Journal of Remote Sensing
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

Exploring the Potential of GEDI in Characterizing Tree Height Composition Based on Advanced Radiative Transfer Model Simulations

Shen Tan[1], Yao Zhang[1,2*], Jianbo Qi[3], Yanjun Su[4,5], Qin Ma[6], and Jinghao Qiu[1]

[1]Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China. [2]Institute of Carbon Neutrality, Peking University, Beijing 100871, China. [3]Center for GeoData and Analysis, State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China. [4]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. [5]University of Chinese Academy of Sciences, Beijing 100049, China. [6]School of Geography, Nanjing Normal University, Nanjing 210023, China.

*Address correspondence to: zhangyao@pku.edu.cn

Citation: Tan S, Zhang Y, Qi J, Su Y, Ma Q, Qiu J. Exploring the Potential of GEDI in Characterizing Tree Height Composition Based on Advanced Radiative Transfer Model Simulations. *J. Remote Sens.* 2024;4:Article 0132. https://doi.org/10.34133/remotesensing.0132

Submitted 23 August 2023
Accepted 5 March 2024
Published 3 April 2024

Copyright © 2024 Shen Tan et al. Exclusive licensee Aerospace Information Research Institute, Chinese Academy of Sciences. Distributed under a Creative Commons Attribution License 4.0 (CC BY 4.0).

Tree height composition describes the relative abundance of trees in different height levels and performs as a critical characteristic for community ecology. The recent launched full-waveform spaceborne LiDAR (Light Detection and Ranging), i.e., Global Ecosystem Dynamics Investigation (GEDI), can map canopy height, but whether this observation reflects tree height composition remains untested. In this study, we firstly conduct numerical simulations to explore to what extent tree height composition can be obtained from GEDI waveform signals. We simulate waveforms for diverse forest scenarios using GEDI simulator coupled with LESS (LargE-Scale remote sensing data and image Simulation), a state-of-the-art radiative transfer model. We devise a minimalistic model, Tree generation based on Asymmetric generalized Gaussian (TAG), for customizing tree objects to accelerate forest scene creation. The results demonstrate that tree objects generated by TAG perform similarly in LiDAR simulation with objects from commercial 3-dimensional software. Results of simulated GEDI waveforms reasonably respond to the variation of crown architectures in even-aged forests. GEDI waveforms have an acceptable ability to identify different height layers within multi-layer forests, except for fir forests with a cone-shaped crown. The shape metric of waveforms reflects the height of each layer, while retrieval accuracy decreases with the increases in height variations within each layer. A 5-m interval between layers is the minimum requirement so that the different height layers can be separated. A mixture of different tree species reduces the retrieval accuracy of tree height layers. We also utilize real GEDI observations to retrieve tree heights in multi-height-layer forests. The findings indicate that GEDI waveforms are also efficient in identifying tree height composition in practical forest scenarios. Overall, results from this study demonstrate that GEDI waveforms can reflect the height composition within typical forest stands.

Introduction

The height distribution of all trees in a forest stand, also denoted as tree height composition, is a crucial aspect that regulates multiple ecosystem processes such as surface roughness, energy transfer, water, and carbon flux exchanges. This ecosystem characteristic is influenced by a complex interplay of ecological succession, including establishment, growth, competition, mortality, and disturbances, and further mediated by human management. The height composition and resulting within-canopy structure of a forest are closely tied to important ecological functions, such as providing habitats for animals [1], influencing the micrometeorological environment, and subsequently impacting biochemical processes [2] and carbon uptake [3]. Therefore, incorporating height composition information into studies that utilize tree heights as inputs, particularly in heterogeneous forests, can be beneficial to ecosystem research [4] and land surface modeling [5]. Specifically, tree height serves as a vital predictor for estimating aboveground biomass (AGB) in forest ecosystems, as allometric equations are widely applied [6]. Neglecting the height heterogeneity can introduce large uncertainties when estimating AGB and quantifying carbon storage, especially when using an established relationship between AGB and tree height at large spatial scales [7].

Downloaded from https://spj.science.org on April 15, 2024

Nevertheless, the height composition is often overlooked or inadequately represented using simple geometric models in ecological studies [8]. This can be attributed to the high costs associated with in situ investigations for large-scale applications and the limited within-canopy structural information provided by passive optical images [9,10]. To address these limitations, the development of an accurate map depicting tree height composition holds great potential in assessing forest biomass and carbon stocks, which are vital for informing effective climate change mitigation strategies and promoting sustainable forest management practices [11].

The Light Detection and Ranging (LiDAR) technology is an active remote sensing technique that utilizes laser pulses with single- or multi-spectral bands to measure the distance between the sensor and targeted objects and is thus used as a powerful tool for accurately measuring the 3-dimensional (3D) structure of forest characteristics [10]. Over the past few decades, LiDAR technology has been extensively utilized to assess forest structural characteristics such as vegetation height [12], canopy structural complexity [13], and foliage profile [14,15] over various spatial ranges. Direct detection of the tree height composition is possible once the height or top of all trees within the stand is known. The airborne LiDAR (ALS) system employs laser pulses with relatively small footprints (approximately at the level of 10^{-1} m in diameter) to penetrate the canopy, thus providing valuable information for detecting forest 3D structures. The canopy height model (CHM) retrieved from ALS point clouds can be used to isolate individual trees by considering local maxima as tree tops [16,17]. Recent advances in LiDAR technology and individual tree segmentation algorithms, including clustering algorithms [18,19] or voxel-based segmentation algorithms [20,21], enhance the performance of quantifying forest height composition. Additionally, the tree height composition can also be indirectly reflected from ALS-derived metrics, i.e., a set of dimensionless scalars known as canopy complexity metrics [22]. While these metrics have been found to explain some forest characteristics and structural traits [13], they have limited capacity to fully characterize the forest vertical profiling and reflect the height compositions. A quantile-based strategy can extract several metrics from the frequency histogram, but these metrics are highly sensitive to the physical structure of the forest stands and are difficult to formulate a priori prediction and exhibit limited capacity to utilize order information of the vertical profiling and to be applied in a wide spatial range [23].

The spaceborne LiDAR sensors, such as the Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and Land Elevation Satellite (ICESat) and the Global Ecosystem Dynamics Investigation (GEDI) onboard the International Space Station, have reinforced the ranging capability [10]. These systems allow for wide-ranging terrestrial sampling with a relatively larger footprint (approximately at the level of 10^{1} m in diameter). By fitting GEDI signals with ALS tree height results and optical signals, several wall-to-wall tree height maps have been produced based on different spatial extrapolation approaches covering a large region [12,13,24]. These products mostly rely on selected relative height (RH) metrics and ignored the full profile of the GEDI waveforms, which contains important information on forest vertical structure. Thus, most of these products only focus on the canopy top height and do not fully consider the tree height layering and height composition within a forest stand. Full utilization of spaceborne waveform LiDAR pulses allows for the representation of crucial forest traits along the vertical dimension, including plant volume density and crown cover profile [25].

To the best of our knowledge, no study has yet demonstrated whether GEDI waveforms can represent tree height composition for forest stands with a diverse forest ecosystem, i.e., different crown architectures (CAs), and varying canopy coverages, both of which influences the shape of waveforms. The main obstacle is the absence of accurate and sufficient in situ tree height investigations and the potential geolocation bias of GEDI samples [26]. Fortunately, radiative transfer models (RTMs) provide a new opportunity to test how GEDI waveform responds to tree height composition under more comprehensive forest conditions. Current RTMs are well-suited to complex scenes and do not oversimplify the interactions between rays and terrestrial objects [27]. Examples of such RTMs include FLiES (Forest Light Environmental Simulator) [28], DIRSIG (Digital Imaging and Remote Sensing Image Generation) [29], DART (Discrete Anisotropic Radiative Transfer) [27,30], and LESS (LargE-Scale remote sensing data and image Simulation) [31]. With these RTMs, researchers are empowered to simulate LiDAR observations within virtual environments that encompass a diverse array of tree objects [13,32–35]. An RTM-based simulation process usually entails the generation of discrete LiDAR point clouds and subsequent aggregation to form large footprint waveforms encompassing both horizontal and vertical energy distributions [36,37]. A comprehensive experiment based on RTM can help identify the bottlenecks and guide future observational based experiments.

However, the computational cost of using RTMs to simulate LiDAR signals and GEDI waveforms across extensive forest conditions persists as a challenge. The issue of unsatisfactory efficiency in simulating reflected signal has been partially mitigated by the current LESS, which enhances computational efficiency through the integration of a state-of-the-art image renderer and a parallel computing module [31,38]. Another challenge is the lack of a computationally efficient method for generating tree objects with continuous CA variation. Tree CAs are important parameters that link the one-dimensional tree height information to 3D individual tree objects, which are required for RTM scene setup. Several methods have been proposed to abstract and describe the crown in RTMs, including the S-crown (simplified crown) method that describes the crown as geometric primitives (e.g., ellipsoid, cone, and cylinder [39]), the H-layer (hybrid layer) method that treats the canopy as one or multiple horizontal layers, and the M-surface (mesh surface) method that can abstract all elements of a single tree (including leaves and branches) by a series of triangles [40]. However, existing methods were not specifically designed for rapid adaptation to variations in tree boundaries, a common natural response to continuously changing climatic conditions [41]. While commercial software commonly customizes object boundaries, it requires extensive manual intervention, limiting its widespread applicability [8]. Therefore, a lightweight method that can generate tree objects with customized boundaries and acceptable efficiency is urgently needed.

In this study, to investigate the extent to which GEDI waveforms can reflect the tree height composition within diverse forest scenarios, we simulate GEDI waveforms by GEDI simulator from discrete ALS point cloud generated by LESS in both even-aged and multi-height-layer forests covering a wide continuous variation of CAs. To facilitate this simulation, we also propose a novel method, namely, Tree generation based on Asymmetric Generalized Gaussian (TAG), that utilizes an asymmetric generalized Gaussian (AGG) function to customize

Downloaded from https://spj.science.org on April 15, 2024

the boundary of tree objects with an improved computational efficiency. This method allows users to generate tree objects with continuous variation, as opposed to using geometric primitives. We designed 4 groups of experiments, aiming to address the following questions: (a) Can tree objects generated by TAG replace 3D tree models in LiDAR point cloud simulation? (b) How do GEDI waveforms respond to variations in CAs in even-aged forests? (c) How do the GEDI waveforms respond to tree height composition and variation in multi-height-layer forests, and can the height composition of these forests be retrieved? (d) Can real GEDI observations be utilized in reflecting tree height in multi-height-layer forests? The answer to these questions from this simulation study provide necessary insights and guidance for future efforts aimed at retrieving forest height compositions covering more comprehensive forest scenarios.

Materials and Methods

GEDI waveform generation

The GEDI signal generation is a 2-step process, each step having undergone separate evaluations [42,43]. The initial step encompasses the simulation of ALS point cloud using LESS. This step necessitates the prior generation of individual tree objects within each pre-designed virtual forest scenario. Each tree object is constructed by TAG, as the representation of diverse and realistic tree structures can be ensured. Subsequently, the second step simulates the full waveform GEDI signal from the aforementioned discrete point cloud using the GEDI simulator. For further details of this workflow, refer to the following text.

Introduction of LESS

We use LESS to simulate discrete LiDAR point clouds from a comprehensive setting of forests in this study. LESS is a ray-tracing-based 3D RTM that can simulate remote sensing signals and images over large-scale realistic scenes (http://lessrt.org/; last access: 2023 June 10). LESS employs a backwards path tracing (BPT) method to generate images or signals (e.g., LiDAR signal in this study). By configuring the required parameters including the sensor parameters, platform parameters, and scenario parameters, LiDAR signals can be generated in batch mode with satisfactory efficiency [31].

Representation of single tree objects

We build a minimalistic model, TAG, to reflect the continuous variation of the CAs for each tree object, rather than employing constant crown geometric primitives (e.g., ellipsoid, cube, and cone in current LESS). TAG considers each tree as a symmetrical object for all horizontal planes. That is, the object is the rotating result of a curve. Each tree can be divided into a trunk part and crown portion. Its outline on a horizontal plane can be described by 4 size metrics and 2 CA metrics (Fig. 1). In this method, tree trunks are abstracted by cylinders. Four size metrics are thus required to describe the structure of an individual tree: tree height, crown height ratio (CHR, the ratio of crown height to tree height), trunk height ratio (THR, the ratio of trunk height to tree height, i.e., 1-CHR), and diameter height ratio (DHR, the ratio of crown diameter to crown height, Fig. 1A).

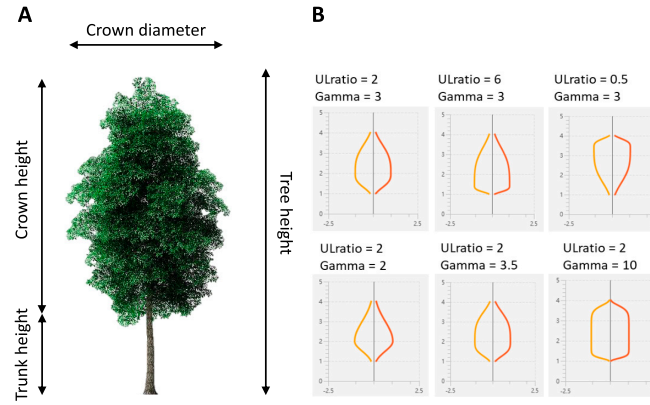

Fig. 1. Diagram describing the outline of a single tree. Panel (A) illustrates the trunk metrics of an ash tree, and panel (B) illustrates how crown outline can be determined by the ULratio (top column) and Gamma (bottom column). All trees in panel (B) have the same size metrics: height = 4 m, crown height = 3 m, and crown diameter = 2 m. The unit of the coordination system in panel (B) is meters.

We use an AGG distribution to describe the outline of the crown portion in TAG. The relative crown diameter $D(i)$ at height i can be written as:

$$(i)=\begin{cases}\dfrac{\text{Gamma}}{(\beta_{\text{upper}}+\beta_{\text{lower}})\Gamma(1/\text{Gamma})}\exp\left(-\left(\dfrac{-i}{\beta_{upper}}\right)^{\text{Gamma}}\right) & (i>h_{\max})\\ \dfrac{\text{Gamma}}{(\beta_{\text{upper}}+\beta_{\text{lower}})\Gamma(1/\text{Gamma})}\exp\left(-\left(\dfrac{-i}{\beta_{\text{lower}}}\right)^{\text{Gamma}}\right) & (i<h_{\max})\end{cases} \quad (1)$$

where

$$\beta_{\text{upper}}=\sigma_{\text{upper}}\sqrt{\frac{\Gamma(1/\text{Gamma})}{\Gamma(3/\text{Gamma})}} \quad (2A)$$

$$\beta_{\text{lower}}=\sigma_{\text{lower}}\sqrt{\frac{\Gamma(1/\text{Gamma})}{\Gamma(3/\text{Gamma})}} \quad (2B)$$

In Eqs. 1 and 2, σ_{upper} and σ_{lower} represent the standard deviations of the upper and lower segments of the Gaussian curves, respectively. These parameters are pivotal in defining the kurtosis of the respective curve segments. It is discerned that the ratio (ULratio, σ_{upper}:σ_{lower}) is instrumental in characterizing the kurtosis of the crown outline. Another crucial metric, Gamma, is employed to regulate the skewness and the contour of the Gaussian curve's tail. Specifically, Gamma influences the curve's slope and elevation near its apex and also modulates the curve's rate of decay. $\Gamma(x)$ is a gamma function. All trees in this study were designed with a single peak within their outline to prevent the occurrence of 2 peaks in a single tree, which could affect height retrieval.

Once the boundary of the canopy has been specified by the above metrics, branches within the canopy can be either neglected or generated with user-specified parameters, including branch height interval, as well as azimuth and zenith angles, and the length of each branch is constrained by the boundary of the crown [44]. These branches are then represented by

Downloaded from https://spj.science.org on April 15, 2024

M-surfaces. All leaves, represented by a series of polygons with specific coordinates, are randomly generated and populated within the tree outline, following the assignment of parameters that describe leaf shapes. During the leaf placement process, if branches are present, leaves are inclined toward the nearest branches. A leaf clumping factor describes the tendency of leaves to cluster close to branches. This method has been included in the latest LESS version.

GEDI simulator

Building upon the approach suggested by Blair and Hofton [36], the generation of GEDI waveforms is conducted by considering the weighted contributions of the point cloud generated by LESS in this study [43]. The point clouds are decomposed to represent the reflected intensity peaks originating from the compact laser pulses with small footprints [45]. When considering the horizontal plane, the weight assigned to each point (denoted as $I_{w,i}$) can be calculated based on its distance from the central point (x_o, y_o) of the footprint (Fig. 2):

$$I_{w,i} = I_i \frac{1}{\sigma_f \sqrt{2\pi}} e^{-\frac{(x_i - x_o)^2 + (y_i - y_o)^2}{\sigma_f^2}} \quad (3)$$

where (x_i, y_i) represents the position of the ith point, and σ_f denotes 1/4 of the width of the GEDI footprint. The I_i is a relative weighting of each point. In this study, we follow Blair and Hofton [36] and adapt this weight to be equal ($I_i = 1$) for all points.

To optimize computational efficiency, we have implemented the vertical convolution after, rather than before, the binning process, and have neglected the noise signal from background light and electronic noise within GEDI waveforms. The GEDI simulator is available online (https://bitbucket.org/StevenHancock/gedisimulator/src/master/; last access: 2023 May 20).

Simulation workflow

To execute each simulation by calling LESS SDK in Python, 5 steps are required.

1. Environment Configuration. This step involves specifying the working space, importing relevant modules, and assigning the central processing unit (CPU) cores to be used during the simulation.

2. Single Tree Simulation. Once the CA parameters are accepted, each individual tree object can be generated. Simulation tasks utilizing TAG will be assigned to separate CPU cores. At this stage, users can choose whether or not to consider and simulate the branches within the canopy.

3. Scene generation. A virtual scene can be generated using a set of parameters. In this study, each scene has a diameter of 25 m matching the GEDI footprint, which is similar to the suggested LiDAR footprint for reflecting forest structures [35]. This step should be repeated until each tree's position is assigned.

4. LiDAR point cloud simulation. The full waveform LiDAR signals are simulated and decomposed into discrete point clouds using a Gaussian decomposition algorithm. We employ the following sensor configuration parameters during the simulation: sensor wavelength = 1,064 nm, beam axial division = 5 representing the cross-section of each laser pulse is divided into 5×5 parts, platform altitude = 800 m, platform resolution = 0.1 m, and platform range = 100 m, representing the vertical range where the return pulses would be recorded. The reflectance and transmittance for leaves are 0.4 and 0.4, respectively, 0.2 and 0.0 for branches, and 0.3 and 0.0 for soil.

5. GEDI waveform simulation. We generate GEDI waveforms using the discrete point clouds obtained in the step 4 following the method in the previous section, with a vertical resolution of 1 m.

GEDI data and processing

To test whether the knowledge we obtained from RTM simulations can guide real-world application, we also carry out an experiment using real GEDI data at the Abby Road ecology site (ABBY) in the United States. This study site is characterized by mixed forests with several canopy layers, spanning approximately 170 km^2 (Fig. 3). A CHM with a 1-m spatial resolution, derived from airborne LiDAR observations, is generated annually for this region, except for the year 2020. We selected the CHM from 2019 in this experiment. The original point cloud records and CHM are accessible for download at https://data.neonscience.org/data-products/explore (last access: January 2024).

We collected the GEDI L2a v002 product for the period from April to October 2019 from https://search.earthdata.nasa.gov/search (last access: January 2024). The GEDI samples were filtered based on following criteria: a "quality_flag" not equal to 0, a "degrade_flag" not equal to 1, and a "sensitivity" greater than 0.95. Subsequently, 174 GEDI samples located in multi-layer forests were selected through visual interpretation. The RH curves from each sample were interpolated into a uniform height interval using a spline function, facilitating the derivation of the waveform from these RH curves. The heights of local maximum intensity (HoIs) within these waveforms were then identified to examine their responsiveness to variations in tree height. The majority of these samples originated from forests with 2 layers (top and bottom layer), while 39 samples represented 3-layer forests (top, middle, and bottom layer). To mitigate the impact of geolocation bias inherent in GEDI signals, we excluded samples exhibiting more than 10 m of absolute bias between the tree height from GEDI observation (i.e., tree height at RH98) and the CHM.

Experimental designation

We carry out 4 experiments in this study.

Experiment 1

This experiment tests whether the tree models generated by the method in this study can be employed in simulating point cloud and GEDI signals. We compare the simulated GEDI signals using simplified tree models with no branches, simplified tree models with branches generated by a B-cluster (boundary cluster) algorithm [44], and virtual 3D tree objects with explicitly described structures for commercial usage (obtained from https://www.cgtrader.com/). We design 3 scenes with different tree species representing typical crown architectures (cone: fir, slim ellipsoid: ash tree, and flat ellipsoid: mango tree, see Fig. 4). The size metrics and CA metrics for the 3 tree species are outlined from the tree objects (see Table 1). The crown cover for each scene is 90%.

Experiment 2

This experiment investigates whether the simulated GEDI waveforms can respond reasonably to variations in size and CA

Downloaded from https://spj.science.org on April 15, 2024

Table 1. Trunk metrics and CA metrics for the 3 tree species in Experiment 1

		Size metrics				Architecture metrics	
Species	Crown shape	Tree height (m)	CHR	THR	DHR	ULratio	Gamma
Fir	Cone	20	0.9	0.1	0.35	6	1.5
Ash	Slim ellipsoid	20	0.65	0.35	0.67	1.25	5
Mango	Flat ellipsoid	20	0.4	0.6	1.5	1.25	3

metrics in even-aged (uniform tree heights) forests and whether these partial relationships are predictable. In this experiment, all trees within each forest scene are set to have the same size and CA metrics reflecting their uniform age. The default settings of the metrics are as follows: tree height = 20 m, THR = 0.25, CHR = 0.75, DHR = 0.4, ULratio = 2, and Gamma = 3. We carry out the simulation with different crown coverages: sparse forest (10%), medium forest (50%), and dense forest (90%). In this experiment, we aim to explore the feasibility of retrieving tree height from GEDI waveforms. To achieve this, the HoIs from each waveform are integrated into a linear regression analysis. HoIs, representing the localized peak return energy, exhibit a strong correlation with crown coverage at specific heights. This correlation yields insightful data regarding forest height for the second or third tree height layers, premised on the theoretical knowledge of CA metrics. During the analysis, HoIs are extracted by identifying values that are larger than their neighbors within a 3-step moving window.

Experiment 3

The purpose of this experiment is to examine the variability of GEDI waveforms in response to multiple height layers, a common occurrence in natural forest conditions. Many forests consist of secondary growth, and thus exhibit multiple height layers, which arise from either regeneration after clearcutting or selective logging. We design 4 forest scenarios, each incorporating various scenes, to assess the waveform response to height variation within each layer, to different height intervals, to diverse crown cover for each height layer, and to the mixing of CAs within each layer (Fig. 4). In Experiment 3, the variation of tree height within each layer is quantified by the coefficient of variation (CV), which reflects the standard deviation in relation to the average height of each layer. A higher CV indicates greater variation in tree height and increased standard error. For instance, in the first set with a CV of 10%, the tree height composition for the 10-m layer is 10±1 m. Waveforms with 4 increasing CV levels—5%, 10%, 15% and 20%—are focused. The uncertainty of the simulated GEDI waveforms in Experiment 3 is evaluated by conducting 10 repeated simulations but with randomly generated tree positions within the footprints.

During the simulation, tree heights are initially generated randomly, adhering to the targeted CV, followed by the calculation of tree quantities for each height layer, given that the crown coverage is predetermined in this experiment. In this phase, all other tree CAs, such as ULratio, Gamma, CHR, and THR, are kept constant. Subsequently, tree positions were randomly generated and allocated within the designated scenes.

For the first scenario, 3 forest height composition scenes are designed with different layer combinations (first: 10 m and 20 m; second: 15 m and 30 m; third: 10 m, 20, and 30 m) for 3 different CAs (fir, ash, and mango; CA metrics are shown in Table 1). The crown cover is 90% for all scenes during simulation. Secondly, we investigate the minimum height interval between layers that GEDI waveforms can detect. To do so, we simulated GEDI waveforms in 2 CAs with varying height intervals of 3 m, 5 m, and 8 m as 3 scenes for different CAs. Thirdly, we aim to assess the capability of the waveforms to differentiate variations in crown cover within each height layer, since under some natural conditions, a layer with a higher crown cover can be located at any part of the canopy. For this purpose, a scenario including 9 forest scenes for different CAs is designed, comprising 3 height layers with varying crown cover (10%, 10%, and 30% of crown cover for the 3 height layers with a total of 50% coverage; 10%, 20%, and 40% with a total of 70% coverage; 10%, 30%, and 50% with a total of 90% coverage). Finally, a comparison is designed for displaying the waveform response to mixing CAs, which are commonly observed in most natural forest ecosystems [41,46]. Each of the scene features 3 ash tree height layers with 10 m interval and introduced mango trees with a proportion of 25% and 50%, respectively. These 2 tree species are selected due to their typical crown architectures found in terrestrial ecosystems.

Another aim of Experiment 3 is to investigate the potential of using the HoI as a predictor for the height of each crown layers, thereby representing the height composition. The relationship between HoI and tree height is determined by the CA metrics, allowing us to establish the HoI-height function in even-aged forests. Leveraging the results obtained from Experiment 2, we estimate the parameters for HoI-height fitting function for each species. Subsequently, using these fitted height functions, we retrieve tree heights for different layers in the 4 scenarios in Experiment 3 from HoIs of the waveforms. The inference results obtained from scenarios 1 and 2 are intended to assess the robustness of the fitted function in retrieving the tree height composition in multi-layer forests based on the height function fitted in even-aged forests. Similarly, the inference results from scenario 3 evaluate the robustness of the height retrieval when confronted with dynamic crown coverage for each layer. Lastly, the inference results from scenario 4 provide insights into the performance of height retrieval in mixed forests.

Experiment 4

This experiment aims to explore the effectiveness of actual GEDI waveforms in representing both top and sub-canopy

Downloaded from https://spj.science.org on April 15, 2024

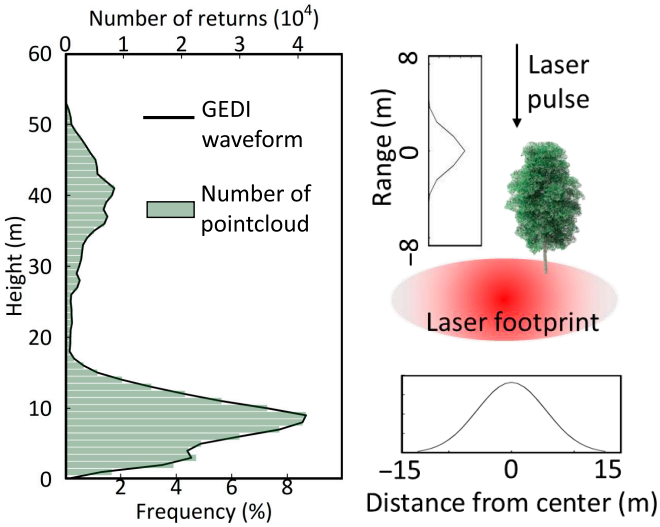

Fig. 2. Diagram of the generation of GEDI waveform (black line in the left panel) from the discrete point cloud at a forest stand with *Pseudotsuga menziesii*. Within a 25-m footprint, there are 10 trees each with a height of 50 m, and 100 shrubs (spherical) each with a height of 2 m. The numbers and frequency of point clouds in each bin (i.e., 1 m in this study) are represented by green columns. The curves in the right panel represent the energy distribution of GEDI instrument on vertical and horizontal dimensions.

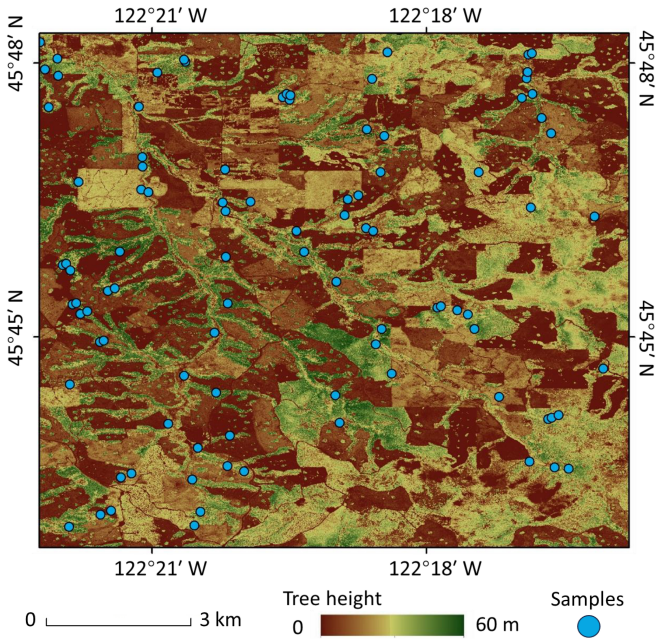

Fig. 3. The CHM covering the ABBY site. The canopy heights are represented by color palette from red to green. The blue circles represent the locations of GEDI samples over multi-layer forests.

tree heights in multiple-height-layer forests, thereby assessing height composition retrieval performance. It seeks to determine if these real-world observations align with findings from prior simulation-based experiments. We also follow the height retrieval strategy using HoIs from GEDI waveforms to predict tree heights extracted from CHM.

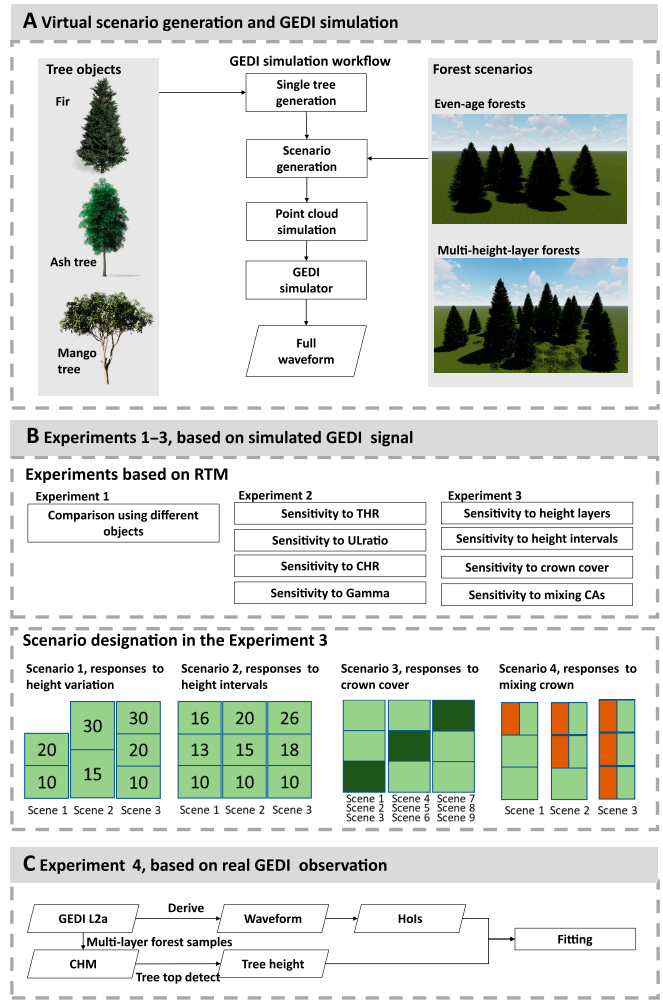

Fig. 4. Schematic representation of the workflow and the structuring of the 4 experiments conducted in this study. (A) The workflow of GEDI simulation. (B) The simulation-based experiments. (C) The real GEDI data-based experiment. Gray blocks symbolize the virtual tree or forest scenarios. In Experiment 3, 4 distinct forest scenarios are depicted, each corresponding to a unique experimental objective. The numerical values within the green blocks indicate the height of each layer. For sensitivity to crown cover tests in Experiment 3, variations in color (dark green and dark orange) signify layer positions with the densest canopy cover and mixed tree species (CAs), respectively. Tree objects illustrated in this figure were crafted using commercial software, with the final rendering accomplished via Lumion (https://lumion.com/, last access: September 2023).

Results

Evaluation of abstracted tree objects from AGG function

The simulated waveforms exhibit limited differences using tree objects from commercial software and using tree objects abstracted by the method in this study (Fig. 5). There are minor differences between waveforms using tree objects for commercial usage (black line in Fig. 5) and generated by TAG (blue and orange line), as TAG is designed to capture a symmetrical crown outline. Including branches within the tree crown also has limited influence on the resulting waveforms. The intensity of GEDI waveforms mainly reflects the cloud point frequency at each height interval (see the diagram in Fig. 2). The limited influence of the trunk and branches on the waveform can be explained by 2 main reasons. Firstly, LiDAR instruments

Downloaded from https://spj.science.org on April 15, 2024

observe the within-canopy structure by penetrating laser pulses through the gaps between leaves; trunks and branches are likely being shaded by top leaves. Furthermore, the surface of the trunk and branches does not allow laser transmission (i.e., transmittance = 0), and their contribution to the point cloud frequency during waveform generation can be deemed negligible.

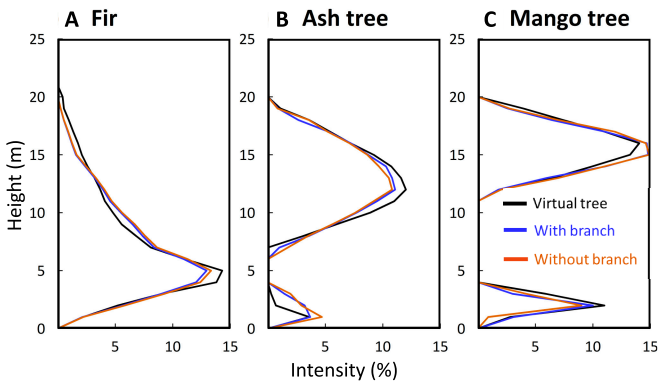

Fig. 5. Simulated waveforms using virtual tree objects (in black), simplified tree objects with branches (in blue), and simplified tree objects without branches (in orange) in 3 scenes [fir, ash tree, and mango tree in panels (A), (B), and (C), respectively].

This is also supported by Qi [44], who reported a high consistency of simulated reflectance using tree objects generated by M-surface and B-cluster methods (the correlation coefficient, R^2, is 0.99). Additionally, by abstracting the crown with an AGG function, the boundaries of the tree object are closer to the natural trees. This improvement overcomes the slight simulation bias caused by using the S-crown method since this method usually mismatches the crown boundary between geometric primitives and natural trees [44].

The waveform responses to canopy architecture in even-aged forests

In the even-aged forest scenes where all trees have similar height, the resulting waveforms can be divided into crown and terrain parts (Fig. 6). The integral intensity of the crown portion (number of crown-intercepted LiDAR returns) has a nonlinear positive relationship with the crown cover. The integrated value of crown portions increased from ~60% to ~75% when the maximum crown cover increased from 10% to 50% but only increased from ~75% to ~95% when the maximum crown cover increased from 50% to 90%, indicating that the LiDAR returns intercepted by the crown increased only slightly after the crown cover reached a certain threshold. Simultaneously, the variations in the trunk and crown height proportions are insulated

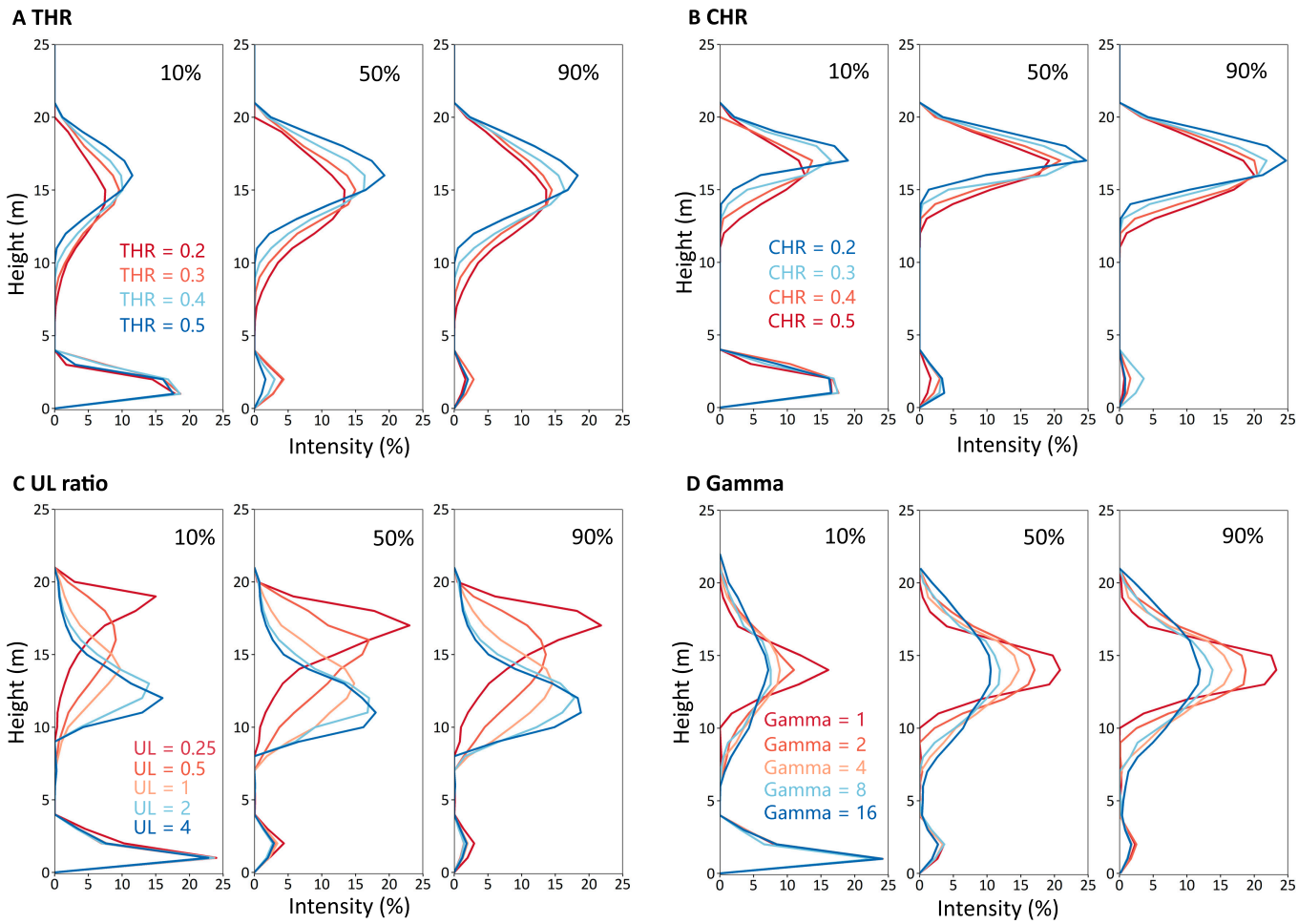

Fig. 6. Simulated waveforms with different trunk metrics [THR in panel (A) and CHR in panel (B)] and crown architecture metrics [ULratio in panel (C) and Gamma in panel (D)]. The simulated results with different metrics are given in different gray colors. The percentage in each subpanel represents the maximum crown cover in this simulation.

Downloaded from https://spj.science.org on April 15, 2024

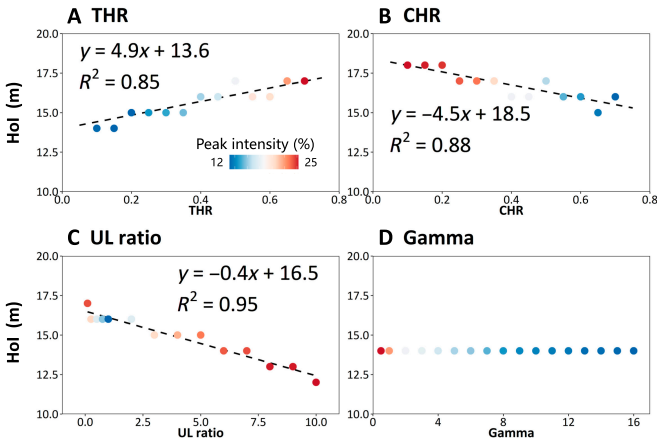

Fig. 7. The sensitivity of the HoI of waveforms (colored circles) to trunk and crown architecture metrics, including (A) THR, (B) CHR, (C) UL ratio, and (D) Gamma. The color of these circles represents the peak intensity. The fitting function and R^2 are labeled in each panel.

from the energy proportions of the crown signal and terrain signal. Since all trees within each scene share the same size and CA metrics, the 2 trunk-related metrics, THR and CHR, determine the ratio of trunk and crown height to the total tree height in GEDI signals (i.e., THR + CHR = 1, Fig. 6). With an increase in THR (decrease in CHR), the range of the crown portion decreases, while the length of the trunk range increases, and vice versa. This phenomenon indicates that we can only use CHR, rather than both CHR and THR, in estimating the sensitivity of HoI to tree height.

The ULratio and Gamma determine the shape of crowns and thus the resulting GEDI waveform (Fig. 6). The skewness of the crown boundary is mainly determined by the ULratio. Its influence can be divided into 3 stages: ULratio < 1 means the crown outline skewed to the upper part; ULratio = 1 means the crown outline is a curve of normalized distribution, at this moment, the boundary of the crown is close to an ellipsoid; and ULratio > 1 means the crown outline skewed to the lower part (Fig. 6C). A reciprocal of ULratios leads to symmetrical crown outlines. Gamma determines the kurtosis of the crown outline and GEDI signals and thus has limited influence on the length of the crown and trunk parts of GEDI waveforms (Fig. 6D).

The sensitivity to the other 2 size metrics, tree height and DHR, is not shown in this experiment. All the trees in this experiment have a height of 20 m, meaning that the top signals of the GEDI waveforms are approximately 20 m. Increasing the tree height can thus increase the range of the resulting GEDI waveforms and vice versa. During the simulation, we kept the crown cover constant (10%, 50%, and 90%). A variation in DHR only influences the tree amount within each scene but not on the crown cover.

Within an even-aged forest characterized by a single tree height layer, we observed statistically significant correlations between the HoI and 3 metrics (Fig. 7). Unlike the peak intensity of the waveform, HoI demonstrates insensitivity to variations in crown cover. The HoI exhibits a positive relationship with THR but a negative relationship with CHR with satisfactory confidence (R^2 = 0.85 and 0.88, respectively, Fig. 7A and B). An increase in ULratio leads to a decrease in the peak height of GEDI waveforms (R^2 = 0.95, Fig. 7C). An increase in Gamma has no effect on the HoI (Fig. 7D).

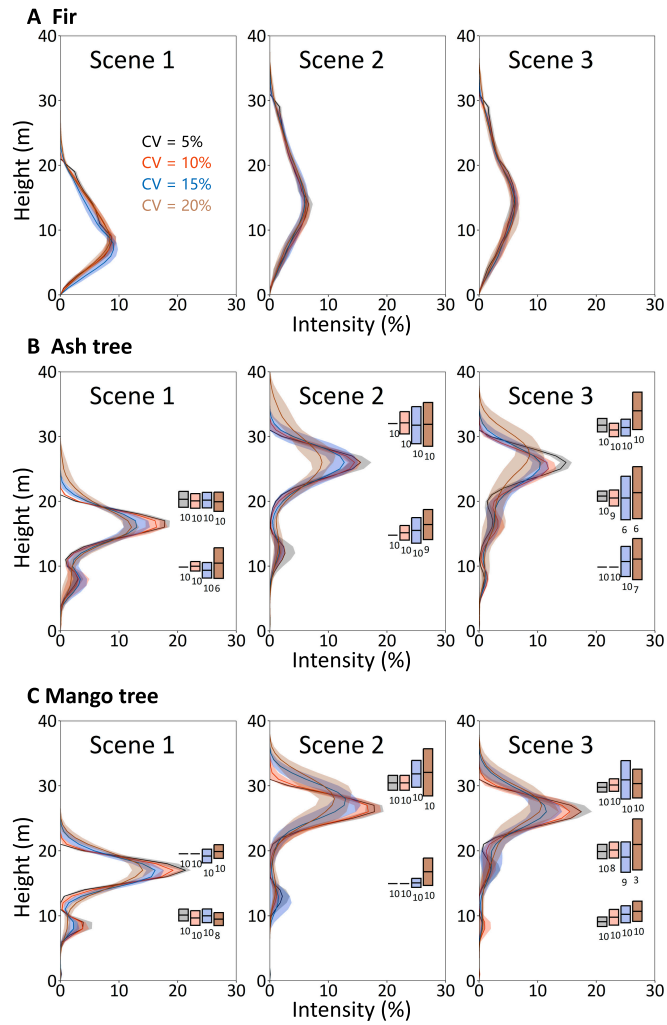

Fig. 8. Diagram of GEDI waveforms in 3 tree height scenes for fir in panel (A), ash tree in panel (B), and mango tree in panel (C). Different CVs are depicted by various colors in the waveforms. The shading within each waveform signifies the standard error across 10 simulations. The height retrieval results from the HoIs for each crown layer are illustrated through colored columns. In these columns (bar plots), the central short line denotes the average prediction, and the error bars convey the 95% confidence range pertinent to the retrievals. Numerical labels on each retrieval reflect the number of waveform peaks detected in a series of 10 simulations. Each height layer displays 4 retrieval results, each symbolizing 1 of the 4 CVs, and is color-matched with the corresponding waveforms.

The waveform responses to tree height variation in multi-height-layer forests

The waveform responses to height variation

The GEDI waveforms demonstrate a noticeable increase in uncertainty as tree height variation amplifies, as indicated by the standard error observed across 10 repeated simulations (Fig. 8). In scenes where tree height variation is limited (i.e., CV = 5%), the maximum height captured by the GEDI waveforms closely approximates the actual tree height. For instance, in scene 1, the maximum height matches a tree height of 20 m, while in scenes 2 and 3, it corresponds to a tree height of 30 m. As the CV increases, indicating a greater likelihood of taller trees existing, the observed maximum waveform height also rises.

The variation in the simulated waveforms is influenced not only by the increasing CV but also by the CAs. Notably, fir

Downloaded from https://spj.science.org on April 15, 2024

forests demonstrate a higher tolerance to tree height uncertainties, resulting in a consistent pattern in simulated waveforms across different CVs. This behavior can be attributed to the relatively short trunk heights (THR = 0.1) of fir trees, which leads to a coupling between the lower portion of taller fir trees and shorter fir trees. However, this coupling effect hampers the accurate detection of waveform peaks from different layers and the determination of height composition within fir forests.

The waveforms obtained from ash and mango forests exhibit a stronger response to tree height variation, enabling the retrieval of each height layer with a considerable degree of confidence, even the tree heights in current scenarios are different from the samples for fitting in Experiment 2 (R^2 for both species are 0.99, see Fig. S1). Since the ULratios for these 2 species are the same and Gamma has no effect on the relationship between HoI and tree height, we estimate the sensitivity of HoI to tree height based on the results in Fig. 7C (Table 2). Due to the relatively shorter crown height of mango trees and a lower CHR, the HoI values for specific height layers in mango forests are higher compared to those in ash forests. With the exception of fir forests, the height of HoI can be utilized to confidently retrieve the tree height for each layer under limited CV conditions. However, as the CV increases, the retrieval uncertainty also increases. When the CV exceeds 15%, some waveform peaks corresponding to medium or lower sections of the canopy become overlapped with adjacent layers, leading to a potential loss of distinction between height layers.

Although the crown cover is identical for each layer, the top layer in the waveforms consistently exhibits the highest intensity compared to the lower layers. This phenomenon stems from the fact that the top layers intercept a greater portion of the downwelling radiative energy during radiative transfer, consequently resulting in a higher likelihood of generating LiDAR returns. As a consequence, there is a possibility of failure in accurately detecting waveform peaks for the lower layers.

The waveform responses to different height intervals

The capability of GEDI waveforms to differentiate canopy layers and their performance in height retrieval are influenced by the spacing between the crown layers and the CA (Fig. 9). We did not conduct this analysis or subsequent analysis for fir forests, as previous findings suggest that different canopy layers cannot be effectively distinguished using GEDI waveforms in those forests. In general, the height retrieval performance in mango tree forests surpasses that in ash tree forests. This discrepancy arises from the relatively shorter crown portions in mango trees, which results in a smaller likelihood of waveform peaks from adjacent layers becoming overlapped. When the interval between layers is 3 m, only one peak can be distinguished within any of the waveforms, making it impossible to detect height layers. There is a significant underestimation in tree height using the HoI and the height function from the "The waveform responses to height variation" section, suggesting different height layers are coupled together. The HoI-height function no longer performs accurately. Similarly, the medium layer (15 m height) in both forests with a 5-m interval cannot be reliably differentiated. On the other hand, the top layer can consistently be detected and retrieved, while the bottom layer can be detected only when the CV is below 10%. Increasing the interval to 8 m allows for successful detection and retrieval of nearly all 3 layers in both forests when the CV is below 10%.

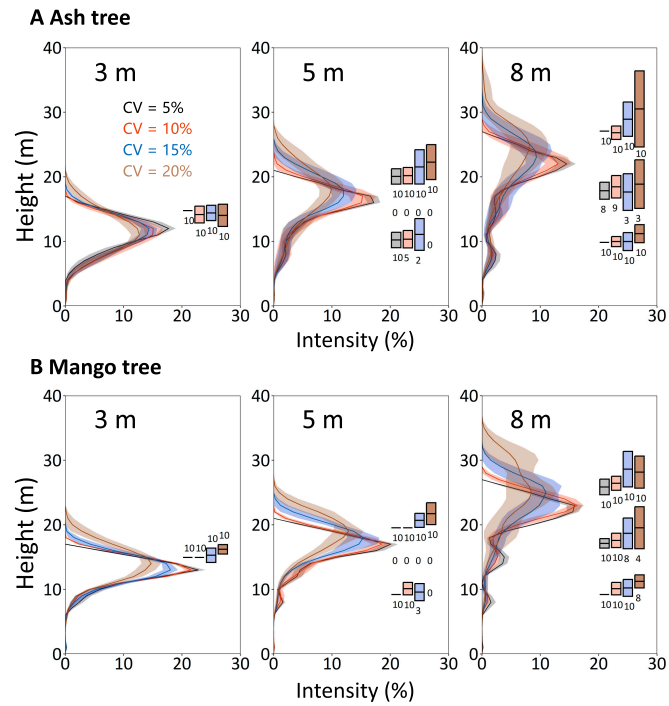

Fig. 9. Diagram of GEDI waveforms in 3 canopy layers with different height resolutions (3 m, 5 m, and 8 m) for ash tree forests in panel (A) and mango forests in panel (B). Other symbols have the same meaning in Fig. 8.

However, this success becomes challenging when the CV exceeds 15%. For instance, in ash tree forests with CV below 10%, only a few failures are observed in detecting the medium layer with an 8-m interval (2 failures for a 5% CV and 1 failure for a 10% CV). Nevertheless, as the CV exceeds 15%, the likelihood of failure increases (7 failures for both 15% and 20% CV).

The waveform responses to different crown cover

The above 2 experiments assume a constant canopy density for each layer (30%), while this canopy cover fraction can be different under actual conditions. We conduct the third experiment to test the performance of GEDI waveform in retrieving height layering when the canopy coverage varies across layers. Our findings reveal that the top layer can always be reliably detected using GEDI waveforms, regardless of the location of the densest canopy locates (Fig. 10). For the middle and bottom layer, the layer height can be successfully retrieved either when the total canopy cover is low or the densest canopy cover is at the bottom layers. These outcomes suggest that waveform intensity may not always accurately reflect the differences in crown cover between layers, given that a significant portion of energy is intercepted by the top layer(s). Similar to the results in previous sections, the height of each layer can also be retrieved by the function for the 2 species respectively, demonstrating its robustness under varying crown cover. The capacity to detect waveform peaks and retrieve tree height using HoI is generally maintained under most circumstances where the CV is less than 10% (at least 8 peaks can be detected in 10 simulations). An increased CV may result in a tree transitioning to an adjacent layer, causing a blurred wave trough between layers and compromising the accuracy of the detection.

Downloaded from https://spj.science.org on April 15, 2024

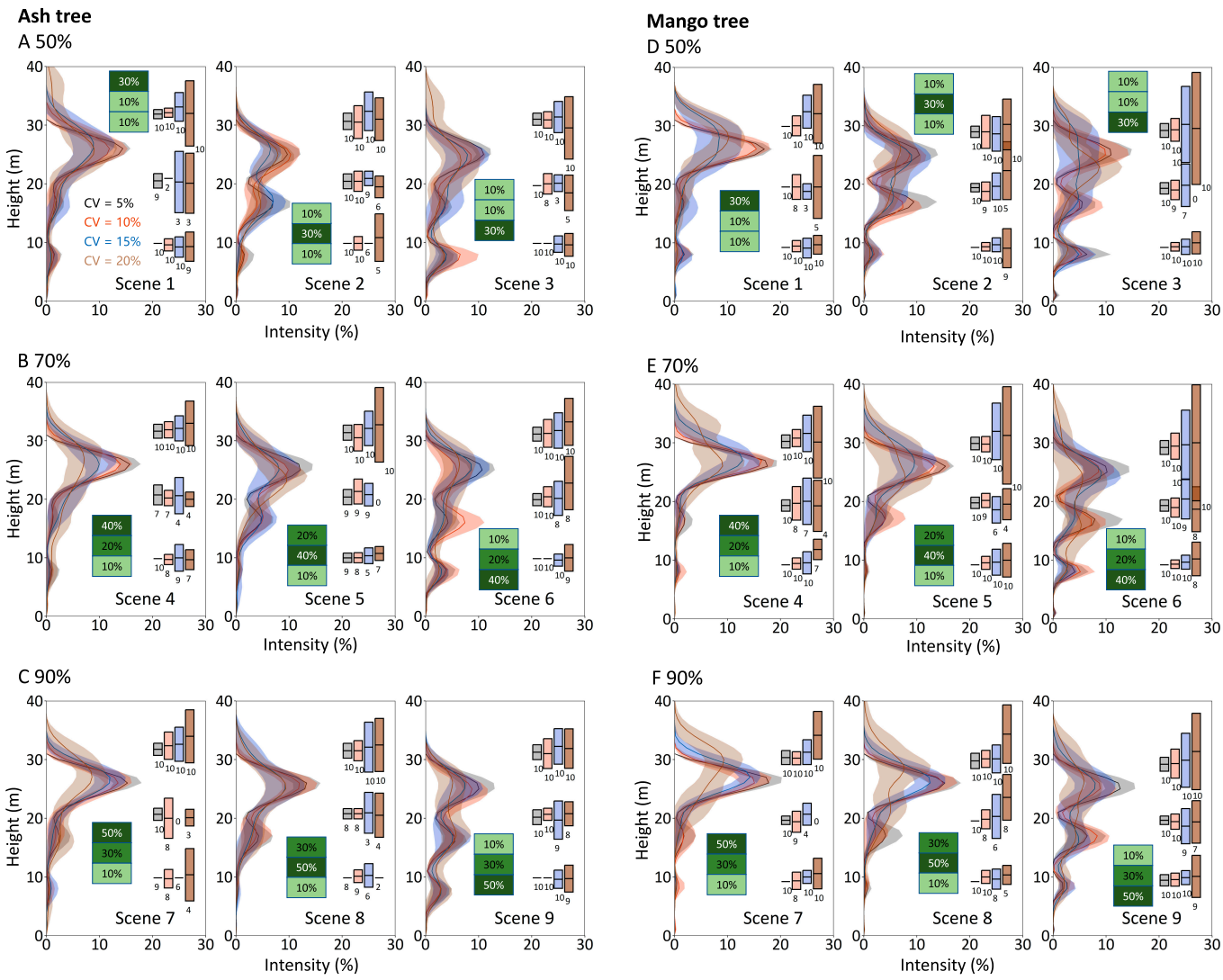

Fig. 10. Diagram of GEDI waveforms in the ash tree (A to C) and mango forests (D to F) with different tree covers [50% in panels (A) and (D), 70% in panels (B) and (E), and 90% in the panels (C) and (F)]. The height interval of different layers is 10 m. The crown cover of each layer is labeled in the figures. Other symbols have the same meaning in Fig. 8.

The waveform responses to mixing crown

Our results demonstrate that even when crown layers consist of 2 different CAs, the resulting waveforms generally exhibit 3 distinct peaks, albeit with a higher degree of uncertainty compared to waveforms obtained from homogeneous forests (Fig. 11). The inclusion of mango trees into ash tree forests leads to limited variations in the width of each peak, since the difference of CHR between the 2 species is limited. Moreover, when the height CV exceeds 15%, the peak of waveforms from some scenes diminishes, mirroring the observations in pure forests.

Notably, inferring the height retrieval functions fitted in pure forests (uniform CA metrics) into mixing forests introduces a greater level of uncertainty in height retrieval than in pure forests, compared with Fig. 8. These findings highlight that the proportion of tree species with distinct crown shapes also influences waveform characteristics and may impact the accurate retrieval of tree height composition. The mixture of different CAs results in a weaker relationship between HoI values and tree heights, thus introducing greater uncertainty in height retrieval. In forest environments where diverse tree species coexist, significant differences in CA metrics may pose challenges in accurately detecting the height composition.

Height retrieval likelihood

We have compiled metrics to evaluate the likelihood of using HoIs in height retrieval (Fig. 12). Our findings indicate that in forests with a height CV under 10%, HoIs are reliably detected, enabling accurate height retrieval. However, as CV increases, the detection rate of HoIs decreases markedly. The success of detection is primarily dependent on the spacing between tree layers. For instance, a 15-m interval between layers yields a success rate exceeding 90% (see the first column in Fig. 12A). If this interval is reduced, the success rate drops significantly (Fig. 12A and B), with intervals less than 5 m resulting in success rates below 50%. Additionally, the formation of HoIs is closely linked to the amount of radiative energy received; thus, a lower height for the densest layer tends to improve the success rate (Fig. 12C).

Height retrieval using real GEDI observation

The analysis reveals that tree heights for both top and bottom layers can be reasonably predicted from HoIs with an acceptable confidence (R^2 values of 0.39 and 0.43, respectively; see Fig. 13). The waveform peaks, originating from the top heights

Downloaded from https://spj.science.org on April 15, 2024

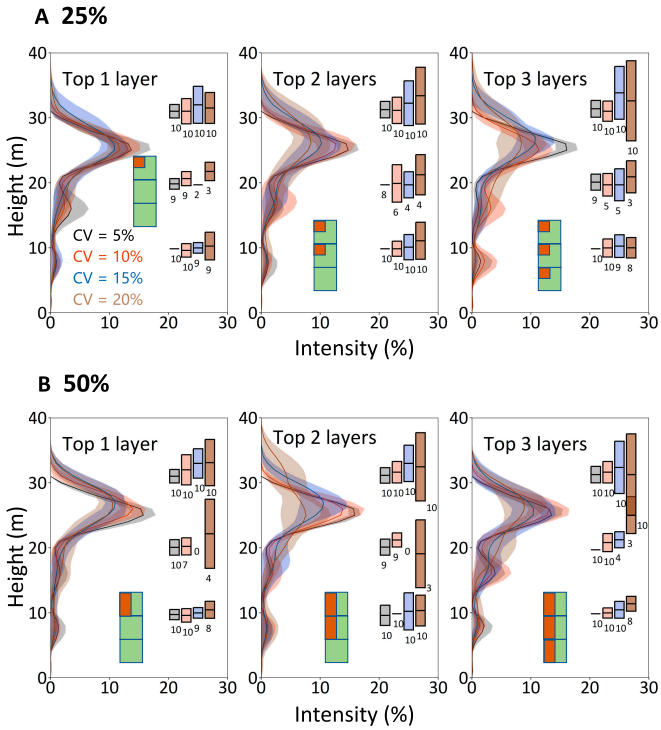

Fig. 11. Simulated GEDI waveforms in mix forests. The original experimental setup consisted of a 3-layered ash tree forest with heights of 10 m, 20 m, and 30 m with the same crown cover of 30%. A gradually increasing proportion of 25% and 50% [panels (A) and (B)] of mango trees are introduced into the original scenario at top 1, top 2, and all 3 layers, respectively (exhibited in the first, second, and third column, respectively). The mixing strategy is visually depicted by an orange block. Other symbols have the same meaning in Fig. 8.

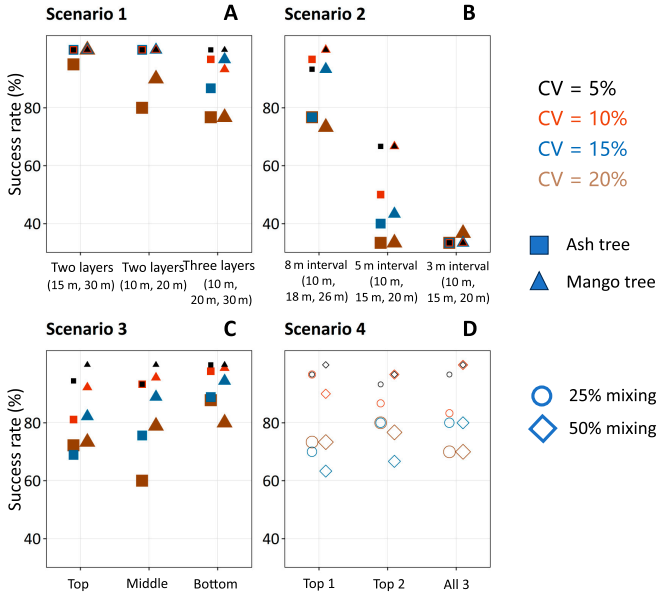

Fig. 12. A synthesis of HoI detection rate across 4 scenarios in Experiment 3. The *x*-axis in each panel delineates different scene configurations: layer combination on panel (A), interval between layers in panel (B), the position of the densest layer in panel (C), and the position of mixing layers in panel (D). The *y*-axis is the success rate of HoI detection. CVs are represented by corresponding colors and increasing size of symbols, to avoid overlapping. The solid block and triangle represent the ash tree and mango tree, respectively, in panels (A), (B), and (C). The hollow circle and rhombus represent the 25% and 50% mixing proportion in panel (D).

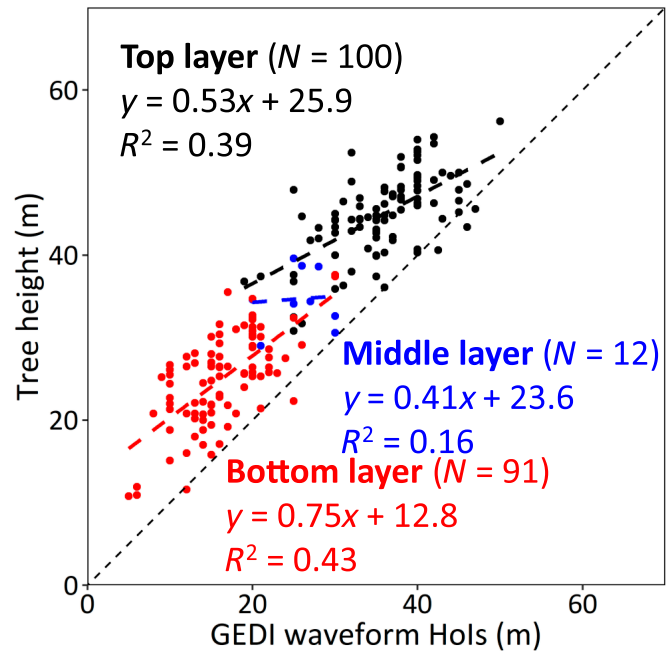

Fig. 13. Scatter plot between HoIs from GEDI waveforms and tree heights from CHM. Top, middle, and bottom layer of the canopy are represented by different colors. Statistical information (sample counts *N* and R^2) and linear fitting functions are also labeled in corresponding colors.

of each layer, reflect the corresponding crown layer heights (Fig. 14A). However, due to geolocation mismatches in GEDI observations, the tree heights in 43% of the GEDI samples did not align with those from the CHM, exhibiting an absolute bias greater than 10 m. Furthermore, there were instances where HoIs of non-top layers were not detected due to the tree height variation within a specific layer, especially in the middle layer, resulting in a decreased success rate in height retrieval for these layers. As shown in the "The waveform responses to tree height variation in multi-height-layer forests" section, layers with a narrow height range often failed to be distinguished in the HoI analysis. This issue is exemplified in Fig. 14B, where, at a specific footprint, tall trees (height > 45m) created 2 waveform peaks, while only one prominent peak was observed below 30 m due to the small height difference between several ~30-m trees (<5 m).

Discussion

Applying LESS in LiDAR signal simulation

This study aims to answer the question of whether and to what extent large footprint LiDAR waveforms can reflect the height composition. Employing RTMs significantly reduces the cost of collecting LiDAR signals from airborne remote sensing and can simulate GEDI waveforms from more comprehensive forest scenarios [27,35,47]. As a state-of-the-art RTM, LESS employs a backward path tracing and parallel computing technique and thus can greatly reduce memory usage compared to forward path tracing models and enhance computational efficiency [38]. For example, in this study, a simulation over a forest scene with 100 pre-constructed 30-m-high trees costs 102.4 ± 2.4 s on a 16-core, 2.20-GHz CPU (10 simulations). Furthermore, this efficiency can be further improved by incorporating graphic

Downloaded from https://spj.science.org on April 15, 2024

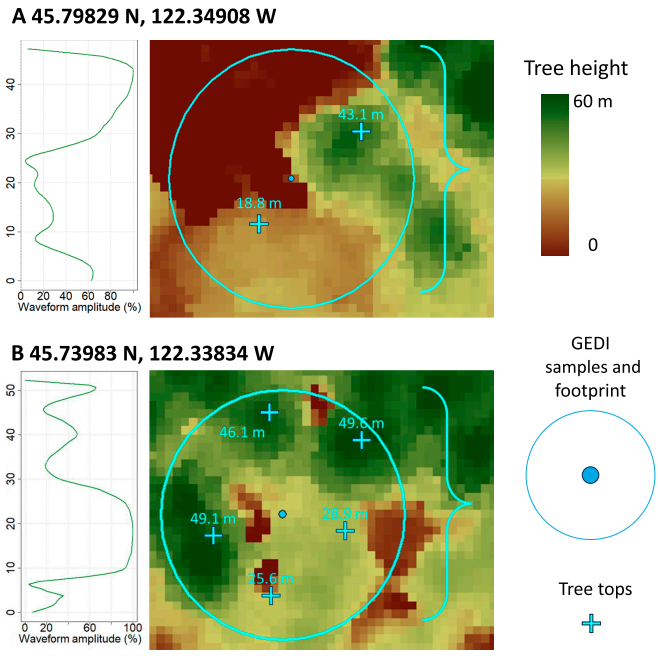

Fig. 14. Two examples (A and B) for GEDI waveforms (left panel) and corresponding footprint CHM (right panel). The tree tops within the GEDI footprint are labeled in the right panels. The positions of these 2 examples are labeled on each figure.

Table 2. Functions for estimating the tree heights of ash tree and mango tree forests

Species	Height function
Ash tree	Height = 1.23 × HoI
Mango tree	Height = 1.15 × HoI

processing units into calculation [48]. To reduce the cost of generating tree objects with customized boundaries, we also design a novel and efficient method, TAG, based on the AGG function. TAG focuses on customizing the outline of each individual tree by a series of CA metrics and can reflect the continuous variation in crown shape, which opposes the S-crown method that employs geometric primitives [39]. The leaves are presented by triangle meshes and are then randomly generated and populated within the crown outline, i.e., follows the M-surface method. The acceptable consistency in Experiment 1 demonstrates that this method can be an ideal substitution for commercial software to customize tree objects. It also reduces the calculation cost by multi-processing: the generation of each single tree occupies only one CPU core; thus, the preparation of the tree object inventory (over 10,000 trees in this study) can be accelerated by employing computing clusters.

The combination of satisfactory model performance and acceptable computational efficiency is crucial for generating a large number of waveforms from diverse forest scenes [49]. A machine learning algorithm that aims to detect forest vertical structure would require an extensive training set that includes diverse forest scenarios [24]). Commercial software often faces challenges when generating tens or even hundreds of tree objects within each LiDAR footprint, which makes it an impractical and costly option to be applied in a wide range [8]. The TAG method presented in this study provides a promising alternative, which can efficiently capture the outline of each tree and produce simulation results that are comparable to commercial software. This method enables us to build an inventory that includes over 1,000,000 tree objects, covering the continuous variation in tree height and crown area that cannot be achieved by other methods aside from our AGG-based method. Although some trees in our study may not exist in nature, this diverse inventory is beneficial for training deep learning networks. Future GEDI simulations that involve more complex forest scenes, such as those with multilayer trees featuring diverse crown areas, can utilize these objects directly.

Implication of this study

HoI as height retrieval feature

In this study, we investigate the applicability and performance of height composition retrieval within diverse forest scenarios by employing HoI as the primary input feature for height retrieval. HoI represents a local maximum of canopy cover along the vertical axis and provides information about peak width and peak intensity of LiDAR waveforms [50]. Consequently, it captures a significant portion of full-waveform shapes [51]. Traditional percentile-based strategies commonly used for top tree height retrieval [12] cannot reflect the tree heights within the second or third layers, leading to biases when estimating AGB or other forest structural traits [52]. Meanwhile, HoI serves as a more robust feature for detecting tree heights within non-top layers. Adams [50] suggested that HoIs exhibit a strong relationship with tree heights through a comprehensive evaluation. The simulation results from Experiment 2 demonstrates that the relationship between HoI and tree heights responds to the variation of CA metrics confidently, and the results from Experiment 3 demonstrates that HoI-height functions are robust when being inferred into the multi-layer forest scenarios. Results from Experiment 4 support the findings from simulation experiments and further demonstrate the applicability of employing HoIs in height retrieval.

It is essential to clarify that in certain scenarios, such as those with significant height CV within each layer, there might be difficulties in detecting HoIs for adjacent layers and consequently in retrieving tree heights (see the examples in the "The waveform responses to tree height variation in multi-height-layer forests" and "Height retrieval using real GEDI observation" sections, especially the limited success rate of retrieving the tree height of the middle layer in Fig. 14). However, this does not imply an inability to reflect tree heights or height compositions. Our current method primarily focuses on identifying discrete waveform peaks, overlooking the waveform broadening caused by adjacent tree layers. Future algorithms, possibly leveraging deep learning on series data, may offer valuable insights.

Height retrieval in multilayer forests

Results from Experiment 3 demonstrate that GEDI waveforms can effectively reflect and retrieve the tree height within multilayer forests, and the capacity of the waveforms to capture the multi-layer height structure is influenced by both the CA characteristics and the interval between different layers. Specifically, the height of different layers in fir forests cannot be reliably

Downloaded from https://spj.science.org on April 15, 2024

detected, and in ash and mango tree forests, layers with a height interval of less than 5 m may not be effectively distinguished. As demonstrated in the "The waveform responses to different height intervals" section, the height functions can accurately detect the tree height composition even with varying height intervals, as long as a greater than 5-m interval between layers is present to generate waveform peaks. It is crucial to acknowledge that the complexity of the canopy can adversely affect the retrieval performance of GEDI waveforms. Typically, in a homogeneous forest with consistent CAs, the height of each layer can be confidently retrieved using HoIs when there is no more than a 10% height variation. However, due to factors such as interspecies competition and intraspecies phenotypic diversity, it is rare to encounter a forest where all crowns possess identical CAs. A comparison between Figs. 8 and 11 reveals that a 25% mixing proportion negatively impacts the successful detection of HoIs, subsequently escalating the uncertainty in height retrieval. An increase in the mixing proportion further amplifies these retrieval uncertainties.

We consolidated the success rates of HoI detection across 4 scenarios in Experiment 3 (see the "Height retrieval likelihood" section), highlighting that under certain conditions, height retrieval accuracy can improve even without HoI detection. The results reveal that HoI detection remains consistent in environments with minimal height variation (e.g., CV < 10%). Additionally, a wider interval between tree layers improves the likelihood of HoI detection, as shown by the resulting trend of reducing height interval from 15 to 3 m. However, when the interval is less than 5 m, the probability of detecting HoIs for non-top layers is considerably reduced.

A major contribution of the numerical simulations (Experiments 1 to 3) in this study is their role in guiding the retrieval of height composition using real GEDI observations, which is the primary focus of Experiment 4. In this experiment, we employed actual GEDI observations to support our numerical simulation conclusions. The findings align with our previous results, confirming that waveform HoIs can effectively discern tree heights in both top and sub-canopy layers, thus accurately representing tree height composition. Additionally, our hypothesis that layers with a height difference of less than 5 m may be indistinguishable is supported by the actual GEDI data (Fig. 14). This outcome reinforces the value of RTM-based simulations in providing insights for future use of GEDI waveforms to determine height compositions. It highlights the importance of considering variables such as height variability, height intervals, crown coverage, and the presence of mixed forest types.

Crown cover retrieval

The findings presented in the "The waveform responses to different crown cover" section demonstrate the effectiveness of HoI-height functions in forest scenarios where there are variations in crown coverage between layers. However, it is important to note that the variations in crown coverage within different layers may not always be accurately reflected in this study, primarily because the top layers tend to intercept a greater amount of laser energy. Consequently, the current method enables us to detect the presence of trees within each layer rather than providing precise information regarding the amount or coverage of trees. Tang [15] proposed a method for retrieving the vertical leaf area index (LAI) profile based on the gap theory [16,53]: the gap probability at a particular height is inversely related to the LAI, as laser energy can only penetrate that height through gaps [54]. This approach enables a confident retrieval of the vertical LAI profile covering wide regions [55]. By incorporating certain prior knowledge, such as CA metrics describing the single crown coverage and total crown coverage from optical imagery, it becomes promising to retrieve more quantitative information about tree height composition.

Uncertainties

In the analysis, we mainly concentrate on the waveform responses to forest structural traits and to diverse height compositions, thus leaving several uncertainties that might influence the results. When using LESS or other RTMs to simulate waveforms, the position of each tree within the forest must be specified [27,31,56]. Our investigation reveals that the positional uncertainty of trees cannot be overlooked, particularly when the crown cover is below 50%. This uncertainty arises due to the non-uniform distribution of energy within the horizontal plate of the footprint (Fig. S2). However, the tree position-induced uncertainty influences the intensity, rather than the HoI of waveforms (Fig. S3). Additionally, we only consider a spherical distribution for leaf angle distribution (LAD). One advantage of full waveform is that it captures the normalized frequency, rather than the absolute amount, of discrete cloud points from Gaussian decomposition, indicating its lower sensitivity to the probability of scattering and more sensitivity to the crown cover of each height slice. The uncertainties introduced by different LADs settings are comparable to the uncertainty caused by tree positions (Fig. S4). Finally, leaf biochemical traits, such as chlorophyll content, carotenoids, and mesophyll structure, can affect the optical properties of leaves, but these factors are not considered in current experiments. The variation in these traits can alter the optical properties of leaves and subsequently affect the transfer process of lasers and waveforms [35,57]. Our simulation indicates that changes in leaf reflectance and transmittance have a minimal impact on simulated waveform (Fig. S5). This is due to their impact on leaf scattering and overall LiDAR returns throughout the canopy, as observed by Yang et al. [58].

The influence of under-canopy topography, particularly slope, within the LiDAR footprint, is currently overlooked in this study. Terrestrial slope impacts the GEDI signal in 2 significant ways. In a genuine large footprint LiDAR system, the received laser energy comprises multiple components: direct reflectance from the terrain and canopy, canopy-diffused energy, and terrain-reflected energy further diffused by the canopy. Notably, terrain slope influences the proportion of received laser energy, altering the overall waveform shape. An increase in terrain slope broadens both the vegetation and ground sub-waveforms [35], significantly affecting wood volume retrieval performance [33,59]. Sloped terrains also introduce a time decay in energy transfer, causing inevitable height variations within homogeneous tree height forest stands. This variation affects the accuracy of retrieved tree heights and height compositions [35]. This study primarily aims to explore the GEDI waveform's potential in reflections where terrain effects are not as pronounced as in wood volume estimation studies [59]. Nevertheless, further research is still essential to quantify terrestrial slope-induced effects and ascertain their manifestation in actual full LiDAR waveforms.

A challenge in Experiment 4 incorporating real GEDI observation is the geolocation bias of GEDI observations, typically exceeding 20 m, leading to a considerable likelihood of mismatches between GEDI-derived tree heights and those from

Downloaded from https://spj.science.org on April 15, 2024

the CHM (43%). This issue also poses a constraint for footprint-scale GEDI-based studies. Beyond the current filtering strategy, implementing a post-processing approach or using an updated product could substantially reduce this adverse impact [60]. Additionally, in this experiment, we manually differentiated tree layers without a strict description of tree height composition within the GEDI footprint. For a more thorough analysis, the application of an automatic tree crown segmentation method would be beneficial.

Future work

In this study, we demonstrate that height composition can be detected by full LiDAR waveforms since the response of the waveforms against the variation in tree metrics and tree height are all shown to be predictable. Future retrieval can be carried out under the help of sequence-based machine learning algorithms, such as the 1D convolutional neural network. The numerical simulation strategy employed in this study demonstrates valuable potential to save the cost of field measurements. The considerable efficiency of TAG and LESS enables the generation of a sufficient training set from diverse forest scenarios for pre-training a complex network. A fine-tuning step transferring this pre-trained network can significantly improve the training efficiency and model performance [61].

Several ecological studies can benefit from a more detailed and accurate understanding of tree height composition. Firstly, an improved estimation of AGB in secondary or managed forests, where trees emerge following natural or human disturbances, can be achieved by considering different height layers [7]. One significant issue that hampers satisfactory AGB estimation in secondary forests is that a considerable portion of AGB (30% to 50%) is hidden by non-top layer trees that are often neglected [62]. Secondly, several recent gridded stand age products cover large regions [63], but only few of them incorporate tree height information as input features, thereby neglecting the intrinsic nexus with each other [64] and its primary contribution to stand age estimation [13]. The current description of height composition undoubtedly can constrain the forest age predictions, thereby enhancing our understanding of forest demography and its role in carbon sequestration [65].

Conclusion

The aim of this study is to investigate whether GEDI waveforms can reflect the tree height composition. We utilize a state-of-the-art RTM, LESS, to simulate discrete point clouds from forest stands and propose a method, TAG, based on the AGG function to create tree objects with continuous variation in CAs while maintaining acceptable efficiency. The GEDI waveforms are aggregated from discrete ALS point cloud considering the energy distribution on horizontal and vertical dimensions by GEDI simulator. Our findings suggest that tree objects can effectively replace virtual tree objects for commercial usage in waveform simulations. In even-aged forest scenes, GEDI waveforms exhibit reasonable responses to CA metrics. In multi-height-layer forest scenes, the shape of the waveforms is determined by the tree height composition and CA metrics. HoIs can be used to effectively retrieve the tree height of each layer. The variations in tree height influence the GEDI waveforms and, thus, the retrieval performance of the height within different layers. A greater than 10% of height CV causes a considerable likelihood of failure to retrieve the height composition. GEDI waveforms can be utilized to identify and retrieve the height of different crown layers with layer intervals greater than 5 m and with limited tree height variation, but cannot confidently project the difference in crown cover between different layers. Finally, a mixing of CAs influences the waveform characteristics and height composition retrieving. Beyond numerical simulation, an experiment integrating actual GEDI observations confirms the satisfactory performance of tree height retrieval. This underscores the vital role of simulation in directing remote sensing applications. Overall, our study highlights the promising potential of using GEDI waveforms to detect the height composition and within-canopy structures. The results of the current study provide necessary guidance for retrieving tree height composition using real GEDI observations covering more comprehensive forest scenarios in the near future.

Acknowledgments

Funding: This study is supported by the National Science Foundation of China (42141005). This work is also supported by the High-performance Computing Platform of Peking University.
Author contributions: S.T.: Methodology, data acquisition, formal analysis, and writing—original draft. Y.Z.: Conceptualization, formal analysis, writing—review and editing, and funding acquisition. J.Q.: Methodology and writing—review and editing. Y.S.: Writing—review and editing and investigation. Q.M.: Writing—review and editing. J.Q.: Writing—review and editing.
Competing interests: The authors declare that they have no competing interests.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary Materials

Figs. S1 to S5

References

1. Vierling KT, Vierling LA, Gould WA, Martinuzzi S, Clawges RM. Lidar: Shedding new light on habitat characterization and modeling. *Front Ecol Environ*. 2008;6:90–98.
2. Antonarakis AS, Saatchi SS, Chazdon RL, Moorcroft PR. Using Lidar and radar measurements to constrain predictions of forest ecosystem structure and function. *Ecol Appl*. 2011;21:1120–1137.
3. Stark SC, Leitold V, Wu JL, Hunter MO, de Castilho CV, Costa FR, McMahon SM, Parker GG, Shimabukuro MT, Lefsky MA, et al. Amazon forest carbon dynamics predicted by profiles of canopy leaf area and light environment. *Ecol Lett*. 2012;15(12):1406–1414.
4. Walker AP, Hanson PJ, De Kauwe MG, Medlyn BE, Zaehle S, Asao S, Dietze M, Hickler T, Huntingford C, Iversen CM. Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO_2 enrichment experiments: Model performance at ambient CO_2 concentration. *J Geophys Res Biogeo*. 2014;119(5):937–964.
5. Braghiere RK, Wang Y, Doughty R, Sousa D, Magney T, Widlowski J-L, Longo M, Bloom AA, Worden J, Gentine P, et al.

Downloaded from https://spj.science.org on April 15, 2024

Accounting for canopy structure improves hyperspectral radiative transfer and sun-induced chlorophyll fluorescence representations in a new generation earth system model. *Remote Sens Environ*. 2021;261:Article 112497.
6. Tang X, Zhao X, Bai Y, Tang Z, Wang W, Zhao Y, Wan H, Xie Z, Shi X, Wu B, et al. Carbon pools in China's terrestrial ecosystems: New estimates based on an intensive field survey. *Proc Natl Acad Sci*. 2018;115(16):4021–4026.
7. Barbosa J, Broadbent E, Bitencourt MD. Remote sensing of aboveground biomass in tropical secondary forests: A review. *International Journal of Forestry Research*. 2014;2014:Article 715796.
8. Widlowski J-L, Mio C, Disney M, Adams J, Andredakis I, Atzberger C, Brennan J, Busetto L, Chelle M, Ceccherini G, et al. The fourth phase of the radiative transfer model intercomparison (RAMI) exercise: Actual canopy scenarios and conformity testing. *Remote Sens Environ*. 2015;169(1-2):418–437.
9. Ma Q, Su Y, Luo L, Li L, Kelly M, Guo Q. Evaluating the uncertainty of Landsat-derived vegetation indices in quantifying forest fuel treatments using bi-temporal LiDAR data. *Ecol Indic*. 2018;95:298–310.
10. Guo Q, Su Y, Hu T, Guan H, Jin S, Zhang J, Zhao X, Xu K, Wei D, Kelly M. Lidar boosts 3D ecological observations and modelings: A review and perspective. *IEEE Geosci Remote Sens Mag*. 2020;9(1):232–257.
11. Duncanson L, Kellner JR, Armston J, Dubayah R, Minor DM, Hancock S, Healey SP, Patterson PL, Saarela S, Marselis S, et al. Aboveground biomass density models for NASA's global ecosystem dynamics investigation (GEDI) lidar mission. *Remote Sens Environ*. 2022;270:Article 112845.
12. Potapov P, Li X, Hernandez-Serna A, Tyukavina A, Hansen MC, Kommareddy A, Pickens A, Turubanova S, Tang H, Silva CE, et al. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens Environ*. 2021;253:Article 112165.
13. Liu X, Ma Q, Wu X, Hu T, Liu Z, Liu L, Guo Q, Su Y. A novel entropy-based method to quantify forest canopy structural complexity from multiplatform lidar point clouds. *Remote Sens Environ*. 2022;282:Article 113280.
14. Hosoi F, Omasa K. Factors contributing to accuracy in the estimation of the woody canopy leaf area density profile using 3D portable lidar imaging. *J Exp Bot*. 2007;58(12):3463–3473.
15. Tang H, Dubayah R, Swatantran A, Hofton M, Sheldon S, Clark DB, Blair B. Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sens Environ*. 2012;124:242–250.
16. Chen Q, Baldocchi D, Gong P, Kelly M. Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogramm Eng Remote Sens*. 2006;72(8):923–932.
17. Zhen Z, Quackenbush LJ, Stehman SV, Zhang L. Agent-based region growing for individual tree crown delineation from airborne laser scanning (ALS) data. *Int J Remote Sens*. 2015;36(7):1965–1993.
18. Gupta S, Weinacker H, Koch B. Comparative analysis of clustering-based approaches for 3-D single tree detection using airborne fullwave lidar data. *Remote Sens*. 2010;2(4):968–989.
19. Maltamo M, Næsset E, Vauhkonen J. Forestry applications of airborne laser scanning. Concepts and case studies. *Manag For Ecosys*. 2014;27:460.
20. Wang Y, Weinacker H, Koch B. A lidar point cloud based procedure for vertical canopy structure analysis and 3D single tree modelling in forest. *Sensors*. 2008;8(6):3938–3951.
21. Heinzel JN, Weinacker H, Koch B. Prior-knowledge-based single-tree extraction. *Int J Remote Sens*. 2011;32(17):4999–5020.
22. McElhinny C, Gibbons P, Brack C, Bauhus J. Forest and woodland stand structural complexity: Its definition and measurement. *For Ecol Manag*. 2005;218(1-3):1–24.
23. Muss JD, Mladenoff DJ, Townsend PA. A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sens Environ*. 2011;115(3):824–835.
24. Lang N, Kalischek N, Armston J, Schindler K, Dubayah R, Wegner JD. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. *Remote Sens Environ*. 2022;268:Article 112760.
25. Dubayah R, Blair JB, Goetz S, Fatoyinbo L, Hansen M, Healey S, Hofton M, Hurtt G, Kellner J, Luthcke S, et al. The global ecosystem dynamics investigation: High-resolution laser ranging of the Earth's forests and topography. *Sci Remote Sens*. 2020;1(G00E09):Article 100002.
26. Roy DP, Kashongwe HB, Armston J. The impact of geolocation uncertainty on GEDI tropical forest canopy height estimation and change monitoring. *Sci Remote Sens*. 2021;4(2):Article 100024.
27. Yang X, Wang Y, Yin T, Wang C, Lauret N, Regaieg O, Xi X, Gastellu-Etchegorry JP. Comprehensive LiDAR simulation with efficient physically-based DART-lux model (I): Theory, novelty, and consistency validation. *Remote Sens Environ*. 2022;272(12):Article 112952.
28. Kobayashi H, Iwabuchi H. A coupled 1-D atmosphere and 3-D canopy radiative transfer model for canopy reflectance, light environment, and photosynthesis simulation in a heterogeneous landscape. *Remote Sens Environ*. 2008;112(1):173–185.
29. Goodenough AA, Brown SD. DIRSIG5: Next-generation remote sensing data and image simulation framework. *IEEE J Sel Top Appl Earth Observ Remote Sens*. 2017;10(11):4818–4833.
30. Gastellu-Etchegorry J-P, Yin T, Lauret N, Grau E, Rubio J, Cook BD, Morton DC, Sun G. Simulation of satellite, airborne and terrestrial LiDAR with DART (I): Waveform simulation with quasi-Monte Carlo ray tracing. *Remote Sens Environ*. 2016;184:418–435.
31. Qi J, Xie D, Yin T, Yan G, Gastellu-Etchegorry J-P, Li L, Zhang W, Mu X, Norford LK. LESS: LargE-scale remote sensing data and image simulation framework over heterogeneous 3D scenes. *Remote Sens Environ*. 2019;221:695–706.
32. Pang Y, Lefsky M, Sun G, Ranson J. Impact of footprint diameter and off-nadir pointing on the precision of canopy height estimates from spaceborne lidar. *Remote Sens Environ*. 2011;115(11):2798–2809.
33. Nie S, Wang C, Xi X, Li G, Luo S, Yang X, Wang P, Zhu X. Exploring the influence of various factors on slope estimation using large-footprint LiDAR data. *IEEE Trans Geosci Remote Sens*. 2018;56(11):6611–6621.
34. Yang W, Ni-Meister W, Lee S. Assessment of the impacts of surface topography, off-nadir pointing and vegetation structure on vegetation lidar waveforms using an extended geometric optical and radiative transfer model. *Remote Sens Environ*. 2011;115(11):2810–2822.
35. Yang X, Wang C, Xi X, Wang Y, Zhang Y, Zhou G. Footprint size design of large-footprint full-waveform LiDAR for forest and topography applications: A theoretical study. *IEEE Trans Geosci Remote Sens*. 2021;59(11):9745–9757.

Downloaded from https://spj.science.org on April 15, 2024

36. Blair JB, Hofton MA. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophys Res Lett*. 1999;26(16):2509–2512.
37. Milenković M, Schnell S, Holmgren J, Ressl C, Lindberg E, Hollaus M, Pfeifer N, Olsson H. Influence of footprint size and geolocation error on the precision of forest biomass estimates from space-borne waveform LiDAR. *Remote Sens Environ*. 2017;200:74–88.
38. Qi J, Yin T, Xie D, Gastellu-Etchegorry J-P. Hybrid scene structuring for accelerating 3D radiative transfer simulations. *Remote Sens*. 2019;11(22):2637.
39. Ni W, Li X, Woodcock CE, Caetano MR, Strahler AH. An analytical hybrid GORT model for bidirectional reflectance over discontinuous plant canopies. *IEEE Trans Geosci Remote Sens*. 1999;37(2):987–999.
40. Janoutová R, Homolová L, Malenovský Z, Hanuš J, Lauret N, Gastellu-Etchegorry J-P. Influence of 3D spruce tree representation on accuracy of airborne and satellite forest reflectance simulated in DART. *Forests*. 2019;10(3):292.
41. Su Y, Hu T, Wang Y, Li Y, Dai J, Liu H, Jin S, Ma Q, Wu J, Liu L, et al. Large-scale geographical variations and climatic controls on crown architecture traits. *Biogeosciences*. 2020;125(2): e2019JG005306.
42. Luo Y, Xie D, Qi J, Zhou K, Yan G, Mu X. LESS LiDAR: A full-waveform and discrete-return multispectral LiDAR simulator based on ray tracing algorithm. *Remote Sens*. 2023;15(18):4529.
43. Hancock S, Armston J, Hofton M, Sun X, Tang H, Duncanson LI, Kellner JR, Dubayah R. The GEDI simulator: A large-footprint waveform lidar simulator for calibration and validation of spaceborne missions. *Earth and Space Sci*. 2019;6(2):294–310.
44. Qi J, Xie D, Jiang J, Huang H. 3D radiative transfer modeling of structurally complex forest canopies through a lightweight boundary-based description of leaf clusters. *Remote Sens Environ*. 2022;283:Article 113301.
45. Zhou T, Popescu SC. Bayesian decomposition of full waveform LiDAR data with uncertainty analysis. *Remote Sens Environ*. 2017;200:43–62.
46. Barbeito I, Dassot M, Bayer D, Collet C, Drössler L, Löf M, del Rio M, Ruiz-Peinado R, Forrester DI, Bravo-Oviedo A, et al. Terrestrial laser scanning reveals differences in crown structure of *Fagus sylvatica* in mixed vs. pure European forests. *For Ecol Manag*. 2017;405:381–390.
47. Li C, Yu Z, Wang S, Wu F, Wen K, Qi J, Huang H. Crown structure metrics to generalize aboveground biomass estimation model using airborne laser scanning data in National Park of Hainan tropical rainforest, China. *Forests*. 2022;13(7):1142.
48. Bian Z, Qi J, Gastellu-Etchegorry J-P, Roujean J-L, Cao B, Li H, Du Y, Xiao Q, Liu Q. A GPU-based solution for ray tracing 3-D radiative transfer model for optical and thermal images. *IEEE Geosci Remote Sens Lett*. 2022;19:1–5.
49. Li L, Mu X, Chianucci F, Qi J, Jiang J, Zhou J, Chen L, Huang H, Yan G, Liu S. Ultrahigh-resolution boreal forest canopy mapping: Combining UAV imagery and photogrammetric point clouds in a deep-learning-based approach. *Int J Appl Earth Obs Geoinf*. 2022;107:Article 102686.
50. Adams T, Beets P, Parrish C. Extracting more data from LiDAR in forested areas by analyzing waveform shape. *Remote Sens*. 2012;4(3):682–702.
51. Wagner W, Ullrich A, Ducic V, Melzer T, Studnicka N. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS J Photogramm Remote Sens*. 2006;60(2):100–112.
52. Francini S, D'Amico G, Vangi E, Borghi C, Chirici G. Integrating GEDI and Landsat: Spaceborne lidar and four decades of optical imagery for the analysis of forest disturbances and biomass changes in Italy. *Sensors*. 2022;22(5):2015.
53. Chen JM, Cihlar J. Plant canopy gap-size analysis theory for improving optical measurements of leaf-area index. *Appl Opt*. 1995;34(27):6211–6222.
54. Ni-Meister W, Jupp DL, Dubayah R. Modeling lidar waveforms in heterogeneous and discrete canopies. *IEEE Trans Geosci Remote Sens*. 2001;39(9):1943–1958.
55. Dubayah R, Tang H, Armston J, Luthcke S, Hofton M, Blair J. *GEDI L2B canopy cover and vertical profile metrics data global footprint level V001*. Washington, DC: NASA; 2020.
56. Wang Y, Kallel A, Yang X, Regaieg O, Lauret N, Guilleux J, Chavanon E, Gastellu-Etchegorry J-P. DART-lux: An unbiased and rapid Monte Carlo radiative transfer method for simulating remote sensing images. *Remote Sens Environ*. 2022;274:Article 112973.
57. Sun J, Shi S, Yang J, Gong W, Qiu F, Wang L, Du L, Chen B. Wavelength selection of the multispectral lidar system for estimating leaf chlorophyll and water contents through the PROSPECT model. *Agric For Meteorol*. 2019;266:43–52.
58. Yang X, Wang C, Xi X. Multiple scattering effect on forest physiological parameters of multi-spectral LiDAR canopy waveforms. In: *Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE; 2019. p. 8467–8469.
59. Fayad I, Baghdadi N, Alcarde Alvares C, Stape JL, Bailly JS, Scolforo HF, Cegatta IR, Zribi M, Le Maire G. Terrain slope effect on forest height and wood volume estimation from GEDI data. *Remote Sens*. 2021;13(11):2136.
60. Tang H, Stoker J, Luthcke S, Armston J, Lee K, Blair B, Hofton M. Evaluating and mitigating the impact of systematic geolocation error on canopy height measurement performance of GEDI. *Remote Sens Environ*. 2023;291:Article 113571.
61. Li R, Wang D, Liang S, Jia A, Wang Z. Estimating global downward shortwave radiation from VIIRS data using a transfer-learning neural network. *Remote Sens Environ*. 2022;274(10):Article 112999.
62. Broadbent EN, Asner GP, Peña-Claros M, Palace M, Soriano M. Spatial partitioning of biomass and diversity in a lowland Bolivian forest: Linking field and remote sensing measurements. *For Ecol Manag*. 2008;255(7): 2602–2616.
63. Maltman JC, Hermosilla T, Wulder MA, Coops NC, White JC. Estimating and mapping forest age across Canada's forested ecosystems. *Remote Sens Environ*. 2023;290(2882-2897): Article 113529.
64. Su R, Wu Q, Yang Y, Hu T. Relationship between diameter at breast height and tree age in populations of a rare and endangered plant. *Polish J Ecol*. 2021;69:84–95.
65. Cook-Patton SC, Leavitt SM, Gibbs D, Harris NL, Lister K, Anderson-Teixeira KJ, Briggs RD, Chazdon RL, Crowther TW, Ellis PW, et al. Mapping carbon accumulation potential from global natural forest regrowth. *Nature*. 2020;585(7826): 545–550.

Downloaded from https://spj.science.org on April 15, 2024