodríguez-Veiga, J. Wheeler, V. Louis, K. Tansey, and H. Balzter, "Quantifying forest biomass carbon stocks from space," *Current Forestry Rep.*, vol. 3, no. 1, pp. 1–18, 2017.

[16] R. O. Dubaya and J. B. Drake, "Lidar remote sensing for forestry," *J. Forestry*, vol. 98, pp. 44–46, 2000.

[17] K. Lim, P. Treitz, M. Wulder, B. St-Onge, and M. Flood, "LiDAR remote sensing of forest structure," *Prog. Phys. Geogr., Earth Environ.*, vol. 27, no. 1, pp. 88–106, 2003.

[18] M. A. Wulder *et al.*, "Lidar sampling for large-area forest characterization: A review," *Remote Sens. Environ.*, vol. 121, pp. 196–209, 2012.

[19] K. Wang, T. Wang, and X. Liu, "A review: Individual tree species classification using integrated airborne LiDAR and optical imagery with a focus on the urban environment," *Forests*, vol. 10, no. 1, 2019, Art. no. 1.

[20] H. Tang *et al.*, "Deriving and validating leaf area index (LAI) at multiple spatial scales through lidar remote sensing: A case study in Sierra National Forest, CA," *Remote Sens. Environ.*, vol. 143, pp. 131–141, 2014.

[21] F. Zhao *et al.*, "Measuring effective leaf area index, foliage profile, and stand height in New England forest stands using a full-waveform ground-based lidar," *Remote Sens. Environ.*, vol. 115, no. 11, pp. 2954–2964, 2011.

[22] C. M. C. S. Listopad, J. B. Drake, R. E. Masters, and J. F. Weishampel, "Portable and airborne small footprint LiDAR: Forest canopy structure estimation of fire managed plots," *Remote Sens.*, vol. 3, no. 7, pp. 1284–1307, 2011.

[23] D. Seidel, S. Fleck, and C. Leuschner, "Analyzing forest canopies with ground-based laser scanning: A comparison with hemispherical photography," *Agricultural Forest Meteorol.*, vol. 154–155, pp. 1–8, 2012.

[24] C. J. Gleason and J. Im, "Forest biomass estimation from airborne LiDAR data using machine learning approaches," *Remote Sens. Environ.*, vol. 125, pp. 80–91, 2012.

[25] R. Nelson *et al.*, "Lidar-based estimates of aboveground biomass in the continental US and Mexico using ground, airborne, and satellite observations," *Remote Sens. Environ.*, vol. 188, pp. 127–140, 2017.

[26] T. Yao *et al.*, "Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar," *Remote Sens. Environ.*, vol. 115, no. 11, pp. 2965–2974, 2011.

[27] H. A. Margolis *et al.*, "Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America," *Can. J. Forest Res.*, vol. 45, pp. 838–855, 2015.

[28] J. B. Abshire *et al.*, "Geoscience laser altimeter system (GLAS) on the ICESat mission: On-orbit measurement performance," *Geophys. Res. Lett.*, vol. 32, no. 21, 2005, Art. no. L21S02.

[29] H. J. Zwally *et al.*, "ICESat's laser measurements of polar ice, atmosphere, ocean, and land," *J. Geodyn.*, vol. 34, pp. 405–445, 2002.

[30] N. Baghdadi *et al.*, "Testing different methods of forest height and aboveground biomass estimations from ICESat/GLAS data in eucalyptus plantations in Brazil," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 290–299, Jan. 2014.

[31] H. Huang, C. Liu, X. Wang, X. Zhou, and P. Gong, "Integration of multi-resource remotely sensed data and allometric models for forest aboveground biomass estimation in China," *Remote Sens. Environ.*, vol. 221, pp. 225–234, 2019.

[32] Y. Yu, X. Yang, and W. Fan, "Estimates of forest structure parameters from GLAS data and multi-angle imaging spectrometer data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 38, pp. 65–71, 2015.

[33] L. Duncanson *et al.*, "Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California," *Remote Sens. Environ.*, vol. 242, 2020, Art. no. 111779.

[34] C. A. Silva *et al.*, "Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping," *Remote Sens. Environ.*, vol. 253, 2021, Art. no. 112234.

[35] G. Sun, K. J. Ranson, D. S. Kimes, J. B. Blair, and K. Kovacs, "Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data," *Remote Sens. Environ.*, vol. 112, no. 1, pp. 107–117, 2008.

[36] J. E. Means *et al.*, "Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western cascades of Oregon," *Remote Sens. Environ.*, vol. 67, no. 3, pp. 298–308, 1999.

[37] M. A. Lefsky, D. Harding, W. B. Cohen, G. Parker, and H. H. Shugart, "Surface lidar remote sensing of basal area and biomass in deciduous forests of Eastern Maryland, USA," *Remote Sens. Environ.*, vol. 67, pp. 83–98, 1999.

[38] J. B. Drake *et al.*, "Estimation of tropical forest structural characteristics using large-footprint lidar," *Remote Sens. Environ.*, vol. 79, pp. 305–319, 2002.

[39] J. B. Drake *et al.*, "Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: Factors affecting the generality of relationships," *Global Ecol. Biogeogr.*, vol. 12, pp. 147–159, 2003.

[40] M. A. Lefsky, M. Keller, Y. Pang, P. B. de Camargo, and M. O. Hunter, "Revised method for forest canopy height estimation from geoscience laser altimeter system waveforms," *J. Appl. Remote Sens.*, vol. 1, Sep. 2007, Art. no. 013537.

[41] M. A. Lefsky, A. T. Hudak, W. B. Cohen, and S. A. Acker, "Patterns of covariance between forest stand and canopy structure in the Pacific Northwest," *Remote Sens. Environ.*, vol. 95, no. 4, pp. 517–531, 2005.

[42] A. Razi, S. Paul, and S. Hensley, "A study of forest biomass estimates from lidar in the northern temperate forests of New England," *Remote Sens. Environ.*, vol. 130, pp. 121–135, 2013.

[43] T. Hu *et al.*, "Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 565.

[44] I. Fayad *et al.*, "Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 52, pp. 502–514, 2016.

[45] H. Chi *et al.*, "Estimation of forest aboveground biomass in Changbai mountain region using ICESat/GLAS and Landsat/TM data," *Remote Sens.*, vol. 9, no. 7, 2017, Art. no. 707.

[46] F. E. Fassnacht *et al.*, "Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass," *Remote Sens. Environ.*, vol. 154, pp. 102–114, 2014.

[47] Y. Pang, M. Lefsky, G. Sun, M. E. Miller, and Z. Li, "Temperate forest height estimation performance using ICESat GLAS data from different observation periods," in *Proc. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 2008, vol. 37, pp. 777–782.

[48] G. Sun *et al.*, "Estimation of tree height and forest biomass from GLAS data," *J. Forest Plan.*, vol. 13, pp. 157–164, 2008.

[49] F. Enßle, J. Heinzel, and B. Koch, "Accuracy of vegetation height and terrain elevation derived from ICESat/GLAS in forested areas," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 31, pp. 37–44, 2014.

[50] M. R. Pourrahmati *et al.*, "Capability of GLAS/ICESat data to estimate forest canopy height and volume in mountainous forests of Iran," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5246–5261, Nov. 2015.

[51] A. A. Borsa, H. A. Fricker, and K. M. Brunt, "A terrestrial validation of ICESat elevation measurements and implications for global reanalyses," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6946–6959, Sep. 2019.

[52] M. A. Hofton, J. B. Minster, and J. B. Blair, "Decomposition of laser altimeter waveforms," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 4, pp. 1989–1996, Jul. 2000.
[53] V. H. Duong, R. Lindenbergh, N. Pfeifer, and G. Vosselman, "Single and two epoch analysis of ICESat full waveform data over forested areas," *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1453–1473, 2008.
[54] G. Mountrakis and Y. Li, "A linearly approximated iterative Gaussian decomposition method for waveform LiDAR processing," *ISPRS J. Photogramm. Remote Sens.*, vol. 129, pp. 200–211, 2017.
[55] A. C. Brenner *et al.*, "The algorithm theoretical basis document for the derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights," Nat. Aeronaut. Space Admin., Washington, DC, USA, Tech. Rep. NASA/TM-2012-208641, Aug. 2012.
[56] Q. Chen, "Assessment of terrain elevation derived from satellite laser altimetry over mountainous forest areas using airborne lidar data," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 111–122, 2010.
[57] J. A. B. Rosette, P. R. J. North, and J. C. Suárez, "Vegetation height estimates for a mixed temperate forest using satellite laser altimetry," *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1475–1493, 2008.
[58] D. J. Selkowitz, G. Green, B. Peterson, and B. Wylie, "A multi-sensor lidar, multi-spectral and multi-angular approach for mapping canopy height in boreal forest regions," *Remote Sens. Environ.*, vol. 121, pp. 458–471, 2012.
[59] M. Simard, N. Pinto, J. B. Fisher, and A. Baccini, "Mapping forest canopy height globally with spaceborne lidar," *J. Geophys. Res., Biogeosci.*, vol. 116, no. G4, 2011, Art. no. G04021.
[60] Y. Wang *et al.*, "A combined GLAS and MODIS estimation of the global distribution of mean forest canopy height," *Remote Sens. Environ.*, vol. 174, pp. 24–43, 2016.
[61] K. Liu, W. Jindi, W.-S. Zeng, and S. Jinling, "Comparison and evaluation of three methods for estimating forest above ground biomass using TM and GLAS data," *Remote Sens.*, vol. 9, no. 4, 2017, Art. no. 341.
[62] I. Chrysafis, G. Mallinis, I. Gitas, and M. Tsakiri-Strati, "Estimating Mediterranean forest parameters using multi seasonal Landsat 8 OLI imagery and an ensemble learning method," *Remote Sens. Environ.*, vol. 199, pp. 154–166, 2017.
[63] Y. Saeys, T. Abeel, and Y. van de Peer, "Robust feature selection using ensemble feature selection techniques," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2008, pp. 313–325.
[64] Y. Fu, H. He, T. Hawbaker, P. Henne, Z. Zhu, and D. Larsen, "Evaluating k-nearest neighbor (kNN) imputation models for species-level aboveground forest biomass mapping in Northeast China," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 2005.
[65] X. Wang, S. Wang, and L. Dai, "Estimating and mapping forest biomass in northeast China using joint forest resources inventory and remote sensing data," *J. Forestry Res.*, vol. 29, no. 3, pp. 797–811, 2018.
[66] A. T. N. Dang, S. Nandy, R. Srinet, N. V. Luong, S. Ghosh, and A. Senthil Kumar, "Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam," *Ecol. Inform.*, vol. 50, pp. 24–32, 2019.
[67] S. Arjasakusuma, S. Swahyu Kusuma, and S. Phinn, "Evaluating variable selection and machine learning algorithms for estimating forest heights by combining lidar and hyperspectral data," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 9, 2020, Art. no. 507.
[68] W. Yang and C. C. Gu, "Selection of important variables by statistical learning in genome-wide association analysis," *BMC Proc.*, vol. 3, Dec. 2009, Art. no. S70.
[69] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011.
[70] C. Wu *et al.*, "Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery," *J. Appl. Remote Sens.*, vol. 10, no. 3, 2016, Art. no. 035010.
[71] Z. Guo, H. Chi, and G. Sun, "Estimating forest aboveground biomass using HJ-1 satellite CCD and ICESat GLAS waveform data," *Sci. China Earth Sci.*, vol. 53, no. S1, pp. 16–25, 2010.
[72] G. M. Foody, "Sample size determination for image classification accuracy assessment and comparison," *Int. J. Remote Sens.*, vol. 30, no. 20, pp. 5273–5291, 2009.
[73] H. Latifi, F. Fassnacht, and B. Koch, "Forest structure modeling with combined airborne hyperspectral and LiDAR data," *Remote Sens. Environ.*, vol. 121, pp. 10–25, 2012.
[74] G. W. Frazer, S. Magnussen, M. A. Wulder, and K. O. Niemann, "Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass," *Remote Sens. Environ.*, vol. 115, no. 2, pp. 636–649, 2011.
[75] J. L. Hernández-Stefanoni *et al.*, "Effects of sample plot size and GPS location errors on aboveground biomass estimates from LiDAR in tropical dry forests," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1586.
[76] L. Cao *et al.*, "Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data," *Remote Sens. Environ.*, vol. 178, pp. 158–171, 2016.
[77] E. Næsset, O. M. Bollandsås, T. Gobakken, S. Solberg, and R. E. McRoberts, "The effects of field plot size on model-assisted estimation of aboveground biomass change using multitemporal interferometric SAR and airborne laser scanning data," *Remote Sens. Environ.*, vol. 168, pp. 252–264, 2015.
[78] G. Matasci *et al.*, "Three decades of forest structural dynamics over Canada's forested ecosystems using Landsat time-series and lidar plots," *Remote Sens. Environ.*, vol. 216, pp. 697–714, 2018.
[79] W. Shen, M. Li, C. Huang, X. Tao, and A. Wei, "Annual forest aboveground biomass changes mapped using ICESat/GLAS measurements, historical inventory data, and time-series optical and radar imagery for Guangdong province, China," *Agricultural Forest Meteorol.*, vol. 259, pp. 23–38, 2018.

Yuzhen Zhang received the Ph.D. degree in global environmental change from Beijing Normal University, Beijing, China, in 2014.

She is currently an Associate Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. Her main research interests include forest biomass mapping from lidar, optical and radar data, assessing the impact of climate and disturbances on forest biomass dynamics, and exploring the climatic effects of forest disturbances and land use land cover changes.

Wenhao Li received the B.S. degree in automation from the Beijing Institute of Petrochemical Technology, Beijing, China, in 2015. He is currently working toward the M.Sc. degree in control science and engineering with the University of Science and Technology Beijing, Beijing.

His research interests include estimating forest biomass and its dynamics from multisource remote sensing data.

Shunlin Liang (Fellow, IEEE) received the Ph.D. degree in geography from Boston University, Boston, MA, USA.

He is currently a Professor with the Department of Geographical Sciences, University of Maryland, College Park, MD, USA. His main research interests include estimating land surface variables from satellite data, global satellite product generation, earth's energy balance, and environmental changes. He has authored or coauthored more than 390 SCI indexed peer-reviewed journal papers, 42 book chapters, and nine special issues of different journals. He authored/edited seven books and four of which were translated into Chinese, such as "Quantitative Remote Sensing of Land Surfaces" (Wiley, 2004), "Advances in Land Remote Sensing: System, Modeling, Inversion, and Application" (Springer, 2008), "Advanced Remote Sensing: Terrestrial Information Extraction and Applications" (Academic Press, 2012 and 2019), "Global Land Surface Satellite Products: Algorithms, Validation, and Analysis" (Springer, 2013), "Land Surface Observation, Modeling, Data Assimilation" (World Scientific, 2013), and "Earth's Energy Budget" (Elsevier, 2017). He has led a team to develop the GLASS products that are freely available (www.glass.umd.edu and www.geodata.cn) and widely used. He was the Editor-in-Chief for the nine-volume books entitled "Comprehensive Remote Sensing" (Elsevier, 2017). He was an Associate Editor for the IEEE TRANSACTION ON GEOSCIENCE AND REMOTE SENSING and is currently an Editor-in-Chief for the Science of Remote Sensing.