Article

Retrieving Forest Canopy Elements Clumping Index Using ICESat GLAS Lidar Data

Lei Cui [1,2,3], Ziti Jiao [1,2,*], Kaiguang Zhao [3,4], Mei Sun [1,2], Yadong Dong [1], Siyang Yin [1,2], Xiaoning Zhang [1,2], Jing Guo [1,2], Rui Xie [1,2], Zidong Zhu [1,2], Sijie Li [1,2] and Yidong Tong [1,2]

1   State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Aerospace Information Research Institute of Chinese Academy of Sciences, Beijing Normal University, Beijing 100875, China; cuil@mail.bnu.edu.cn (L.C.); sunm@mail.bnu.edu.cn (M.S.); dongyd@aircas.ac.cn (Y.D.); yinsy@mail.bnu.edu.cn (S.Y.); xnzhang@mail.bnu.edu.cn (X.Z.); guojing0404@mail.bnu.edu.cn (J.G.); xierui@mail.bnu.edu.cn (R.X.); zhuzidong@mail.bnu.edu.cn (Z.Z.); lisijie@mail.bnu.edu.cn (S.L.); 202021051064@mail.bnu.edu.cn (Y.T.)
2   Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
3   Ohio Agricultural Research and Development Center, School of Environment and Natural Resources, The Ohio State University, Wooster, OH 44691, USA; zhao.1423@osu.edu
4   Environmental Science Graduate Program, School of Environment and Natural Resources, The Ohio State University, Columbus, OH 43210, USA
*   Correspondence: jiaozt@bnu.edu.cn

Abstract: Clumping index (CI) is a canopy structural variable important for modeling the terrestrial biosphere, but its retrieval from remote sensing data remains one of the least reliable. The majority of regional or global CI products available so far were generated from multiangle optical reflectance data. However, these reflectance-based estimates have well-known limitations, such as the mere use of a linear relationship between the normalized difference hotspot and darkspot (NDHD) and CI, uncertainties in bidirectional reflectance distribution function (BRDF) models used to calculate the NDHD, and coarse spatial resolutions (e.g., hundreds of meters to several kilometers). To remedy these limitations and develop alternative methods for large-scale CI mapping, here we explored the use of spaceborne lidar—the Geoscience Laser Altimeter System (GLAS)—and proposed a semi-physical algorithm to estimate CI at the footprint level. Our algorithm was formulated to leverage the full vertical canopy profile information of the GLAS full-waveform data; it converted raw waveforms to forest canopy gap distributions and gap fractions of random canopies, which was used to estimate CI based on the radiative transfer theory and a revised Beer–Lambert model. We tested our algorithm over two areas in China—the Saihanba National Forest Park and Heilongjiang Province—and assessed its relative accuracies against field-measured CI and MODIS CI products. We found that reliable estimation of CI was possible only for GLAS waveforms with high signal-to-noise ratios (e.g., >65) and at gentle slopes (e.g., <12°). Our GLAS-based CI estimates for high-quality waveforms compared well to field-based CI (i.e., $R^2 = 0.72$, RMSE = 0.07, and bias = 0.02), but they showed less correlation to MODIS CI (e.g., $R^2 = 0.26$, RMSE = 0.12, and bias = 0.04). The difference highlights the impact of the scale effect in conducting comparisons of products with huge differences resolution. Overall, our analyses represent the first attempt to use spaceborne lidar to retrieve high-resolution forest CI and our algorithm holds promise for mapping CI globally.

Keywords: GLAS; lidar; model; clumping index (CI); Beer-Lambert law; gap fraction; forest

1. Introduction

Clumping index (CI) characterizes the level of foliage grouping within a distinct canopy structure relative to a random distribution, which is closely related to radiation interception and distribution within canopies [1,2]. Given the effectiveness of CI in describing the distribution of canopy elements and solar radiation within the canopy, it has been

widely used in remote sensing field. For example, CI is a necessary input parameter for mapping true leaf area index (LAI) [3–5], and is frequently used to determine the proportions of sunlit and shaded leaves in ecosystem process models [6–8]. In addition, previous studies found that the proper consideration of CI could obviously improve the estimation accuracy of gross primary productivity (GPP) and evapotranspiration (ET) [9,10].

The importance of CI leads to various efforts on estimating CI from different data sources. Multiangle remote sensing data has proven to be an effective data for estimating CI, partly because reflectances observed at different angles can capture the effect of canopy clumpiness to certain extents [11]. Existing models mainly relate the CI to an angular index named the normalized difference between hotspot reflectance and darkspot reflectance (NDHD) by constructing a linear relationship model, based on the simulation of a so-called 4-scale geometric-optical model [12]. Hotspot and darkspot reflectances, which are used to calculate NDHD, are usually reconstructed using bidirectional reflectance distribution function (BRDF) model [12,13]. Based on linear relationships between CI and NDHD, global and regional CI products have been generated from the available multiangle remote sensing data, such as POLDER [14], MODIS [15–17], and MISR [18].

Despite the wide use of NDHD-based CI estimation methods, these kinds of methods have some weaknesses that probably limit their applications. Firstly, the calculation of NDHD is sensitive to the choice of BRDF models [19], especially to hotspot effect [12,13,20]. Thus far, no census has concluded which BRDF model is the most appropriate for CI estimation. Furthermore, passive optical sensors have limited observation abilities in 3D objects, such as forest, because of low penetration characteristic and the spectral signal saturation problem [21–23]; most observations can only provide structural information on the horizontal distribution of observed objects. Thus, using observations from passive optical sensors to retrieve 3D vegetation structure information is rather limited, such as CI [24]. Another drawback is the coarse spatial resolution of CI products from multiangle remote sensing data.

Lidar provides a promising alternative technology to estimate CI remotely and avoids the limitations associated with satellite optical data. In addition, unlike optical imagery that captures horizontal canopy clumpiness, lidar data characterizes the full 3D distribution of phytoelements and the clumping along both vertical and horizontal directions [25–27]. Terrestrial and airborne discrete lidar data are the most widely used lidar data sources to estimate CI. At present, using these lidar data to retrieve forest canopy CI mainly includes two categories: statistically-based method and physically-based method. Statistically-based method was developed based on a regression relationship between lidar-obtained metrics and field-measured CI [28]. Although statistically-based methods are easy to conduct and computationally efficient in operation, they do not fully use the 3-D measurement information contained by lidar data. Furthermore, this kind of statistically-based method has poor universality and is usually used only at the site scale. In view of problems of statistically-based methods, researchers turn to develop physically-based methods for CI estimation. Physically-based methods are usually implemented by taking lidar data as input for LX (finite-length averaging method), CC (gap-size distribution method), or CLX (combination of gap-size distribution and finite-length averaging methods) these kinds of CI estimation methods [24,29]. Summarizing these existing physical-based methods, we found that most of these methods have to conduct lidar data processing steps before CI estimation, and different data processing strategies usually give different CI values. Therefore, the uncertainty caused by lidar data processing may be a major limitation for this kind of method and limit their further applications. Recently, a CI inversion method based on the hypothesis of random distribution theory and the CC method was proposed [30], which could avoid uncertainty caused by data processing. However, the ground-based validation did not perform well, which might be caused by the low point density of airborne laser scan (ALS) data.

Although CI estimation from lidar data is a good attempt to obtain fine resolution CI, existing methods still have some aspects for improvement, such as the uncertainty

caused by data processing, using low-density discrete lidar data cannot capture complete vegetation structure information, etc. Furthermore, all existing methods are applied only to regional or site scales because of airborne and terrestrial lidar equipment's limited observation ability. In the above sense, we try to determine a way to solve existing problems. Spaceborne lidar usually uses digital waveform to record observation and has global observation capabilities, which can capture more details than low-density discrete lidar data from ALS [31,32]. The characteristics of spaceborne lidar make it a suitable data source for estimating CI, but CI estimation methods based on spaceborne lidar do not exist yet, even though spaceborne lidar data, such as Geoscience Laser Altimeter System (GLAS), have been available for many years. Generally, digital waveforms recorded by spaceborne lidar can be taken as a function of the Beer–Lambert law. In the Beer–Lambert law, CI is a variable nonidentifiable from other variables, such as leaf area index and leaf area density [2]. This nonidentification makes it impossible to retrieve CI from lidar waveforms alone—a subtle theoretical difficulty that has not been articulated or explored before in the literature but can be potentially resolved by introducing auxiliary data and additional model assumptions to constrain the CI inversion. This theoretical gap is an important impetus that motivates this work.

Given that spaceborne lidar can provide 3D forest structure information at a high resolution globally, CI research based on spaceborne lidar data can broaden the perspective for understanding the structural characteristics of forests in global ecosystem studies. In this study, we developed a semi-physical algorithm to retrieve high-resolution forest CI from GLAS full-waveform lidar data under a hypothesis of random distribution theory. First, we characterized the canopy gap fraction as a ratio between ground echoes and total echoes. Second, the vertical gap distribution within the canopy was derived from GLAS lidar data based on radiative transfer theory, which was used to calculate the canopy gap fraction corresponding to a random structure. Finally, the obtained canopy gap information from the above two steps was utilized as inputs to calculate CI based on the revised Beer–Lambert law. The performance of our method was evaluated by using ground-based measurements and MODIS CI products, and several factors affecting the accuracy of our method were analyzed.

2. Materials

2.1. Study Area

Our study areas are located in northern China, which includes two sites (Figure 1). One site is Heilongjiang Province, which is in northeast China with abundant forest resources. The major forest types include larch (*Larix gmelinii* (Rupr.) Rupr.) and birch (*Betula platyphylla* Suk). Heilongjiang is a mountainous area, and ground elevations are variable ranging from approximately 300~1000 m [33]. Another site is Saihanba National Forest Park, which is an important forest ground investigation site in China. Forest is mainly composed of larch and birch. Birch is a natural forest and usually grows on hillsides or mountain tops. Larch is a plantation forest, and most of them are planted in relatively flat areas.

2.2. GLAS Data

The GLAS sensor onboard the ICESat satellite platform is a lidar instrument that is widely used to monitor vegetation. According to the information provided by GLAS measurements, the NSIDC (National Snow and Ice Data Center) produced a total of 15 related products (GLA01-GLA15) for all users. In this study, the GLA01, GLA05, and GLA14 products, which were acquired from laser campaigns L2c (18 May 2004–21 June 2004), L3c (20 May 2005–23 June 2005), and L3f (24 May 2006–26 June 2006), were used to estimate CI; the key information extracted from GLAS products is shown in Table 1.
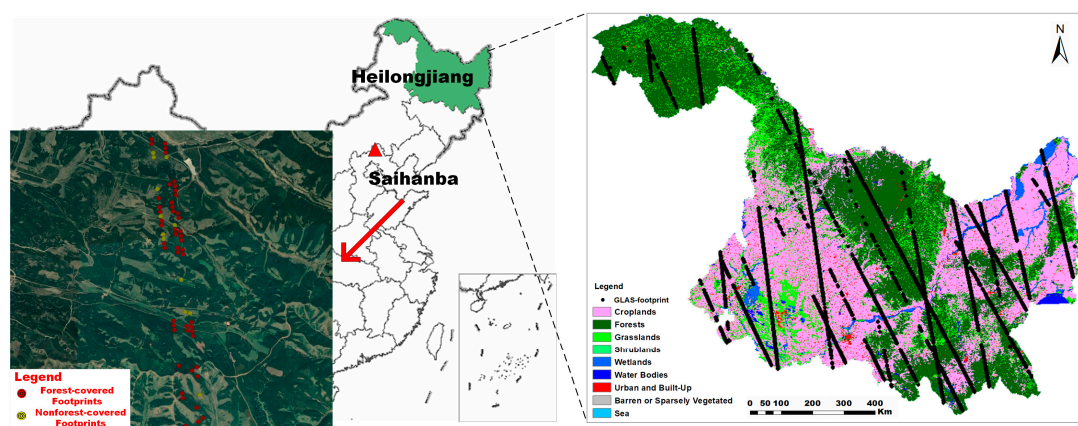
Figure 1. Geoscience Laser Altimeter System (GLAS) footprints distribution in study areas of Saihanba National Forest Park and Heilongjiang Province. The base map of Saihamba National Forest Park is the high-resolution image extracted from Google Maps. The base map of Heilongjiang Province is the ground surface landcover image provided by the GlobeLand30 landcover product.

Table 1. Key information extracted from GLAS products.

Product	Attributes	Record Name in GLAS File
GLA01 & GLA05 & GLA14	Record identification	i_rec_ndx
	count	i_shot_count
GLA01	Transmitted pulse waveform	r_tx_wf
	Received pulse waveform	r_rng_wf
	Starting address of the transmit pulse sample	i_TxWfStart
	Ending address of the range response	i_RespEndTime
GLA05	Gain value used for received pulse	i_gval_rcv
	Gain value used for transmitted pulse	i_gval_tx
GLA14	Latitude	i_lat
	Longitude	i_lon
	Standard deviation of background noise	i_sDevNsObl
	Maximum amplitude of signal	i_maxRecAmp
	Reflectivity	d_reflctUC
	Reflectivity correction factor for atmospheric effects	d_reflCor_atm

2.3. Ancillary Data

We used the 30-m spatial resolution globe landcover mapping product (GlobeLand30) as reference data to identify forest-covered GLAS data. The classification system of GlobeLand30 contains ten land cover types (Figure 1) with an accuracy of 83% on a global scale [34,35]. The accuracy of the vegetation structure parameters derived from lidar measurements is usually affected by terrain slope [36]. Therefore, such a terrain slope factor may introduce errors to our results. To analyze the effect of terrain slope on our method, the slope maps for two study areas were produced using void-filled 90 m resolution Shuttle Radar Topography Mission (SRTM) elevation data for global version 4 [37].

2.4. MODIS CI Products

We used two kinds of monthly MODIS CI products, which have the same temporal with GLAS data, over Heilongjiang Province as comparative data to validate our method. These two CI products were generated based on the method from Jiao et al. [17]—V005 and V006 versions of MCD43A1 BRDF products were used to capture clumping information, and the MODIS International Geosphere Biosphere Programme (IGBP) land-cover product (MCD12Q1) was used to determine vegetation types. In this study, we used MODIS CI with

good quality (i.e., quality label = 0 or 1) and corresponding to forest (i.e., IGBP = 1,2,3,4) as comparative data. Throughout this paper, we termed the MODIS CI generated from the V005 version of the MCD43A1 product as V005-MODIS-CI, and termed the MODIS CI generated from the V006 version of the MCD43A1 product as V006-MODIS-CI.
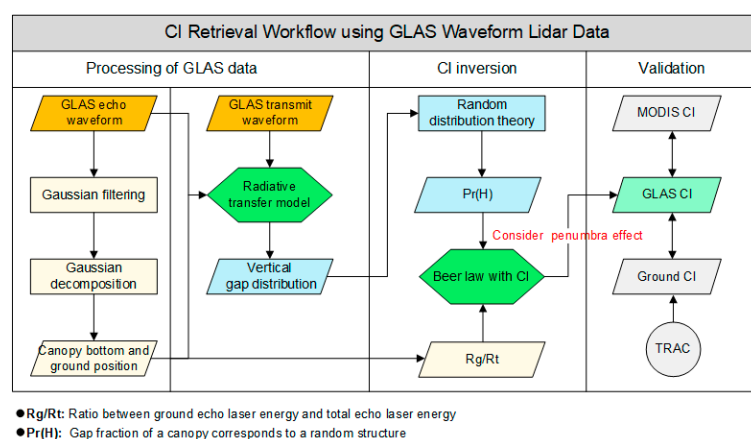
2.5. Data Used for Scale Effect Analysis

When conducting a comparison between MODIS CI and GLAS CI, the scale effect should not be ignored. Here, we used the collection 6 MODIS vegetation continuous fields (VCF) product and 30 m clear sky Landsat surface reflectance data to conduct scale effect analysis between MODIS CI and GLAS CI: (1) the VCF product provides the percentage of the three components that make up the pixel: percent tree cover, percent non-tree vegetation and percent bare [38]. We used the percentage of tree cover information to analyze MODIS pixel heterogeneity. (2) The clear sky Landsat surface reflectance data provides the fraction of incoming solar radiation that is reflected from the earth's surface to the Landsat sensor, it has been widely used as intermediaries to conduct spatial representativeness analysis [39,40]. Following previous research, we selected clear sky Landsat surface reflectance data as an intermediary to establish the spatial representativeness of GLAS measurements to assess the uncertainty caused by a direct comparison between GLAS CI and MODIS CI.

2.6. Ground-Based Data

CI from tracing radiation and architecture of canopies (TRAC) measurements is considered to be the most appropriate validation data source for CI retrievals from satellite data [41]. Therefore, we collected 50 ground-based CIs across the Saihanba National Forest Park by the TRAC in August 2018 to validate our method (Table A1). For each sampling, we first used a GPS equipment (i.e., Trimble GEO7X handheld GPS) to identify the GLAS footprint position with an accuracy of 1 cm (Figure 1). Then, we used the TRAC to obtain the GLAS footprint-scale CI based on a numerical gap-removal technique [42].

3. Methods

In this section, we elaborate on the framework of retrieving CI from the GLAS lidar data. Figure 2 shows the framework of this study, which includes three modules: processing of GLAS lidar data, GLAS CI inversion, and GLAS CI validation; details of each module are introduced in Sections 3.1–3.3.

●**Rg/Rt:** Ratio between ground echo laser energy and total echo laser energy
●**Pr(H):** Gap fraction of a canopy corresponds to a random structure

Figure 2. Workflow diagram for retrieving clumping index (CI) from GLAS data.

3.1. Processing of GLAS Data

3.1.1. Extraction of Canopy Bottom and Ground Position from GLAS Received Waveform

The canopy bottom and ground positions need to be identified from the GLAS received waveform to calculate CI using our method. To identify these two key positions, a Gaussian decomposition method was used to process the raw GLAS received waveform, which

included three steps. First, the raw GLAS received waveform was denoised by the Gaussian filtering method. Then, the Gaussian decomposition method was applied to decompose the denoised GLAS received waveform into multiple Gaussian components [43,44], as shown in Equation (1):

$$C(\text{t}) = \varepsilon + \sum_{i=1}^{n} A_i e^{\frac{-(h-h_i)^2}{2\sigma_i^2}} \tag{1}$$

where $C(\text{t})$ is the Gaussian component, h is the elevation, ε is determined by background noise, n is the number of Gaussian peaks, and A_i, h_i and σ_i are the amplitude, elevation position and waveform half width, respectively.

The third step is to extract the canopy bottom and ground positions. The decomposed Gaussian components represent the detected objects at different height levels. For forest measurements, the last decomposed component was generally considered to correspond to the ground surface, and the remaining components were canopy returns [45,46]. Therefore, we took the position of the last peak as the ground surface and took the end position of the last component of the canopy as the bottom of the canopy (Figure 3).
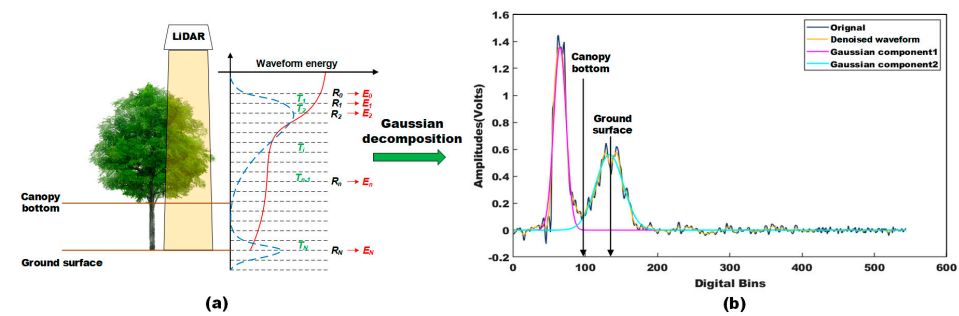
Figure 3. An example of processing the raw GLAS received waveform to obtain canopy bottom and ground positions by the Gaussian decomposition method. (**a**) is an example of GLAS observation for forest area; the blue dotted line is the GLAS received waveform (R_0, R_1 ... , R_n), and the red line is the transmitted energy profile (E_0, E_1, ... , E_n); T_i is the gap fraction between two successive recorded layers, which is equal to E_{i+1}/E_i. (**b**) is the result of Gaussian decomposition; the raw GLAS received waveform is shown by the blue line, the denoised waveform is shown by the orange line, and the decomposed canopy and ground components are shown in pink and cyan, respectively. The positions of two arrows represent the detected canopy bottom and the ground position, respectively.

3.1.2. Calculation of Forest Vertical Gap Distribution Using GLAS data

Forest vertical gap distribution is another important variable needed in our method, which will be used to calculate the gap fraction corresponding to a random distribution canopy. We calculated the forest vertical gap distribution from GLAS data based on the radiative transfer theory, which is shown as follows:

$$R_n = E_n[1 - \exp(-k * LAI_n)]\omega \cdot S \tag{2}$$

where ω is the reflectance of the detected material; k is the extinction coefficient; R_n is the lidar sensor received echo intensity reflected by record layer n; and E_n is the lidar sensor transmitted pulse energy intensity received by record layer n, which is given as Equation (3); S is a transform parameter, determined by the lidar sensor configuration and atmospheric transmittance. Appendix A describes the process for calculating S.

$$E_{n+1} = \exp(-k * LAI_{n-1})E_n \tag{3}$$

Using Equations (2) and (3), the relationship of the lidar transmitted pulse energy intensity between two successive record layers can be written as follows:

$$E_n = \frac{R_n}{S\omega} + E_{n+1} \tag{4}$$

Based on Equation (4), the lidar transmitted pulse energy intensity received by each recording layer (E_0, E_1, ... , E_n) can be expressed as follows:

$$PRE = \begin{pmatrix} E_0 = \frac{R_0}{S\omega} + E_1 \\ E_1 = \frac{R_1}{S\omega} + E_2 \\ \cdots \\ E_{n-1} = \frac{R_{n-1}}{S\omega} + E_n \end{pmatrix} \tag{5}$$

where *PRE* is the transmitted pulse energy intensity profile, E_0 is the lidar sensor emitted pulse energy intensity, and E_n is the transmitted pulse energy intensity received by the ground surface. The terms E_0, S and R_i ($i = 0, 1, 2, \ldots , n$) are known, as shown in Table 2. Therefore, if we know the reflectance ω, the lidar sensor transmitted pulse energy intensity received by each record layer (E_1, E_2, ... , E_n) can be calculated based on Equation (5).

Table 2. The instrument parameters and waveform information of the GLAS lidar.

Parameter	Symbol	Unit	Values
GLAS sensor configuration parameter	S	dimensionless	Calculation process of S is shown in Appendix A.
GLAS sensor emitted total pulse energy intensity	E_0	volts	E_0 is calculated by summing up the effective GLAS transmitted waveform recorded in the GLAS product (i.e., r_tx_wf).
GLAS sensor received echo intensity at each recorded layer	R_i	volts	R_i is provided by the GLAS received waveform recorded in the GLAS product (i.e., r_rng_wf).

To calculate ω, we first perform a summing operation for *PRE* and obtain the following results:

$$E_0 = \frac{\sum\limits_{i=0}^{n-1} R_i}{S\omega} + \frac{R_n}{S\omega_g} \tag{6}$$

where R_n is the echo intensity from the ground and ω_g is the ground reflectance. As shown in Equation (6), if we know the ground reflectance ω_g, the foliage reflectance ω can be obtained. Here, we adopted a reliable strategy to obtain ω_g, that is, using the mean value of the atmospherically corrected GLAS reflectance (i.e., GLAS reflectance = d_reflctUC * d_reflCor_atm) corresponding to no-tree covered footprints in the study area as ω_g. This strategy can make ω_g close to the actual reflectance of the GLAS laser for the ground. Finally, we obtained the atmospherically corrected ω_g, which is 0.21.

When obtaining the lidar transmitted pulse energy intensity profile (i.e., E_1, E_2, ... , E_n), the forest vertical gap distribution (i.e., T_1, T_2, ... , T_n) with a resolution of 0.15 m can be calculated by Equation (7):

$$T_i = \frac{E_{i+1}}{E_i} \tag{7}$$

where T_i is the gap fraction between two successive record layers (i.e., layer i and layer $i + 1$).

3.2. CI Inversion

We calculated the gap fraction corresponding to the random distribution canopy based on the vertical gap distribution of the canopy from GLAS observations. In GLAS observations, the GLAS sensor recorded the return energy intensity with a time step of 1 ns, which means that the thickness between two successive record layers is 0.15 m. In such a thin space, there is no mutual shading of foliage elements for the laser sensor. In other words, there is no clumping effect for foliage elements within two successive record layers. If the patterns of canopy elements in different elementary layers were independent, the gap fraction of the canopy corresponding to a random structure can be calculated by the following equation:

$$P_r = \prod_{i=1}^{l\_bottom} T_i \tag{8}$$

where *l_bottom* is the bottom of the canopy, which has been given in Section 3.1.1.

According to the Beer-Lambert law revised by Nilson [2], the CI is given as follows:

$$P(\theta) = \exp(-\Omega_E G(\theta) LAI / \cos(\theta)) \tag{9}$$

where $P(\theta)$ is the gap fraction in the θ direction of solar radiation; $G(\theta)$ is the projection coefficient of the foliage; and Ω_E is the CI. Since the energy emission direction of the GLAS sensor is close to zenith, the above formula can be simplified as follows:

$$P(0) = \exp(-\Omega_E G(0) LAI) \tag{10}$$

The ratio of the ground return energy to that of total energy added to a reflectance revised factor has been widely used as the gap fraction of the canopy [45,46], which is shown as follows:

$$P(0) = \frac{rG}{V + rG} \tag{11}$$

where G and V are the ground return energy and canopy return energy, respectively; r is the reflectance revised factor, which equals ω / ω_g in this study. Here, the canopy return energy was calculated using GLAS received bins above *l_bottom*, and the ground return energy was calculated using GLAS received bins less than *l_bottom*.

The gap fraction shown in Equation (8) is a result corresponding to a random structure, which means the clumping index $\Omega_E = 1$. Therefore, combining Equations (8) and (10) enables estimation of the clumping parameter Ω_E in a forest stand:

$$\Omega_E = \frac{\ln[P(0)]}{\ln[P_r]} \tag{12}$$

For the needleleaf forest, optical instruments such as lidar pulses (1064 nm) have difficulty measuring the needle shadow within the shoots because of the penumbra effect. Therefore, Ω_E characterizes the clumping effects at the shoot or leaf level. To characterize the smaller scale clumping effects for needleleaf forests, Chen et al. [47] used the needle-to-shoot area ratio γ measured in the laboratory for transformation. Here, we also adopted the same strategy, which is shown as follows:

$$\Omega = \frac{\Omega_E}{\gamma} \tag{13}$$

3.3. Validation

3.3.1. Comparison of GLAS CI and TRAC-Measured CI

TRAC-measured CI was used as a reference data to compare with GLAS CI to evaluate the performance of our method in Saihanba National Forest Park. The criteria for evaluation are the coefficient of determination (R^2), root-mean-square error (RMSE), and bias. When applying Equation (13), the needle-to-shoot area ratio γ needs to be set. Here, γ was set

according to a database collected from a review of the literature [17,41], which is shown in Table A1.

3.3.2. Comparison of GLAS CI and MODIS CI

We implemented our method in Heilongjiang province. Due to the limitation of the ground-based information in this study area, we set γ for needleleaf forest based on the dominant tree species (i.e., larch: $\gamma = 1.5$). Here, the needleleaf forest is identified by the MODIS IGBP classification system. To evaluate the performance of our method in Heilongjiang Province, we used two kinds of ~500 m MODIS CI products as the comparison data. It can be noted that there is a big difference in resolution between ~70 m GLAS footprint and ~500 m MODIS pixel. Vast differences in spatial resolution create several challenges for direct comparison, such as uncertainty caused by the scale effect.

To eliminate the scale effect as much as possible, we take some measures. First, we used the VCF data to select MODIS pixels with tree cover >60%, and then the corresponding CI was used as the comparative data. This measure ensures that the tree is the main cover within the MODIS pixel, thereby reducing the uncertain error introduced by the heterogeneity of the pixel. Second, we used an assessment method developed by Roman et al. [40] based on finer resolution satellite data (i.e., Landsat-TM images) to consider the spatial representativeness of GLAS observations compared to MODIS pixels. Following previous conclusions [48], we take the GLAS footprint with a still value $<5.0 \times 10^{-4}$ as the spatially representative site, which will be used as the comparison site. Figure 4 shows an example of the spatial representativeness analysis. We can note that the variogram estimators (point values) for these three spatial elements are more closely aligned within 500 m; the still values of three spatial elements (i.e., 1 km, 1.5 km, and 2 km) in this example site are 2.17×10^{-4}, 2.90×10^{-4}, and 2.97×10^{-4}, respectively, which indicate that the GLAS CI for this site is representative of the MODIS CI at ~500 m spatial resolution.
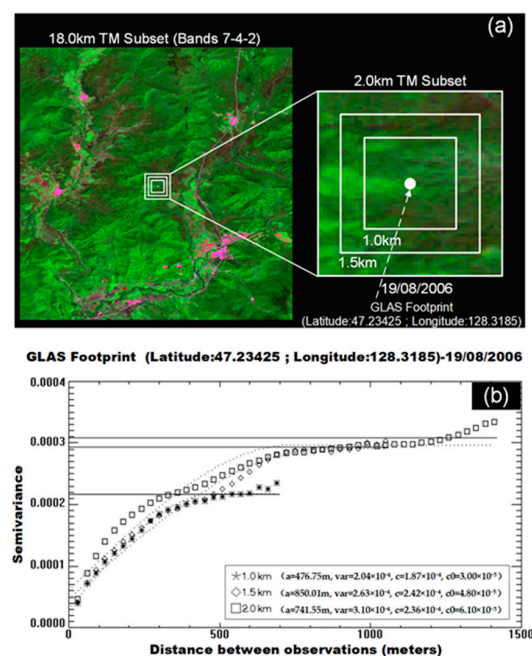

Figure 4. (**a**) Shortwave bidirectional reflectance factor (BRF) composites (Landsat-TM 7-4-2 band) centered on a GLAS footprint (lat: 47.23425; lon: 128.3185) in Heilongjiang, China, at ground level. Trees are indicated by the shades of green, bare areas are shown in pale-pink, and water is black. (**b**) Variogram estimators (points), spherical model (dotted curves), and sample variance (solid straight lines) obtained over this GLAS footprint using surface reflectance derived from the TM scene on 19 August 2006 with spatial thresholds of 1 km, 1.5 km and 2.0 km as a function of distance from the center of the GLAS observation.

4. Results and Analysis

4.1. GLAS CI vs. TRAC-Measured CI

There is a moderate agreement between GLAS CI and TRAC-measured CI (Figure 5), with R^2 = 0.34, RMSE = 0.12, and bias = 0.02. We can note that despite the low bias and low RMSE, the correlation between them is not very strong. The scatterplot shows some outliers in the comparison, which cause this low correlation. We therefore try to find the reason for these outliers.
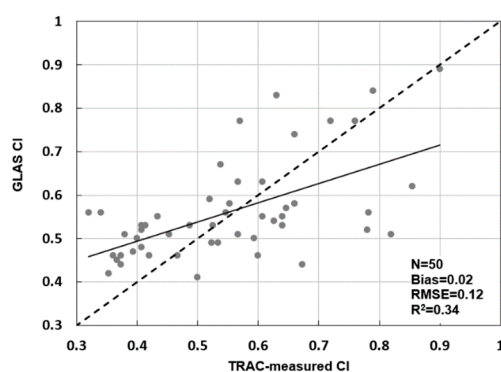

Figure 5. Evaluation of the GLAS CI using TRAC-measured CI. The solid line represents the linear fits of the TRAC-measured CI with GLAS CI.

The influence of terrain slope is a nonnegligible factor in lidar measurements, which has been well documented [49–51]. For example, the forest canopy and sloped ground can coalesce into a single broad return, which cause difficulties for data preprocessing [52]. To quantitatively investigate the role of the terrain slope on our method, we used the slope information derived from SRTM data to see how this factor influences the accuracy of GLAS CI.

Figure 6a shows the slope information of GLAS footprints in Saihanba National Forest Park. We can note that some GLAS observations are located in sloped areas, such as slope > 15° [45]. Figure 6b presents a GLAS waveform with a large slope (i.e., slope = 16°). It is obvious that the canopy echoes nearly merge with the ground echoes. Although we can obtain different Gaussian components through the data preprocessing step (Figure 6b), these Gaussian components may not represent the true canopy structure or ground information because of the influence of terrain slope [31]. For example, the Gaussian component which corresponding to the canopy structure is mixed with ground information due to the influence of large slope. The mixed information contained within the Gaussian component will easily lead to inaccurate canopy and ground position information in the preprocessing step. These inaccurate input parameters for our method will lead to the inaccurate canopy gap fraction result, which in turn lead to the inaccurate CI. We counted the variation of error with slope (Figure 6c); the error here means the difference between GLAS CI and TRAC-measured CI. The result indicated that GLAS CI and TRAC-measured CI was consistent well when the slope was less than 12°—almost all validation sites had errors less than 0.1, and the average error was 0.07. However, the error increases significantly when the slope is greater than 12°—the average error is 0.18. The above evidence indicates that terrain slope is an external factor that influences the accuracy of our method.

From Figure 6c, we note that there are two obvious outliers in comparison sites with gentle slopes (i.e., slope < 12°), which are site 31 (slope = 5°, error = 0.24) and site 4 (slope = 11°, error = 0.32). GLAS waveforms of these two sites provide reasons for these unideal results. As shown in Figure 7a, site 31 corresponds to a forest with a multilayer distribution canopy. In this type of forest, the lidar pulse interacts with multiple layers of different heights and shows multiple peaks in the GLAS waveform. Although this kind of multi-peak waveform does not have the problem of waveform broadening just like waveform at large slope, the characteristic of multiple peaks makes it more difficult to

process. Usually, the obtained information from multi-peak waveforms through automatic preprocessing methods, such as Gaussian decomposition method, cannot truly reflect the actual situation of the sample site, such as ground position [53]. Using inaccurate ground position information for our method will lead to an incorrect canopy gap fraction, which in turn will affect the final accuracy of GLAS CI.
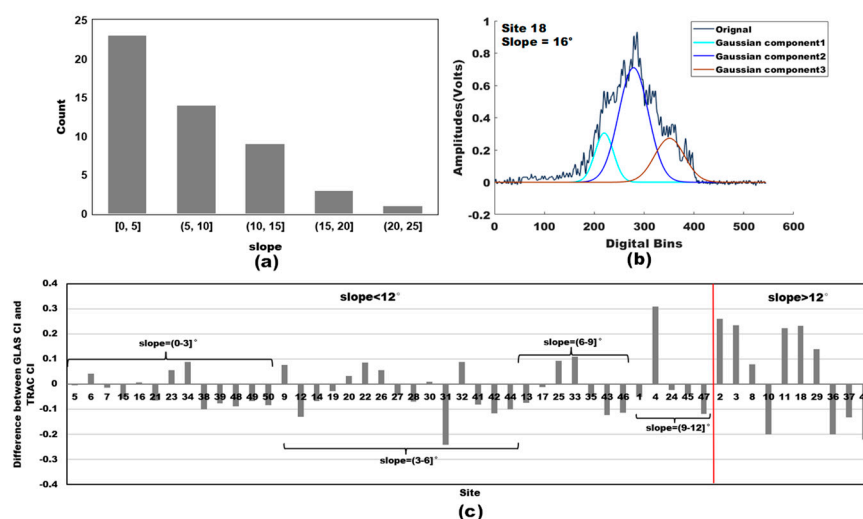

Figure 6. (**a**) Slope information of GLAS footprints over Saihanba National Forest Park. (**b**) An example of GLAS received waveform with a large slope (i.e., slope = 16°); three Gaussian components represent different detected objects. (**c**) The variation of error with slope; error means the difference between GLAS CI and TRAC-measured CI.
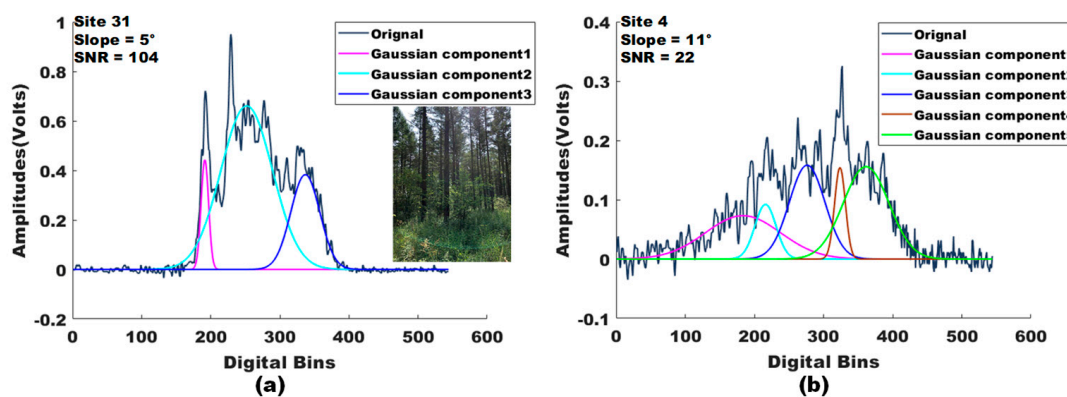
Figure 7. Two GLAS received waveforms corresponding to forest-covered footprints. (**a**) A GLAS received waveform corresponds to the forest with a multilayer distribution canopy. (**b**) A GLAS received waveform with a low signal-to-noise ratio.

Site 4 corresponds to the GLAS observation with a low signal-to-noise ratio (SNR = 22)—GLAS data have bad data quality. As shown in Figure 7b, the received waveform with a low SNR can no longer correctly reflect the canopy structure due to noise. Data quality determines the accuracy of results associated with remote sensing data processing and applications [54]. Therefore, the big bias at this site could potentially be explained by the noise factor. Once outliers, which are caused by terrain factor, data quality, and multilayer distribution canopy, are identified and removed from the validation dataset, the validation results have significantly improved with a determination coefficient of 0.72, RMSE of 0.07, and bias of 0.02 (Figure 8).
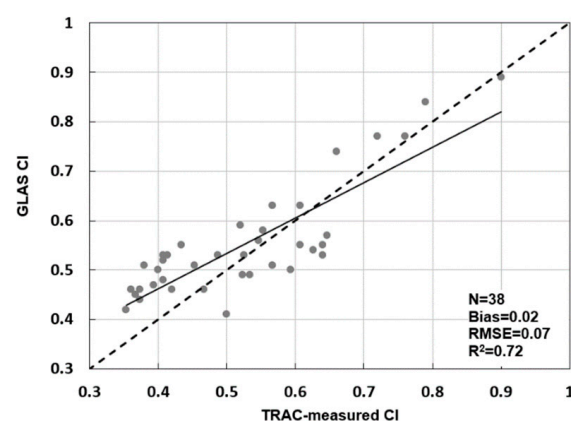
Figure 8. Evaluation of the GLAS CI derived from GLAS data with low slope and good data quality using TRAC-measured CI. The solid line represents the linear fits of the TRAC measurements with GLAS CI.

4.2. CI retrieval in Heilongjiang Province

We implemented our method in Heilongjiang Province to further check the reliability of our method. Here, we only used the GLAS data with slope < 12° as input for CI retrieval because the result in Section 4.1 showed that our method performed well based on the GLAS data with slope < 12°. Finally, a total of 1942 GLAS CIs for forests were obtained, and the distribution of GLAS CI was shown in Figure 9a. Note that there are some obviously unreasonable CI values, such as CI > 1 for forest. Our ground validation analysis in Section 4.1 indicated that CI derived from the GLAS data with low SNR was questionable and unreliable. Therefore, the SNR may be the reason for these unreasonable results. To prove our conjecture, we counted the SNR of used GLAS data (Figure 9b); the result showed that there were GLAS data with bad quality, such as GLAS data with SNR < 60 [55]. We further analyzed the relationship between unreasonable CI values and SNR; the result showed an exciting finding: these unreasonable CI values are closely related to SNR—the number of unreasonable CI values gradually decreases as the signal quality improves; when the SNR > 65, these unreasonable CI values disappear (Figure 9c). Therefore, we can reasonably consider that our method is suitable for GLAS data with SNR > 65.
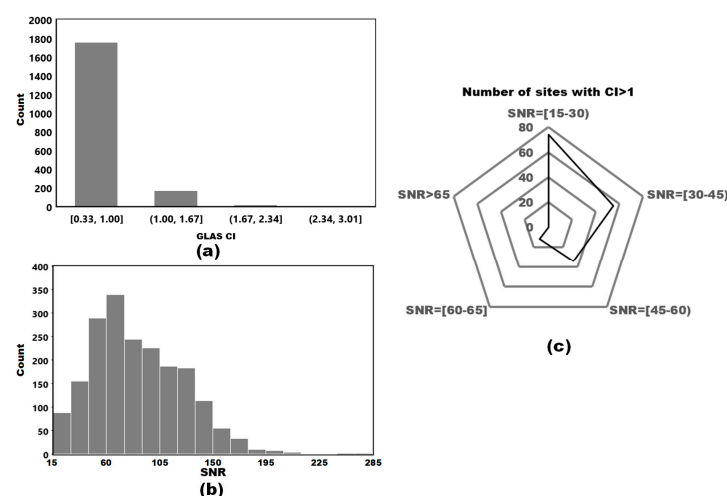
Figure 9. (**a**) The GLAS CI distribution in Heilongjiang Province. (**b**) The SNR (i.e., signal-to-noise ratio = i_maxRecAmp/i_sDevNsObl) distribution of GLAS data in Heilongjiang Province. (**c**) The relationship between unreasonable CI values and SNR, the number of unreasonable CI values gradually decreases as the signal quality improves; when SNR > 65, the unreasonable CI values disappear.

4.2.1. GLAS CI vs. MODIS CI

GLAS CI with low slope (i.e., slope < 12°) and good data quality (i.e., SNR > 65) was compared with two MODIS CI products with good quality (i.e., V005-MODIS-CI and V006-MODIS-CI). The result showed that GLAS CI was poorly correlated with V005-MODIS-CI (R^2 = 0.03), and V006-MODIS-CI (R^2 = 0.05) (Figure 10), although the root mean square error was relatively low (RMSE = 0.17 for V005-MODIS-CI and RMSE = 0.18 for V006-MODIS-CI).
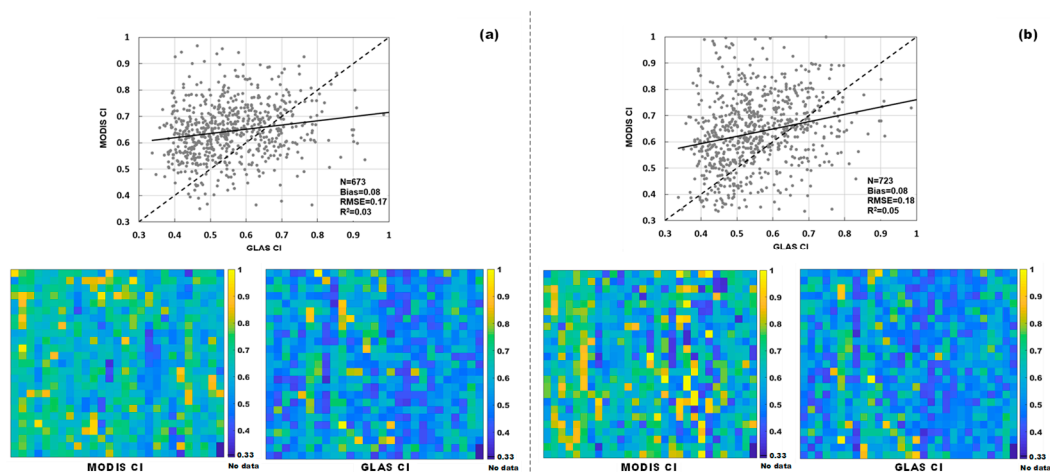
Figure 10. Scattering plots of the comparison between GLAS CI and MODIS CI and the corresponding CI distribution pattern. (**a**) is the comparison between GLAS CI and V005-MODIS-CI; (**b**) is the comparison between GLAS CI and V006-MODIS-CI.

In terms of resolution, GLAS CI and MODIS CI have vast differences, so the direct comparison may have uncertainty due to the scale effect, such as uncertainty caused by pixel heterogeneity. Tree cover information, provided by 250 m MODIS VCF data, showed that MODIS CIs correspond to the MODIS pixels with tree coverage between 40% and 80% (Figure 11). Due to the object described by GLAS CI is forest; therefore, the scale effect caused by pixel heterogeneity existed, which should lead to inconsistency between GLAS CI and MODIS CI.
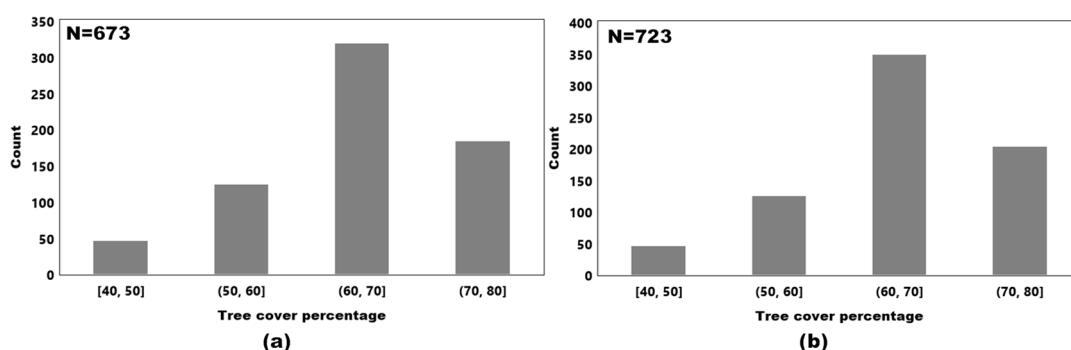

Figure 11. (**a**,**b**) are the distribution of tree cover percentage of MODIS pixel corresponding to the used V005-MODIS-CI and V006-MODIS-CI, respectively.

To minimize uncertainty in the comparison between GLAS CI and MODIS CI caused by the scale effect, we take two measures, as introduced in Section 3.3.2, to identify inappropriate comparison data and remove them. Here, the inappropriate comparison data means that MODIS CI from the MODIS pixel with tree coverage less than 60%, GLAS CI from the site without spatial representativeness. We then used the remaining data to conduct the comparison analysis, and it could be noted that some obvious outliers disappeared; however, the comparison accuracy was not greatly improved (Figure 12): the result be-

tween GLAS CI and V005-MODIS-CI is R^2 = 0.09, RMSE = 0.15, bias = 0.07, and the result between GLAS CI and V006-MODIS-CI is R^2 = 0.26, RMSE = 0.12, bias = 0.04. Interestingly, the V006-MODIS-CI has better consistency with the GLAS CI than the V005 MODIS-CI.
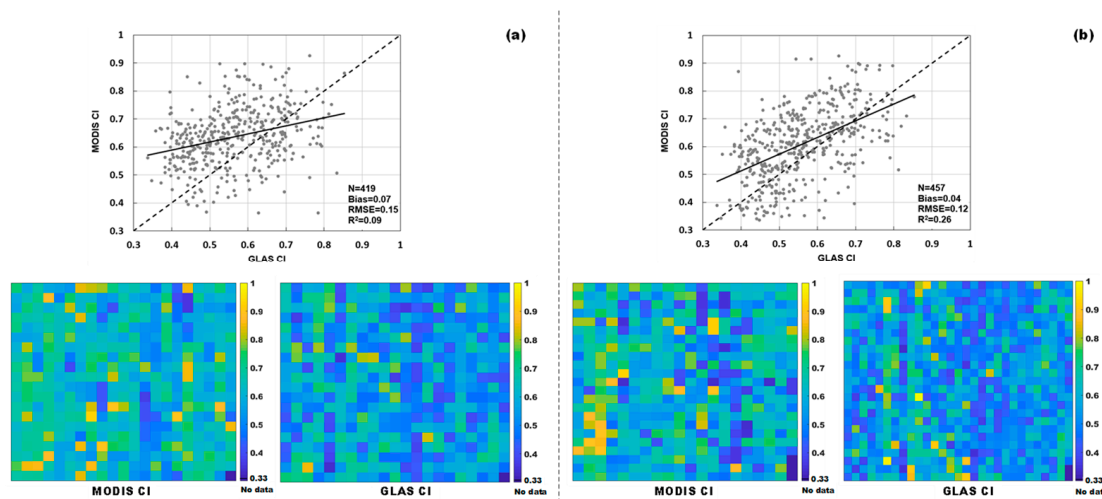
Figure 12. Scattering plots of the comparison between GLAS CI and MODIS CI and the corresponding CI distribution pattern, which are the results after removing the uncertainty comparison data caused by scale effect. (**a**) is the comparison between GLAS CI and V005-MODIS-CI; (**b**) is the comparison between GLAS CI and V006-MODIS-CI.

Ideally, the MODIS CI corresponding to MODIS pixels with 100% forest coverage is the most suitable comparison data for GLAS CI; however, such pixels are nonexistent in our study area. Figure 13 shows two typical MODIS pixels with tree coverage greater than 60%, and the corresponding CI was used as comparison data. We can note that there are elements such as grass and bare land within the MODIS pixels in addition to trees. These non-forest elements should be one reason for the inconsistency between the GLAS CI and MODIS CI because the GLAS CI describes the clumpiness of the forest canopy.


Figure 13. Two typical MODIS pixels with tree coverage greater than 60% and CI obtained from these two pixels were used as comparative data for GLAS CI. (**a**) is a MODIS pixel with tree coverage of 73%; (**b**) is a MODIS pixel with tree coverage of 61%. It can be noted that there are elements such as grass and bare land in these pixels in addition to trees.

4.3. Model Uncertainty Analysis

The ground reflectance is an essential input parameter in our method, but due to the limitation of observation conditions, we cannot give absolutely correct ground reflectance for each GLAS observation—incorrect ground reflectance may introduce error to the output CI. To see what role the ground reflectance played in our method, we conducted a sensitivity analysis between input ground reflectance and output CI: we varied ground reflectance from 0.18 to 0.24 to evaluate its effect on CI (Figure 14). We found that CI range variation was less than 0.1 for a moderate CI (0.6–0.7). Recall that the ground validation for our

method had RMSE values of 0.07; thus, it is possible that much of our average error could be explained by spatial variation in ground reflectance.
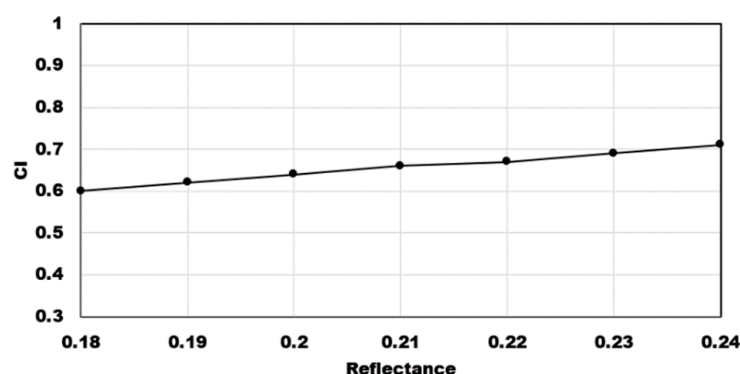
Figure 14. An example of CI-derived GLAS data with different input ground reflectances. As the ground reflectance increases, the CI value increases.

5. Discussion

The spaceborne lidar system GLAS can capture detailed canopy gap information on a global scale; thus, GLAS lidar data may be an ideal data source for estimating CI globally. However, no one has tried CI estimation using GLAS data at present. This study developed a method to estimate CI using GLAS data, which is the first time to test the feasibility of deriving CI from spaceborne lidar data. In addition, our method is physically based, not statistically based regression methods, as is commonly done. More importantly, our method provided an opportunity to retrieve high-resolution CI globally.

TRAC measurements were used as a standard value to validate the performance of GLAS CI. However, there is a significant difference between GLAS CI and TRAC CI—GLAS CI from vertical gap information, TRAC CI from horizontal gap information. Although this difference exists between them, these two CI products show a strong correlation, which should be closely related to the characteristics of CI. CI is a vegetation structure parameter that describes the light distribution within the canopy, which is determined by the gap distribution within the canopy [42]. The gap information used by TRAC is obtained under the canopy along the horizontal direction, which can be taken as the accumulation of the vertical gap. Therefore, the CI calculated based on these two kinds of gap distribution information should have the same function.

The different acquisition times of GLAS data and TRAC measurements is a major limitation in our ground validation. Despite this limitation, our validation is still valid because our ground validation is to evaluate the ability of GLAS lidar to capture spatial patterns and heterogeneity in CI. Here, the spatial variability, which is a signature that should be independent of time, contributed to most of the variation in both the in situ and GLAS CI. Therefore, our validation should provide useful quantification of the GLAS accuracies over space.

We conducted our method in Heilongjiang Province as well. Due to the lack of measured data in Heilongjiang Province, we used MODIS CI products provided by Jiao et al. [17], which are one of the most widely used global CI products, as validation data. The comparison results show a not good agreement. There are some possible reasons for the inconsistency between GLAS CI and MODIS CI. First is the influence of the scale effect; although we have taken some measures to eliminate the impact of the scale effect as much as possible, its influence still exists, such as problem caused by pixel heterogeneity as described in Section 4.2.1. Incorrect needle-to-shoot area ratio setting for GLAS CI should be another reason for the inconsistency between GLAS CI and MODIS CI. The GLAS CI for needleleaf forests needs to set the needle-to-shoot area ratio in our method. In general, the needle-to-shoot area ratio was measured in the lab. However, it is challenging to conduct needle-to-shoot area ratio measurements at a provincial scale due to workforce and cost.

Here, we used the needle-to-shoot area ratio of dominant tree species in Heilongjiang Province as the input for needleleaf forest over the entire province. Although this strategy could ensure the accuracy of the needle-to-shoot area ratio setting in most validation sites, there would be inaccurate situations in some sites, which in turn introduces uncertainty in the comparison between GLAS CI and MODIS CI.

One interesting finding in the comparison between GLAS CI and MODIS CI is that GLAS CI has better consistency with V006-MODIS-CI than V005-MODIS-CI (R^2 = 0.25 vs. 0.09). The quality of MODIS CI products is mainly governed by the quality of the input MCD43A1 BRDF product. The Collection V006 MCD43 BRDF product has a better performance than the previous Collection V005 product due to the use of the improved MODIS reflectance, cloud masking, and an enhanced daily inversion algorithm [56]. Therefore, the quality of MODIS CI, which is produced by the Collection V006 MCD43 BRDF product, is theoretically better than that from the Collection V005 MCD43 BRDF product. A previous study reinforced our conjecture that the MODIS CI derived from the latest Collection V006 MCD43A1 BRDF daily product has better agreement with the field measurements than the MODIS CI from the Collection V005 MCD43A1 BRDF product [15]. Therefore, the comparison results between GLAS CI and MODIS CI can illustrate the reliability of our proposed CI inversion method to certain extents.

At present, although there are several physical-based methods for CI estimation that have been developed and widely used, such as CC, LX, and CLX, existing studies show these methods have some limitations. The CC method does not consider the problem of inconsistent paths when light passes through the canopy, which may lead to errors in CI estimation [42]. The LX method is through averaging logarithms of gap probabilities over segments of finite length to estimate CI. However, there is a major limitation in the averaging process in LX method, that is, ignore the clumping effect within the segment [57]. To address the problem in LX method, a new method combining CC and LX method called CLX was proposed. In the CLX method, the use of CC method can solve problem of non-randomness within segment in LX method to certain extents [58]. However, some studies found that CLX method was sensitive to the choice of segment length, which had an impact on the CI estimation [59]. Our proposed method utilizes the gap information in the height direction extracted from the full-waveform lidar data, which can avoid problem of inconsistent paths. In addition, our method doesn't need to consider the choice of segment length. Despite there are uncertainties in our method, such as mentioned in Section 4.3, it is obviously that our new method can avoid limitations of existing methods, which provides a new perspective for improving the accuracy of CI estimation. Furthermore, our method is developed based on the full-waveform lidar data, which gives our method an opportunity to be applied to new full-waveform spaceborne lidar equipment in the future, such as the latest generation of spaceborne lidar equipment Global Ecosystem Dynamics Investigation (GEDI) [32]. The use of the latest spaceborne lidar equipment observation provides unprecedented opportunities to improve the accuracy of global scale CI estimation.

6. Conclusions

As surface process modeling becomes complex and refined, there is an urgent need for detailed information about canopy structure, such as canopy clumping information. However, no appropriate high-resolution multiangle reflectance satellite data are currently available for traditional methods to produce such products. In this study, we developed a method for CI estimation from GLAS lidar data, which fills the theoretical gap of using full-waveform lidar data to estimate CI and provides an opportunity to use spaceborne full-waveform lidar data to produce high-resolution CI globally. We concluded based on our findings that our method could accurately retrieve forest canopy CI using GLAS data with gentle slope (e.g., <12°) and high signal-to-noise ratio (e.g., >65). However, our GLAS CI from high-quality waveforms showed less correlation with MODIS CI (e.g., R^2 = 0.26, RMSE = 0.12, and bias = 0.04): The difference highlights the impact of the scale effect in

conducting comparisons of products with huge differences resolution. Due to the simple configuration and the feasibility of operational applications, our proposed method can be a basis for retrieving continental or global CI datasets from spaceborne full-waveform lidar systems, such as GLAS or GEDI. Such a dataset, once available, will not only be an important input in modeling land surface processes and terrestrial carbon cycle but can also serve as reliable reference data for validating moderate-resolution CI products over a large scale. Despite several limitations in our method, the results point toward significant opportunities ahead.

Author Contributions: Conceptualization, L.C., Z.J. and K.Z.; data curation, L.C., Y.T., S.Y., Z.Z. and X.Z.; formal analysis, L.C., Z.J., J.G., S.L. and R.X.; funding acquisition, Z.J. and Y.D.; software, L.C. and M.S.; writing—original draft, L.C.; and writing—review and editing, Z.J., K.Z., M.S. and Y.D. All authors have read and agreed to the published version of the manuscript.

Informed Consent Statement: Not applicable.

Data Availability Statement: All satellite remote sensing data used in this study are openly and freely available. The GLAS data is available at https://nsidc.org/data/icesat (accessed on 20 January 2020). The GlobeLand30 data is available at http://www.globallandcover.com (accessed on 20 January 2020). The SRTM data is available at http://srtm.csi.cgiar.org/srtmdata/ (accessed on 20 January 2020). The MODIS BRDF product (MCD43A1), MODIS VCF product (MOD44B), and MODIS landcover product (MCD12Q1) are available at https://search.earthdata.nasa.gov/search (accessed on 20 January 2020). The clear sky Landsat surface reflectance data is available at https://earthexplorer.usgs.gov/ (accessed on 20 January 2020).

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

Appendix A

The GLAS sensor received pulse energy is determined by the sensor transmitted pulse energy, ground surface optical properties and atmospheric transmittance, and the corresponding mathematical formula is shown as follows [60]:

$$E_{echo} = E_{tx}\tau_{opt}\frac{A_{telescope}}{R^2} \cdot \frac{\rho_{sur}}{\Omega}\tau_{atm} \tag{A1}$$

where E_{echo} is the sensor received pulse energy, E_{tx} is the transmitted pulse energy, $\tau_{opt} = 0.67$ is the optics transmission, $A_{telescope} = 0.709$ m^2 is the sensor telescope area, and R is the sensor altitude [R = (i_RespEndTime − i_TxWfStart) · C/2, where C is the light transmission speed.], ρ_{sur} is the surface reflectance, τ_{atm} is the roundtrip atmosphere transmission (i.e., d_reflCor_atm), and $\Omega = \pi$ is the scattering solid angle.

The relationship between the pulse energy and digitized pulse waveforms is given as follows:

$$E_g = \frac{\alpha}{\eta_e\eta R_{Det}G}\sum_{i=1}^{N} W(\text{i})\Delta t \tag{A2}$$

where E_g represents the pulse energy (i.e., transmitted or received pulse energy), α is the calibration parameter ($\alpha_{trans} = 1.21$ is for transmitted laser pulse; $\alpha_{rec} = 1.00$ for the return laser pulse), $\eta_e = 0.923$ is the electronic throughput, η is the optical throughput (return laser pulse: $\eta_{rec} = 0.67$; transmitted laser pulse: $\eta_{trans} = 2.97 \times 10^{-4}$ is for laser 1, $\eta_{trans} = 2.79 \times 10^{-4}$ is for laser 2, $\eta_{trans} = 2.79 \times 10^{-4}$ is for laser 3), $R_{Det} = 2.28 \times 10^7$ V/W is the detector responsibility, G is the gain value for GLAS sensor (received pulse: G_{rec} = i_gval_rcv/255; transmitted pulse: G_{trans} = i_gval_tx/255), $W(\text{i})$ is the pulse waveform in volts, N is the digital bins of the waveform, and $\Delta t = 1$ ns is the sampling time interval.

Combining Equations (A1) and (A2), we can obtain a relationship between the lidar sensor received digitized pulse waveform and transmitted digitized pulse waveform, which is written as follows:

$$E_{rec-waveform} = E_{trans-waveform} \cdot \frac{A_{telescope} \cdot \tau_{opt}\tau_{atm}}{\Omega R^2} \cdot \frac{\alpha_{trans}\eta_{rec}G_{rec}}{\alpha_{rec}\eta_{trans}G_{trans}} \cdot \rho_{sur} \tag{A3}$$

where $E_{rec-waveform}$ is the received pulse waveform and $E_{trans-waveform}$ is the transmitted pulse waveform. According to Equation (A3), a parameter S can be defined; S is determined by the lidar sensor configuration and atmospheric transmittance, which is shown as Equation (A4). As all the variables are known for Equation (A4); therefore, S for each GLAS laser can be calculated. S is a necessary parameter to understand the characteristics of the observed surface reflectivity based on the lidar digitized pulse waveform.

$$S = \frac{A \cdot \tau_{opt}\tau_{atm}}{\Omega R^2} \cdot \frac{\alpha_{trans}\eta_{rec}G_{rec}}{\alpha_{rec}\eta_{trans}G_{trans}} \tag{A4}$$

Appendix B

Table A1. Characteristics of 50 validation sites. GLAS ID is the index of GLAS footprint in the product file; vegetation types are from the field investigation; γ is the needle-to-shoot area ratio; and $\Omega_e(\theta)$ denotes the clumpiness of foliage elements.

| | GLAS ID | | | | | | | |
Site	i_rec_ndx	i_shot_cout	Latitude	Longitude	Vegetation Types	γ	$\Omega_e(\theta)$	Field CI
1	219109634	5	42.357381	117.35471	birch	1	0.77	0.77
2	219109634	11	42.366674	117.35291	birch	1	0.52	0.52
3	219109634	12	42.368221	117.35260	birch	1	0.62	0.62
4	219109634	18	42.377491	117.35078	birch	1	0.51	0.51
5	219109634	26	42.389872	117.34837	larch	1.5	0.8	0.53
6	219109634	27	42.391422	117.34806	larch	1.5	0.74	0.49
7	219109634	28	42.392973	117.34776	larch	1.5	0.84	0.56
8	219109639	4	42.419341	117.34263	larch	1.5	0.87	0.58
9	219109639	5	42.417794	117.34293	larch	1.5	0.86	0.57
10	219109639	9	42.427073	117.34111	birch	1	0.77	0.77
11	219109639	10	42.425525	117.34141	birch	1	0.56	0.56
12	219109639	12	42.430165	117.3405	larch	1.5	0.77	0.51
13	219109639	13	42.431711	117.3402	larch	1.5	0.72	0.48
14	219109639	15	42.434803	117.33959	larch	1.5	0.66	0.44
15	219109639	27	42.453392	117.33596	larch	1.5	0.77	0.51
16	219109639	28	42.454943	117.33566	larch	1.5	0.69	0.46
17	377874826	1	42.36017	117.36226	birch	1	0.77	0.77
18	377874826	10	42.37410	117.35957	larch	1.5	0.66	0.44
19	377874826	12	42.37718	117.35896	larch	1.5	0.87	0.58
20	377874826	19	42.387999	117.35687	larch	1.5	0.74	0.49
21	377874826	20	42.389544	117.35657	larch	1.5	0.95	0.63
22	377874826	21	42.391089	117.35627	larch	1.5	0.81	0.54
23	377874826	22	42.392634	117.35597	larch	1.5	0.77	0.51
24	377874826	37	42.415832	117.35146	larch	1.5	0.95	0.63
25	377874826	38	42.41738	117.35115	larch	1.5	0.75	0.50

Table A1. *Cont.*

	GLAS ID		Latitude	Longitude	Vegetation Types	γ	$\Omega_e(\theta)$	Field CI
Site	i_rec_ndx	i_shot_cout						
26	377874826	40	42.420475	117.35055	larch	1.5	0.83	0.55
27	377874831	1	42.422022	117.35024	larch	1.5	0.8	0.53
28	377874831	2	42.423569	117.34994	larch	1.5	0.89	0.59
29	377874831	5	42.428203	117.34904	larch	1.5	0.69	0.46
30	377874831	6	42.42975	117.34874	birch	1	0.89	0.89
31	377874831	10	42.435944	117.347562	larch	1.5	0.84	0.56
32	377874831	11	42.437491	117.34726	larch	1.5	0.62	0.41
33	377874831	12	42.439036	117.34696	larch	1.5	0.8	0.53
34	377874831	13	42.44058	117.34666	larch	1.5	0.82	0.55
35	537071457	1	42.358912	117.36027	birch	1	0.84	0.84
36	537071457	4	42.363587	117.35937	birch	1	0.83	0.83
37	537071457	12	42.375961	117.356948	mixed forest	1.43	0.96	0.67
38	537071457	21	42.389854	117.35421	larch	1.5	0.75	0.50
39	537071457	22	42.391403	117.35391	larch	1.5	0.7	0.47
40	537071457	38	42.416399	117.3491	larch	1.5	0.84	0.56
41	537071457	39	42.41796	117.3488	birch	1	0.74	0.74
42	537071462	2	42.422629	117.34788	mixed forest	1.43	0.79	0.55
43	537071462	3	42.424179	117.34758	larch	1.5	0.8	0.53
44	537071462	7	42.430353	117.34636	larch	1.5	0.69	0.46
45	537071462	8	42.431895	117.34606	larch	1.5	0.69	0.46
46	537071462	9	42.433434	117.34576	larch	1.5	0.78	0.52
47	537071462	13	42.439595	117.34455	larch	1.5	0.8	0.53
48	537071462	20	42.450446	117.34242	larch	1.5	0.69	0.46
49	537071462	21	42.452004	117.34212	larch	1.5	0.63	0.42
50	537071462	22	42.453564	117.34182	larch	1.5	0.68	0.45

References

1. Chen, J.M.; Black, T.A. Foliage Area and Architecture of Plant Canopies from Sunfleck Size Disributions. *Agric. Forest Meteorol.* **1992**, *60*, 249–266. [CrossRef]
2. Nilson, T. Theoretical Analysis of Frequency of Gaps in Plant Stands. *Agric. Meteorol.* **1971**, *8*, 25. [CrossRef]
3. Xiao, Z.; Liang, S.; Wang, J.; Chen, P.; Yin, X.; Zhang, L.; Song, J. Use of General Regression Neural Networks for Generating the Glass Leaf Area Index Product from Time-Series Modis Surface Reflectance. *IEEE Trans. Geosci. Remote* **2014**, *52*, 209–223. [CrossRef]
4. Deng, F.; Chen, J.M.; Plummer, S.; Chen, M.; Pisek, J. Algorithm for Global Leaf Area Index Retrieval Using Satellite Imagery. *IEEE Trans. Geosci. Remote* **2006**, *44*, 2219–2229. [CrossRef]
5. Liu, Y.; Liu, R.; Chen, J.M. Retrospective Retrieval of Long-Term Consistent Global Leaf Area Index (1981–2011) From Combined Avhrr and Modis Data. *J. Geophys. Res. Biogeosci.* **2012**, *117*, G04003. [CrossRef]
6. Chen, J.M.; Liu, J.; Cihlar, J.; Goulden, M.L. Daily Canopy Photosynthesis Model through Temporal and Spatial Scaling for Remote Sensing Applications. *Ecol. Model.* **1999**, *124*, 99–119. [CrossRef]
7. He, L.; Chen, J.M.; Gonsamo, A.; Luo, X.; Wang, R.; Liu, Y.; Liu, R. Changes in the Shadow: The Shifting Role of Shaded Leaves in Global Carbon and Water Cycles Under Climate Change. *Geophys. Res. Lett.* **2018**, *45*, 5052–5061. [CrossRef]
8. Ryu, Y.; Baldocchi, D.D.; Kobayashi, H.; van Ingen, C.; Li, J.; Black, T.A.; Beringer, J.; van Gorsel, E.; Knohl, A.; Law, B.E.; et al. Integration of Modis Land and Atmosphere Products with a Coupled-Process Model to Estimate Gross Primary Productivity and Evapotranspiration From 1 Km to Global Scales. *Glob. Biogeochem. Cycles* **2011**, *25*, GB4017. [CrossRef]

9. Chen, J.M.; Mo, G.; Pisek, J.; Liu, J.; Deng, F.; Ishizawa, M.; Chan, D. Effects of Foliage Clumping on the Estimation of Global Terrestrial Gross Primary Productivity. *Glob. Biogeochem. Cycles* **2012**, *26*, 1–18.

10. Chen, B.; Liu, J.; Chen, J.M.; Croft, H.; Gonsamo, A.; He, L.; Luo, X. Assessment of Foliage Clumping Effects on Evapotranspiration Estimates in Forested Ecosystems. *Agric. Forest Meteorol.* **2016**, *216*, 82–92. [CrossRef]

11. Chen, J.M.; Menges, C.H.; Leblanc, S.G. Global Mapping of Foliage Clumping Index Using Multi-Angular Satellite Data. *Remote Sens. Environ.* **2005**, *97*, 447–457. [CrossRef]

12. Chen, J.M.; Leblanc, S.G. A Four-Scale Bidirectional Reflectance Model Based on Canopy Architecture. *IEEE Trans. Geosci. Remote* **1997**, *35*, 1316–1337. [CrossRef]

13. Jiao, Z.; Schaaf, C.B.; Dong, Y.; Román, M.; Hill, M.J.; Chen, J.M.; Wang, Z.; Zhang, H.; Saenz, E.; Poudyal, R.; et al. A Method for Improving Hotspot Directional Signatures in Brdf Models Used for Modis. *Remote Sens. Environ.* **2016**, *186*, 135–151. [CrossRef]

14. Pisek, J.; Chen, J.M.; Lacaze, R.; Sonnentag, O.; Alikas, K. Expanding Global Mapping of the Foliage Clumping Index with Multi-Angular Polder Three Measurements: Evaluation and Topographic Compensation. *ISPRS J. Photogramm.* **2010**, *65*, 341–346. [CrossRef]

15. Wei, S.; Fang, H.; Schaaf, C.B.; He, L.; Chen, J.M. Global 500 M Clumping Index Product Derived from Modis Brdf Data (2001–2017). *Remote Sens. Environ.* **2019**, *232*, 111296. [CrossRef]

16. He, L.; Chen, J.M.; Pisek, J.; Schaaf, C.B.; Strahler, A.H. Global Clumping Index Map Derived from the Modis Brdf Product. *Remote Sens. Environ.* **2012**, *119*, 118–130. [CrossRef]

17. Jiao, Z.; Dong, Y.; Schaaf, C.B.; Chen, J.M.; Román, M.; Wang, Z.; Zhang, H.; Ding, A.; Erb, A.; Hill, M.J.; et al. An Algorithm for the Retrieval of the Clumping Index (Ci) From the Modis Brdf Product Using an Adjusted Version of the Kernel-Driven Brdf Model. *Remote Sens. Environ.* **2018**, *209*, 594–611. [CrossRef]

18. Pisek, J.; Ryu, Y.; Sprintsin, M.; He, L.; Oliphant, A.J.; Korhonen, L.; Kuusk, J.; Kuusk, A.; Bergstrom, R.; Verrelst, J.; et al. Retrieving Vegetation Clumping Index from Multi-Angle Imaging Spectroradiometer (Misr) Data at 275M Resolution. *Remote Sens. Environ.* **2013**, *138*, 126–133. [CrossRef]

19. Wei, S.; Fang, H. Estimation of Canopy Clumping Index from Misr and Modis Sensors Using the Normalized Difference Hotspot and Darkspot (Ndhd) Method: The Influence of Brdf Models and Solar Zenith Angle. *Remote Sens. Environ.* **2016**, *187*, 476–491. [CrossRef]

20. Maignan, F.; Breon, F.M.; Lacaze, R. Bidirectional Reflectance of Earth Targets: Evaluation of Analytical Models Using a Large Set of Spaceborne Measurements with Emphasis on the Hot Spot. *Remote Sens. Environ.* **2004**, *90*, 210–220. [CrossRef]

21. Brovelli, M.A.; Crespi, M.; Fratarcangeli, F.; Giannone, F.; Realini, E. Accuracy Assessment of High Resolution Satellite Imagery Orientation by Leave-One-Out Method (Article). *ISPRS J. Photogramm.* **2008**, *63*, 427–440. [CrossRef]

22. Goetz, S.J. Synergistic Use of Spaceborne Lidar and Optical Imagery for Assessing Forest Disturbance: An Alaska Case Study. *J. Geophys. Res. Biogeosci.* **2010**, *115*, G00E07. [CrossRef]

23. Mitchell, J.J.; Shrestha, R.; Spaete, L.P.; Glenn, N.F. Combining Airborne Hyperspectral and Lidar Data Across Local Sites for Upscaling Shrubland Structural Information: Lessons for Hyspiri. *Remote Sens. Environ.* **2015**, *167*, 98–110. [CrossRef]

24. García, M.; Gajardo, J.; Riaño, D.; Zhao, K.; Martín, P.; Ustin, S. Canopy Clumping Appraisal Using Terrestrial and Airborne Laser Scanning. *Remote Sens. Environ.* **2015**, *161*, 78–88. [CrossRef]

25. Li, Y.; Guo, Q.; Su, Y.; Tao, S.; Zhao, K.; Xu, G. Retrieving the Gap Fraction, Element Clumping Index, and Leaf Area Index of Individual Trees Using Single-Scan Data from a Terrestrial Laser Scanner. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 308–316. [CrossRef]

26. Ma, L.; Zheng, G.; Wang, X.; Li, S.; Lin, Y.; Ju, W. Retrieving Forest Canopy Clumping Index Using Terrestrial Laser Scanning Data. *Remote Sens. Environ.* **2018**, *210*, 452–472. [CrossRef]

27. Zhao, K.; Ryu, Y.; Hu, T.; Garcia, M.; Li, Y.; Liu, Z.; Londo, A.; Wang, C.; Chisholm, R. How to Better Estimate Leaf Area Index and Leaf Angle Distribution from Digital Hemispherical Photography? Switching to a Binary Nonlinear Regression Paradigm. *Methods Ecol. Evol.* **2019**, *10*, 1864–1874. [CrossRef]

28. Thomas, V.T.V.E.; Noland, T.; Treitz, P.; McCaughey, J.H. Leaf Area and Clumping Indices for a Boreal Mixed-Wood Forest: Lidar, Hyperspectral, and Landsat Models. *Int. J. Remote Sens.* **2011**, *32*, 8271–8297. [CrossRef]

29. Zhao, F.; Strahler, A.H.; Schaaf, C.L.; Yao, T.; Yang, X.; Wang, Z.; Schull, M.A.; Román, M.O.; Woodcock, C.E.; Olofsson, P.; et al. Measuring Gap Fraction, Element Clumping Index and Lai in Sierra Forest Stands Using a Full-Waveform Ground-Based Lidar. *Remote Sens. Environ.* **2012**, *125*, 73–79. [CrossRef]

30. Kuusk, A.; Pisek, J.; Lang, M.; Märdla, S. Estimation of Gap Fraction and Foliage Clumping in Forest Canopies. *Remote Sens. Basel* **2018**, *10*, 1153. [CrossRef]

31. Cui, L.; Jiao, Z.; Zhao, K.; Sun, M.; Dong, Y.; Yin, S.; Li, Y.; Chang, Y.; Guo, J.; Xie, R.; et al. Retrieval of Vertical Foliage Profile and Leaf Area Index Using Transmitted Energy Information Derived from Icesat Glas Data. *Remote Sens. Basel* **2020**, *12*, 2457. [CrossRef]

32. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dy-namics Investigation: High-Resolution Laser Ranging of the Earth'S Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002. [CrossRef]

33. Heilongjiang Forestry and Grassland Bureau. Available online: http://hljforest.com/ (accessed on 18 December 2020).

34. National Geomatics Center of China (NGCC). 30-Meter Global Land Cover Dataset (GlobeLand30)-Product Description Data. 2014. Available online: http://www.globallandcover.com/GLC30Download/index.aspx (accessed on 20 January 2020).
35. Jun, C.; Ban, Y.; Li, S. Open Access to Earth Land-Cover Map. *Nature* **2014**, *514*, 434. [CrossRef] [PubMed]
36. Pang, Y.; Li, Z.; Lefsky, M.; Guoqing, S.; Yu, X. Model Based Terrain Effect Analyses on ICEsat GLAS Waveforms. In Proceedings of the 2006 IEEE International Symposium on Geoscience and Remote Sensing, Denver, CO, USA, 31 July–4 August 2006; p. 3232.
37. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, 2005RG000183. [CrossRef]
38. Hansen, M.C.; DeFries, R.S.; Townshend, J.R.G.; Carroll, M.; Dimiceli, C.; Sohlberg, R.A. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the Modis Vegetation Continuous Fields Algorithm. *Earth Interact.* **2003**, *7*, 1–15. [CrossRef]
39. Wang, Z.; Schaaf, C.B.; Strahler, A.H.; Chopping, M.J.; Roman, M.O.; Shuai, Y.; Woodcock, C.E.; Hollinger, D.Y.; Fitzjarrald, D.R. Evaluation of Modis Albedo Product (Mcd43a) Over Grassland, Agriculture and Forest Surface Types During Dormant and Snow-Covered Periods. *Remote Sens. Environ.* **2014**, *140*, 60–77. [CrossRef]
40. Roman, M.O.; Schaaf, C.B.; Woodcock, C.E.; Strahler, A.H.; Yang, X.; Braswell, R.H.; Curtis, P.S.; Davis, K.J.; Dragoni, D.; Goulden, M.L.; et al. The Modis (Collection V005) Brdf/Albedo Product: Assessment of Spatial Representativeness over Forested Landscapes. *Remote Sens. Environ.* **2009**, *113*, 2476–2498. [CrossRef]
41. Pisek, J.; Chen, J.M.; Nilson, T. Estimation of Vegetation Clumping Index Using Modis Brdf Data. *Int. J. Remote Sens.* **2011**, *32*, 2645–2657. [CrossRef]
42. Chen, J.M.; Cihlar, J. Plant Canopy Gap-Size Analysis Theory for Improving Optical Measurements of Leaf-Area Index. *Appl. Opt.* **1995**, *34*, 6211–6222. [CrossRef]
43. Duong, V.H.; Lindenbergh, R.; Pfeifer, N.; Vosselman, G. Single and Two Epoch Analysis of Icesat Full Waveform Data Over Forested Areas. *Int. J. Remote Sens.* **2008**, *29*, 1453–1473. [CrossRef]
44. Hofton, M.A.; Minster, J.B.; Blair, J.B. Decomposition of Laser Altimeter Waveforms. *IEEE Trans. Geosci. Remote* **2000**, *38*, 1989–1996. [CrossRef]
45. Yang, X.; Wang, C.; Pan, F.; Nie, S.; Xi, X.; Luo, S. Retrieving Leaf Area Index in Discontinuous Forest Using Icesat/Glas Full-Waveform Data Based on Gap Fraction Model. *ISPRS J. Photogramm.* **2019**, *148*, 54–62. [CrossRef]
46. Luo, S.; Wang, C.; Li, G.; Xi, X. Retrieving Leaf Area Index Using Icesat/Glas Full-Waveform Data. *Remote Sens. Lett.* **2013**, *4*, 745–753. [CrossRef]
47. Chen, J.M. Optically-Based Methods for Measuring Seasonal Variation of Leaf Area Index in Boreal Conifer Stands. *Agric. Forest Meteorol.* **1996**, *80*, 135–163. [CrossRef]
48. Wang, Z.; Schaaf, C.B.; Sun, Q.; Kim, J.; Erb, A.M.; Gao, F.; Roman, M.O.; Yang, Y.; Petroy, S.; Taylor, J.R.; et al. Monitoring Land Surface Albedo and Vegetation Dynamics Using High Spatial and Temporal Resolution Synthetic Time Series from Landsat and the Modis Brdf/Nbar/Albedo Product. *Int. J. Appl. Earth Obs.* **2017**, *59*, 104–117. [CrossRef] [PubMed]
49. Huang, H.; Liu, C.; Wang, X.; Biging, G.S.; Chen, Y.; Yang, J.; Gong, P. Mapping Vegetation Heights in China Using Slope Correction Icesat Data, Srtm, Modis-Derived and Climate Data. *ISPRS J. Photogramm.* **2017**, *129*, 189–199. [CrossRef]
50. Tang, H.; Dubayah, R.; Brolly, M.; Ganguly, S.; Zhang, G. Large-Scale Retrieval of Leaf Area Index and Vertical Foliage Profile from the Spaceborne Waveform Lidar (Glas/Icesat). *Remote Sens. Environ.* **2014**, *154*, 8–18. [CrossRef]
51. Wang, X.; Huang, H.; Gong, P.; Liu, C.; Li, C.; Li, W. Forest Canopy Height Extraction in Rugged Areas with Icesat/Glas Data. *IEEE Trans. Geosci. Remote* **2014**, *52*, 4650–4657. [CrossRef]
52. Harding, D.J. Icesat Waveform Measurements of within-Footprint Topographic Relief and Vegetation Vertical Structure. *Geophys. Res. Lett.* **2005**, *32*, L21S10. [CrossRef]
53. Popescu, S.C.; Zhao, K.; Neuenschwander, A.; Lin, C. Satellite Lidar Vs. Small Footprint Airborne Lidar: Comparing the Accuracy of Aboveground Biomass Estimates and Forest Structure Metrics at Footprint Level. *Remote Sens. Environ.* **2011**, *115*, 2786–2797. [CrossRef]
54. Nie, S.; Wang, C.; Li, G.; Pan, F.; Xi, X.; Luo, S. Signal-to-Noise Ratio-Based Quality Assessment Method for Icesat/Glas Waveform Data. *Opt. Eng.* **2014**, *53*, 103104. [CrossRef]
55. Wang, M.; Sun, R.; Xiao, Z. Estimation of Forest Canopy Height and Aboveground Biomass from Spaceborne Lidar and Landsat Imageries in Maryland. *Remote Sens. Basel* **2018**, *10*, 344. [CrossRef]
56. Wang, Z.; Schaaf, C.B.; Sun, Q.; Shuai, Y.; Roman, M.O. Capturing Rapid Land Surface Dynamics with Collection V006 Modis Brdf/Nbar/Albedo (Mcd43) Products. *Remote Sens. Environ.* **2018**, *207*, 50–64. [CrossRef]
57. Lang, A.; Xiang, Y.Q. Estimation of Leaf-Area Index from Transmission of Direct Sunlight in Discontinuous Canopies. *Agric. Forest Meteorol.* **1986**, *37*, 229–243. [CrossRef]
58. Leblanc, S.G.; Chen, J.M.; Fernandes, R.; Deering, D.W.; Conley, A. Methodology Comparison for Canopy Structure Parameters Extraction from Digital Hemispherical Photography in Boreal Forests. *Agric. Forest Meteorol.* **2005**, *129*, 187–207. [CrossRef]
59. Yan, G.; Hu, R.; Luo, J.; Weiss, M.; Jiang, H.; Mu, X.; Xie, D.; Zhang, W. Review of Indirect Optical Measurements of Leaf Area Index: Recent Advances, Challenges, and Perspectives. *Agric. Forest Meteorol.* **2019**, *265*, 390–411. [CrossRef]
60. Sun, X.; Abshire, J.B.; Borsa, A.A.; Fricker, H.A.; Yi, D.; DiMarzio, J.P.; Paolo, F.S.; Brunt, K.M.; Harding, D.J.; Neumann, G.A. Icesat/Glas Altimetry Measurements: Received Signal Dynamic Range and Saturation Correction. *IEEE Trans. Geosci. Remote* **2017**, *55*, 5440–5454. [CrossRef] [PubMed]