Article

Canopy Height Layering Biomass Estimation Model (CHL-BEM) with Full-Waveform LiDAR

Jinyan Tian 1,2, Le Wang 3,*, Xiajuan Li 1,2, Dameng Yin 3, Huili Gong 1,2, Sheng Nie 4, Chen Shi 1,2, Ruofei Zhong 1,2, Xiaomeng Liu 1,2 and Ronglong Xu 1,2

1 Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, China; b308@cnu.edu.cn (J.T.); lixiaojuan@cnu.edu.cn (X.L.); 6506@cnu.edu.cn (H.G.);
2 College of Resources Environment and Tourism, Capital Normal University, Beijing 100048, China;
3 Department of Geography, The State University of New York at Buffalo, Buffalo, NY 14261, USA;
4 Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; niesheng@radi.ac.cn
* Correspondence: lewang@buffalo.edu

Received: 9 May 2019; Accepted: 14 June 2019; Published: 18 June 2019

Abstract: Forest biomass is an important descriptor for studying carbon storage, carbon cycles, and global change science. The full-waveform spaceborne Light Detection And Ranging (LiDAR) Geoscience Laser Altimeter System (GLAS) provides great possibilities for large-scale and long-term biomass estimation. To the best of our knowledge, most of the existing research has utilized average tree height (or height metrics) within a GLAS footprint as the key parameter for biomass estimation. However, the vertical distribution of tree height is usually not as homogeneous as we would expect within such a large footprint of more than 2000 m², which would limit the biomass estimation accuracy vastly. Therefore, we aim to develop a novel canopy height layering biomass estimation model (CHL-BEM) with GLAS data in this study. First, all the trees with similar height were regarded as one canopy layer within each GLAS footprint. Second, the canopy height and canopy cover of each layer were derived from GLAS waveform parameters. These parameters were extracted using a waveform decomposition algorithm (refined Levenberg–Marquardt—RLM), which assumed that each decomposed vegetation signal corresponded to a particular canopy height layer. Third, the biomass estimation model (CHL-BEM) was established by using the canopy height and canopy cover of each height layer. Finally, the CHL-BEM was compared with two typical biomass estimation models of GLAS in the study site located in Ejina, China, where the dominant species was Populus euphratica. The results showed that the CHL-BEM presented good agreement with the field measurement biomass ($R^2 = 0.741$, RMSE = 0.487, %RMSE = 24.192) and achieved a significantly higher accuracy than the other two models. As a whole, we expect our method to advance all the full-waveform LiDAR development and applications, e.g., the newly launched Global Ecosystem Dynamics Investigation (GEDI).

Keywords: biomass; full-waveform LiDAR; GLAS; GEDI; UAV LiDAR

1. Introduction

Forest biomass, commonly defined as the total amount of organic matter (dry weight) per unit ground surface area [1,2], includes both aboveground biomass and underground biomass. It should be noted that biomass refers to aboveground biomass in this study. Accurate forest biomass information plays vital roles in evaluating carbon sequestration capacity and carbon storage, as well as studying carbon cycle processes and global change science [2–5].
Remote sensing technology provides a convenient and cost-effective way to estimate forest biomass [6,7], especially Light Detection And Ranging (LiDAR) [8,9]. The successful estimation of the biomass in forests with both airborne and spaceborne platforms has been reported in several studies [10–18]. However, it is logistically difficult to employ airborne LiDAR when large-scale forest biomass needs to be monitored [12,19–22]. Fortunately, Ice, Cloud and land Elevation Satellite (ICESat)/Geoscience Laser Altimeter System (GLAS), as the first full-waveform spaceborne LiDAR system, has demonstrated its advantage in collecting extensive and repetitive forest biomass information, as GLAS data is free, has global coverage, and has repeatable observations [23–30].

Almost all of the GLAS-based forest biomass estimation methods are statistical methods of waveform parameters, which can be grouped into two categories: Direct statistical methods and indirect statistical methods. The direct statistical methods estimate the biomass by building a regression model directly between GLAS waveform parameters and field measured (or airborne LiDAR-derived) biomass [19,21,31]. For example, Zhang et al. [21] acquired as many GLAS metrics as possible, according to the existing research. Then, support vector regression (SVR) was implemented to build the relationship between GLAS metrics, Moderate Resolution Imaging Spectroradiometer (MODIS) vegetation indices, and field biomass data. It should be noted that, with the direct statistical method, it is normally difficult to find a universally applied statistical model that fits different study sites with complex terrain and various vegetation.

The second group of methods for estimating forest biomass is that of the indirect statistical methods, which currently encompass the mainstream approach. Indirect statistical methods estimate the biomass by first utilizing GLAS waveform parameters to obtain the average or maximum forest height within a GLAS footprint, and then they develop a regression relationship between forest height and field measured (or airborne LiDAR-derived) biomass [2,3,18,20,21,32–34]. For example, Lefsky et al. [33] employed LEE (the leading edge extent), GE (ground extent, i.e., the terrain index) and WE (the waveform extent) from GLAS waveform and Shuttle Radar Topographic Mission (SRTM) data to calculate the maximum forest height within a GLAS footprint. Results showed that the ICESat-derived forest heights achieved a high correlation with field-estimates of biomass ($R^2 = 73\%$, RMSE = 58.3 Mg·ha$^{-1}$). Though the indirect statistical methods tend to overestimate biomass over complex terrain [35], they are easier to understand in principle and more applicable than the direct statistical methods. Nevertheless, we should note that both the indirect and direct statistical methods attempt to acquire average or maximum forest height or height metrics within a GLAS footprint. However, the tree height is usually not homogeneous within such a large footprint (>2000 m$^2$), which greatly limits biomass estimation accuracy.

This study aims to develop a novel canopy height layering biomass estimation model (CHL-BEM) with GLAS data, one which belongs to the indirect statistical method category. The objectives of this study are three-fold: (1) To acquire the canopy height and canopy cover of each height layer from GLAS waveform decomposition parameters; (2) to establish a novel biomass estimation model with the canopy height and canopy cover of multiple height layers within a GLAS footprint; (3) to compare and assess the benefits and disadvantages of the CHL-BEM and two typical GLAS biomass estimation models. Compared to the existing biomass estimation methods with the GLAS, our innovation mainly lies in that we employed multiple canopy height layers instead of one canopy height layer. Ultimately, we expect that the CHL-BEM will improve the biomass estimation accuracy and advance all full-waveform LiDAR development and applications, e.g., estimating biomass from the newly launched Global Ecosystem Dynamics Investigation (GEDI).

2. Materials

2.1. Study Site

The study site is located in the Ejina Banner in the northwest area of Inner Mongolia, China (39°52′N–42°47′N and 97°10′–103°7′E), which covers an overall area of 1300 km$^2$ (Figure 1).
Ejina has a north temperate continental arid climate characterized by drought, less rain, large evaporation within each year [36]. The terrain of the study area is quite flat (the average slope is less than 3°), with the elevation ranging from 820 to 1127 m above sea level. The study site is dominated by three vegetation species: *Populus euphratica*, *Tamarix ramosissima*, and *Sophora alopecuroides*, which function as the first barriers protecting northwest China from sandstorms. The top height of *Populus euphratica* in this study site ranges from 5 to 16 m [37]. The vertical structure of most of these trees is regular, despite a few exceptions due to artificial or natural factors.

Figure 1. The study site and its geographical location. The zoomed-in figure of Ejina Oasis is a screenshot image.

2.2. GLAS Data

In total, 79 GLAS waveform data, randomly sampled, were acquired and preprocessed to test the new biomass estimation model proposed in this study (CHL-BEM). In this study site, each GLAS footprint can be considered as a circle with a diameter of 53 m. The data and related processing software were provided by NASA (http://nsidc.org/data/icesat). The raw GLAS waveform data were gleaned by combining the GLA-01 and GLA-14 products. Detailed processing steps of the GLAS data can be found in [37,38]. These 79 GLAS waveform data were collected in March 2004, 2005, and 2007 and in November 2003, 2005, and 2006, respectively. It should be noted that there exists a large difference between GLAS data acquisition time and field measurement time. Fortunately, the growth rate of *Populus euphratica* is slow because (1) most of them are trees that are a few hundred years old; (2) the growth rate of *Populus euphratica* has a strong positive correlation with soil water content [39], and our study site is located in an extremely arid desert area. Moreover, the spatial distribution of *Populus euphratica* changes marginally because this study site was the National *Populus euphratica* natural reserves, which are well protected. In this study, we separated the 79 GLAS waveform data into two parts: 30 GLAS waveform data were used to calculate the parameters in the biomass estimation models,
whereas the remaining 49 GLAS waveform data were utilized for validation. The GLAS footprints for both training and testing were distributed as evenly as possible in the whole study area (Figure 2).

Figure 2. QuickBird imagery of the study area with 79 Geoscience Laser Altimeter System (GLAS) footprints.

2.3. Field Biomass

In order to test the performance of the proposed CHL-BEM, we collected 79 field-estimate biomass data at the GLAS footprint level. The precise extent of each GLAS footprint was located using a hand-held high-precision GPS (Global Positioning System) (precision: 1 m) and a QuickBird image (spatial resolution: 0.6 m). Dong et al. [40] obtained the dry weight of Populus euphratica by the whole harvest method and established an empirical model among tree height, breast diameter (diameter of the trunk at 1.3 m above the ground), canopy cover, and biomass at the individual tree level. Because we could not acquire the biomass of Populus euphratica by means of felling and weighing in this study site, the empirical model by Dong et al. [40] was utilized to calculate the biomass at the individual tree level. Therefore, the field-estimate biomass at the GLAS footprint level can be described according to the following Equations (1)–(4). Equations (1)–(3) were used to calculate the biomass for each individual tree, and the biomass at the GLAS footprint level was calculated using Equation (4)

\[ M_{\text{Trunk}} = 125.546 \times (D^2 \times H)^{0.654} \]  
\[ M_{\text{Crown}} = 0.016 \times A^2 + 2.291 \times A + 11.084 \]
\[ M_{\text{individual}} = M_{\text{Trunk}} + M_{\text{Crown}} \]  

(3)

\[ B_{\text{GLAS}} = \frac{\sum_{i=1}^{n} M_{i}^{\text{individual}}}{S_{\text{GLAS}}} \]  

(4)

where \( i \) is the \( i \)th individual tree; \( n \) is the total number of trees within a GLAS footprint; \( M_{\text{Trunk}}, M_{\text{Crown}}, \) and \( M_{\text{individual}} \) are the dry weight of tree trunk, tree crown, and individual tree, respectively; \( H, D, \) and \( A \) are the maximum tree height, breast diameter, and canopy cover, respectively; and \( S_{\text{GLAS}} \) and \( B_{\text{GLAS}} \) are the area and biomass of each GLAS footprint.

According to the empirical model mentioned above, we should acquire field measurements of parameters including tree height, breast diameter, and canopy cover. From 2013 to 2015, field survey methods were employed to measure these three parameters. Recently, unmanned aerial vehicle (UAV) remote sensing has opened the door to new sources of data to effectively characterize vegetation metrics at very high spatial resolutions [41,42] and with very high accuracy [43]. LiDAR data were used to test tree height and canopy cover measured from a field survey (in only three GLAS footprints). Results showed that the tree height and canopy cover from field survey matched that of UAV LiDAR. In addition, the breast diameter was measured with a tape. As a whole, the tree height, breast diameter, and canopy cover of 1430 individual trees were collected within the 79 GLAS footprints. Details about field survey methods can be seen in our previous work [37]. For each GLAS footprint, the biomass was calculated according to Equations (1)–(4) (see Figure 3).

![Figure 3. Field measurements of biomass at the GLAS footprint level.](image)

3. Methods

Three key steps were engaged in our developed method for biomass estimation at the GLAS footprint level: GLAS preprocessing, refined Levenberg–Marquardt (RLM) Gaussian decomposition, and biomass estimation with canopy height and canopy cover from different height layers. The overall flowchart is shown in Figure 4.
3.1. **GLAS Preprocessing**

To suppress the influence of system noise and background noise, a waveform filtering algorithm was essential for acquiring reliable GLAS waveform data. In this study, the Savitzky–Golay filtering algorithm was employed, as it had been successfully applied to GLAS waveform data processing \[37,44,45\]. Besides waveform filtering, detecting the signal start, end, and peak locations was the other key step of GLAS preprocessing. Following the report by Sun et al. \[38\], we identified the signal start and end locations according to the standard deviation of GLAS waveform return volts. Moreover, a second-order derivative was employed to search for the peak location. Details about the two specific GLAS preprocessing steps can be found in our previous works \[37,44\].

3.2. **RLM Gaussian Decomposition Method**

To obtain the canopy height and canopy cover of each height layer within a GLAS footprint, we adopted a waveform decomposition algorithm \[46–51\], RLM, assuming that each canopy height layer forms a vegetation Gaussian signal. RLM embodies the morphological characteristics of forest vertical structures which serve as restrictions for the parameters’ initialization and criteria for iteration termination. RLM not only injects physical meaning in the Gaussian decomposition method but also overcomes the overfitting problem in traditional methods where only geometric characteristics were considered. The waveform decomposition result of RLM is shown in Figure 5. Details about the RLM Gaussian decomposition method can be found in our previous work \[37\].
3.3. Canopy Height Layering Biomass Estimation Model (CHL-BEM)

Most of the existing research utilized average tree height within a GLAS footprint as the only input parameter for biomass estimation. However, the tree height is usually not homogeneous because each footprint is over 2000 m$^2$, omitting which will lower the biomass estimation accuracy. To this end, in this study, we proposed the CHL-BEM, which considers different canopy height layers. Two key steps were engaged: (1) Obtaining the canopy height and canopy cover of each height layer from the decomposed waveform data (see Section 3.3.1); (2) establishing a biomass estimation model with canopy height and canopy cover from different height layers (see Section 3.3.2).

3.3.1. Canopy Height and Canopy Cover

As shown in Figure 5, RLM decomposed the GLAS waveform into three Gaussian signals which corresponded to the ground and two canopy height layers. For each canopy layer, the canopy top height (CTH) was determined by the signal start and the ground peak positions, which can be defined as Equation (5):

$$CTH_i = \frac{T_G - T_{S,i}}{2} \times c$$

(5)

where $i$ represents the $i$th the canopy signal, $T_G$ is the ground peak position, $T_{S,i}$ is the canopy signal start position; and $c$ is the speed of light ($3 \times 10^8$ m/s).

Besides CTH, canopy cover (CC) served as the other key parameter in the proposed biomass estimation model. For each canopy layer, CC was defined as the ratio of each canopy layer return energy to total waveform energy, which can be simplified as Equation (6):

$$CC_i = \frac{CE_i}{\sum_{i=1}^{N} CE_i + GE}$$

(6)

where $CE$ represents the canopy echo energy, $GE$ represents the ground echo energy, and $N$ is the number of canopy height layers.
3.3.2. Biomass Estimation

To obtain the biomass of each GLAS footprint, the dry weights of each canopy layer were acquired first according to Equation (7). Therefore, the biomass at a GLAS footprint level can be calculated by Equation (8).

\[
M_i = a \times (CTH_i)^b \times (CC_i)^c + d
\]  

\[
B_{GLAS} = \frac{\sum_{i=1}^{N} M_i}{S_{GLAS}}
\]

where \(M\) represents the dry weight of each layer; \(a\), \(b\), \(c\), and \(d\) were the constant parameters to be determined; and \(S_{GLAS}\) was the area of a GLAS footprint.

3.4. Two Other Biomass Models for Comparison

In addition to the CHL-BEM, we also implemented the other two biomass estimation models to use for comparison. They were: The canopy top height biomass estimation model (CTH-BEM) and the canopy top height & canopy cover biomass estimation model (CTH&CC-BEM). The CTH-BEM only employs CTH as the key input parameter for biomass estimation, whereas CC was added in the CTH&CC-BEM in addition CTH. It should be noted that one CTH value and one CC value were acquired within a GLAS footprint because these two models treated all trees as a whole without stratification. Equations (9) and (10) for the CTH-BEM and CTH&CC-BEM were defined as:

\[
B_{CTH-BEM}^{GLAS} = \frac{a_1 \times (CTH)^{b_1} + c_1}{S_{GLAS}}
\]

\[
B_{CTH&CC-BEM}^{GLAS} = \frac{a_2 \times (CTH)^{b_2} \times (CC)^{c_2} + d_2}{S_{GLAS}}
\]

where \(a_1\), \(b_1\), \(c_1\), \(a_2\), \(b_2\), \(c_2\), and \(d_2\) were constants to be determined, and \(S_{GLAS}\) was the area of a GLAS footprint.

4. Results and Discussion

4.1. CHL-BEM Estimated Biomass

In Equation (7), four constant parameters were estimated with 30 of the 79 GLAS footprints by using the optimization software package 1stOpt (\(a = 1983.916\), \(b = 1.050\), \(c = 1.237\), \(d = 1444.028\)), in which \(R^2\) was 0.767 and RMSE was 846.131. The remaining 49 GLAS footprints and field-estimated biomass were used to validate the CHL-BEM estimated biomass.

Figure 6 shows that the CHL-BEM obtained a relatively high accuracy in the estimation of biomass at the GLAS footprint level (\(R^2 = 0.741\), RMSE = 0.487, %RMSE = 24.192). Moreover, the fitted line is very close to the 1:1 line. Though we acquired a satisfactory biomass estimation result at the GLAS footprint level, it should be noted that some limitations still exist.

First, a limitation from the RLM Gaussian decomposition method. Though RLM has shown success in addressing the challenging hypothesis that each vegetation Gaussian signal corresponds to a particular canopy height layer, it failed in a few GLAS footprints where the tree height was evenly distributed in the vertical direction. In addition, some GLAS waveform data were recorded in winter when the shape of Populus euphratica was irregular, which caused the canopy layer to not form a Gaussian signal. Therefore, the RLM Gaussian decomposition method still needs to be improved to take the above two issues into account.
Second, a limitation from the parameter fitting in the CHL-BEM. In Section 3.3.2, $a$, $b$, $c$, and $d$ were set the same across canopy height layers. Though this reduced the complexity and improved the applicability of the model, there should exist some difference in those parameters for each canopy layer. Therefore, an automated parameter setting method for different canopy height layers would help to improve the biomass estimation accuracy.

Third, a limitation of the universality in fitting study sites with complex terrain and various vegetation. Our study site was located in a very flat region where the only dominant species was *Populus euphratica*, at least in the 79 GLAS footprints. It should be noted that the CHL-BEM needs to be refined when the terrain becomes not as flat. On the other hand, the CHL-BEM should be improved to accommodate more forest types, especially for mixed forests that consist of a variety of vegetation species.

Finally, we have to admit that some errors were introduced due to the different data acquisition times between the field survey and the GLAS waveform data. However, we could not avoid this problem because (1) the latest GLAS data were available in 2009; (2) the spatial distribution of GLAS footprints and *Populus euphratica* are sparse, and, therefore, very few GLAS data could be used to validate our method. Though the biomass of *Populus euphratica* continued to grow during those years, the impact of this growth on the accuracy assessment of the statistical biomass estimation models would not be significant because all *Populus euphratica* trees were growing at the same time.

![Figure 6](image_url)

**Figure 6.** The relationship between the canopy height layering biomass estimation model (CHL-BEM) estimated biomass and field-estimated biomass.

### 4.2. CHL-BEM vs. Other Two Biomass Models

The comparison results of the three different biomass estimation models at the GLAS footprint level are shown in Table 1, which indicates that the CHL-BEM ($R^2 = 0.741$, RMSE = 0.487, %RMSE = 24.192) obtained higher accuracy than both the CTH-BEM ($a_1 = 9.025 \times 10^{-4}$, $b_1 = 2.670$, $c_1 = 1.075$, $R^2 = 0.210$, RMSE = 0.847, %RMSE = 40.025) and the CTH&CC-BEM ($a_2 = 2060.602$, $b_2 = 0.844$, $c_2 = 0.996$, $d_2 = 1508.274$, $R^2 = 0.522$, RMSE = 0.661, %RMSE = 32.829). Accordingly, three conclusions can be drawn:
The CTH-BEM obtained the lowest accuracy among the three models. This was mainly caused by the distribution characteristics of *Populus euphratica* in the horizontal and vertical directions. In the vertical direction, the height of *Populus euphratica* ranges from 5 to 19 m within some GLAS footprints. The CTH-BEM only utilizes the maximum height for biomass estimation, which leads to poor accuracy. In the horizontal direction, the spatial distribution of *Populus euphratica* is quite sparse and uneven. Canopy cover not being considered in the CTH-BEM also leads to a poor biomass estimation accuracy. Therefore, the canopy height layering methods for both canopy height and canopy cover must be considered to address the abovementioned issues.

The CTH&CC-BEM obtained a relatively higher accuracy than the CTH-BEM, but it was still not ideal. Results showed that the CTH&CC-BEM gained a higher accuracy than the CTH-BEM, which was mainly because the canopy cover was considered as an additional key input parameter. As we all know, breast diameter is a key indicator of biomass estimation, but we cannot acquire this information from GLAS data. Some existing research has confirmed that a positive correlation exists between canopy cover and breast diameter [17,52,53]. Therefore, considering the canopy cover improved the biomass estimation results.

The CHL-BEM obtained the highest accuracy among the three models. Though there still exist some limitations, as mentioned in Section 4.1, it achieved a satisfactory accuracy in biomass estimation. This was mainly because the CHL-BEM divided all trees within a GLAS footprint into multiple height layers and used the layered canopy height and canopy cover for biomass estimation. Therefore, we suggest that canopy height layering methods should be further explored to improve forest biomass estimation accuracy in future work. The newly developed CHL-BEM can be applied on other full-waveform LiDAR data such as GEDI.

Table 1. Comparison of three different GLAS biomass estimated models. In this table, \( x \) represents the model-estimated biomass, and \( y \) represents the field-estimated biomass.

<table>
<thead>
<tr>
<th>Models</th>
<th>Input</th>
<th>Fitting Equation</th>
<th>( R^2 )</th>
<th>RMSE</th>
<th>%RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTH-BEM</td>
<td>Height</td>
<td>( y = 1.0475x - 0.0508 )</td>
<td>0.210</td>
<td>0.847</td>
<td>42.025</td>
</tr>
<tr>
<td>CTH&amp;CC-BEM</td>
<td>Height and CC</td>
<td>( y = 1.0584x - 0.0723 )</td>
<td>0.522</td>
<td>0.661</td>
<td>32.829</td>
</tr>
<tr>
<td>CHL-BEM</td>
<td>Layered Height and CC</td>
<td>( y = 0.9691x + 0.105 )</td>
<td>0.741</td>
<td>0.487</td>
<td>24.192</td>
</tr>
</tbody>
</table>

5. Conclusions

This research developed a novel biomass estimation model (CHL-BEM) with full-waveform LiDAR (ICESat/GLAS) that could utilize the metrics from multiple canopy height layers as key input parameters. The CHL-BEM not only considered the vertical structure information of each canopy height layer but also took the canopy cover at the horizontal direction into account. By comparison with the CTH-BEM and CTH&CC-BEM models, we confirmed that the canopy height layering step played vital roles in biomass estimation for large footprint full-waveform LiDAR data. As a whole, we expect that our method will advance all full-waveform LiDAR development and applications, e.g., the newly launched GEDI.

**Author Contributions:** J.T. designed, collected data, implemented the manual measurements, and drafted the initial manuscript; L.W., X.L., H.G., S.N., D.Y., C.S., and R.Z. provided suggestions on discussion and structure of the manuscript. X.L. and R.X. made a contribution to modify the reviewers’ comments. All authors read the manuscript and gave valuable suggestions to improve the manuscript.

**Funding:** This study was supported by “First-class funding from the post-doctoral fund project (17540730001)”, “Special Funded Project of China Postdoctoral (18540340004)”, and “National Natural Science Foundation of China (41801331)”.

**Acknowledgments:** The authors thank Huabing Dai, Wenjie Ji, Sai Zhang, Lu Xun and Yuming Liu for assistance with collection of field data. We are also deeply indebted to the reviewer for the thoughtful and constructive comments on improving the manuscript.

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


20. Fatoyinbo, T.E.; Simard, M. Height and biomass of mangroves in Africa from ICESat/GLAS and SRTM. *Int. J. Remote Sens.* 2013, **34**, 668–681. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).