

Progress on methods for waveform processing and target feature inversion based on full-waveform LiDAR

Jiajie Fang, Yongxiang Wang, Liang Shi, Xurong Duan, Shilong Xu, Nanxiang Zhao, Yuhao Xia, Xiao Dong, Xinyuan Zhang and Yihua Hu

Citation: Fang JJ, Wang YX, Shi L, et al. Progress on methods for waveform processing and target feature inversion based on full-waveform LiDAR. *Opto-Electronics Plus* 2, 260001(2026).

Received: 4 January 2026; Accepted: 10 February 2026; Published online: 31 March 2026

Related articles

Optical addressing enables a new architecture for spatial light modulators

Xiangyu Huang, Na Liu

Opto-Electronic Advances 2026, 9(3): 260049 doi: [10.29026/oea.2026.260049](https://doi.org/10.29026/oea.2026.260049)

Narrow beam and low-sidelobe electro-optic beam steering on thin-film lithium niobate optical phased array

Yang Li, Shiyao Deng, Xiao Ma, Ziliang Fang, Shufeng Li, Weikang Xu, Fangheng Fu, Xu Ouyang, Yuming Wei, Tiefeng Yang, Heyuan Guan, Huihui Lu

Opto-Electronic Science 2026, 5(2): 260002 doi: [10.29026/oes.2026.260002](https://doi.org/10.29026/oes.2026.260002)

Parallel all-optical encoded CDMA-driven anti-interference LiDAR for 78 MHz point acquisition

Shujian Gong, Peng Tian, Yinghui Guo, Xiaoyin Li, Mingbo Pu, Qi Zhang, Yanqin Wang, Heping Liu, Xiangang Luo

Opto-Electronic Technology 2025, 1(3): 250005 doi: [10.29026/oet.2025.250005](https://doi.org/10.29026/oet.2025.250005)

Fast three-dimensional scanning imaging using single pixel single photon lidar under extremely low signal to noise ratio conditions

Yun Jiang, Zixun Wang, Bo Liu, Taibei Liu, Feifan He, Lan Luo, Yongqi Yang, Fengyun Huang, Bin Zhao

Intelligent Opto-Electronics 2025, 1(2): 250013 doi: [10.29026/ioe.2025.250013](https://doi.org/10.29026/ioe.2025.250013)

More related articles in Opto-Electronic Journals Group website 



Progress on methods for waveform processing and target feature inversion based on full-waveform LiDAR

Jiajie Fang^{1,2,3}, Yongxiang Wang^{1,2}, Liang Shi^{1,4*}, Xurong Duan^{1,2}, Shilong Xu^{1,2}, Nanxiang Zhao^{1,2}, Yuhao Xia^{1,2}, Xiao Dong^{1,2}, Xinyuan Zhang^{1,2,5*} and Yihua Hu^{1,2*}

Abstract: Full-waveform LiDAR employs detectors with high-speed sampling and storage capabilities, expanding the information dimensions of signals by completely recording the return waveforms. In earth observation, both terrain and vegetation structural parameters (e.g., canopy height, canopy cover, leaf area index (LAI), and foliage height diversity) modulate the laser echoes, endowing this technology with the unique capability to perceive three-dimensional fine-scale surface structures and retrieve key ecological parameters. This technical advantage provides unprecedented opportunities for earth observation, significantly advancing our understanding of ecosystem function and architecture. This review systematically summarizes recent progress in full-waveform LiDAR technology, focusing on processing of echo waveform and retrieval of ecological information. Waveform processing covers waveform simulation, analytical formula, and waveform decomposition and waveform deconvolution. Retrieval methods of ecological information include waveform simulation, analytical formula, waveform decomposition, and waveform metrics. It elaborates on the origins of these methods, and their latest applications in extracting ecological parameters such as forest canopy height, aboveground biomass (AGB), LAI, clumping index, structural complexity, fuel load and canopy bulk density, burn severity, and ecosystem productivity. Finally, this review summarized the advantage and disadvantages of different methods, outline future research directions for full-waveform LiDAR feature retrieval, with particular emphasis on target identification based on waveform attributes and its emerging applications in spatial/spaceborne remote sensing.

Keywords: full-waveform LiDAR; waveform simulation; waveform decomposition; target feature inversion; relative height; aboveground biomass

DOI: [10.67704/oep.2026.260001](https://doi.org/10.67704/oep.2026.260001)

Citation: Fang JJ, Wang YX, Shi L et al. Progress on methods for waveform processing and target feature inversion based on full-waveform LiDAR. *Opto-Electron Plus* 2, 260001 (2026).

1 Introduction

High-precision remote sensing detection and identification technology supports various fields including ecological security, resource exploration, disaster warning, topographic mapping, target recognition and classification, precision guidance, and deep space exploration¹⁻¹¹. It represents a critical strategic requirement for national aerospace security and cutting-edge technological development. Among various remote sensing techniques, full-waveform LiDAR stands

out as particularly distinctive. Originating from laser ranging technology, it is capable of capturing both the three-dimensional (3D) structure and related properties of a target in a single observation^{5,6}. In laser ranging, a 1 ns echo time delay corresponds to a 15 cm laser travel distance. When measuring targets such as forest canopies or buildings, it is necessary to compare two independent ranging measurements to obtain vertical height information. Its resolution is independent of detection distance and determined solely by

Received: 4 January 2026

Accepted: 10 February 2026

Published online: 31 March 2026

¹State Key Laboratory of Pulsed Power Laser Technology, National University of Defense Technology, Hefei 230037, China;

²Advanced Laser Technology of Anhui Province, National University of Defense Technology, Hefei 230037, China; ³Nanhu Laser Laboratory, National University of Defense Technology, Hefei 230037, China; ⁴China Certification and Inspection Group Tanovo Information Technology (Anhui) Limited Company, Hefei 230061, China; ⁵University of Chinese Academy of Sciences, Hangzhou Institute for Advanced Study, Hangzhou 310024, China.

*Correspondence: L Shi, E-mail: shi1983218@126.com; XY Zhang, E-mail: skl_zxy@163.com; YH Hu, E-mail: skl_hyh@163.com

system parameters such as laser pulse width and detector bandwidth. While full-waveform LiDAR employs detectors with high-speed sampling and storage capabilities to digitally record the energy-time variation curves, i.e., echo waveforms, modulated by targets at different locations within the laser footprint, as shown in Fig. S1. Multi-dimensional target characteristics such as 3D structure, length, orientation, position, and surface optical properties collectively influence the echo waveform. By analyzing features such as peak intensity, pulse width, and waveform shape, it is possible to extract deep information beyond distance, including 3D geometry, attitude, and material properties. It thus achieves a technological leap from inferring elevation via multiple measurements to retrieving multi-dimensional attributes in a single acquisition.

Over the past two decades, full-waveform LiDAR has evolved into an efficient, high-resolution remote sensing technique for capturing multi-dimensional properties of targets^{5,6,12–15}. In the domain of spaceborne earth observation, NASA launched the Ice, Cloud, and land Elevation Satellite (ICESat) in 2003, which carried the Geoscience Laser Altimeter System (GLAS) designed to quantify ice sheet mass balance, assess impacts of atmospheric and climate changes on polar ice and global sea levels, and map land topography, sea ice, and vegetation cover¹². Furthermore, in 2018, NASA deployed the Global Ecosystem Dynamics Investigation (GEDI) mission to access 3D forest structure, analyze the effects of disturbances on forests, model future forest dynamics, and map the spatiotemporal distribution of habitat structure and its influence on habitat quality and biodiversity^{5,13}. More recently, China has launched several satellites equipped with full-waveform LiDAR payloads, including the ZY-3 03, GaoFen-7, and the Terrestrial Ecosystem Carbon Inventory satellite (TECIS)^{6,14}. These developments mark a period of vigorous growth in spaceborne full-waveform LiDAR for earth observation.

In contrast to passive optical and SAR imaging, which reconstruct target features directly from images, full-waveform LiDAR retrieves target characteristic parameters by interpreting the features of laser echo waveforms. Even after removing interference from factors such as atmospheric scattering, background light noise and detector thermal noise, multiple characteristics of the target, including the 3D structure, size, position, orientation, and surface optical properties, collectively influence the laser echo waveform^{16–22}. Moreover, different characteristics affect the waveform through distinct mechanisms, resulting in a complex echo waveform. For example, the longitudinal features of a target determine both the echo intensity and time delay, while its transverse features affect only the intensity²². Under these conditions, it is difficult to directly invert multi-feature information of targets within the laser footprint using echo waveform information through empirical, numerical, or simple repeated observations. The complex mapping relationship from the features of the target to the echo wave-

form poses significant challenges for the precise retrieval of multi-dimensional target information.

In recent years, substantial research has been devoted to processing echo waveform data from full-waveform LiDAR and retrieving ecological parameters from the derived waveform features. However, a systematic review and in-depth discussion of these methodologies remain limited. In 2019, Zhang et al. summarized advances in full-waveform LiDAR data processing techniques, but that work is now over six years old and primarily focused on challenges facing future Chinese spaceborne systems like the GaoFen-7 payload²³. In 2024, Huang et al. reviewed current spaceborne LiDAR observation systems, emphasizing the retrieval of forest structural parameters, product standards, accuracy assessment in forestry applications, and lidar parameter design²⁴. This review focuses on recent progress in echo waveform processing and target feature retrieval mechanisms within the full-waveform LiDAR framework. It does not cover hardware system innovations, for example, pre-launch calibration, geometric processing, laser parameters (e.g., power, beam quality, repetition frequency), footprint size optimization, sampling strategies (such as cross-track wide coverage and along-track discontinuous observation), day/night detection modes, beam configuration/sensitivity, etc^{6,25–34}. Instead, it concentrates on the underlying logic of waveform processing and multi-feature retrieval, encompassing their origins and latest applications. As shown in Fig. 1, the waveform processing includes waveform simulation, waveform decomposition and deconvolution, and analytical formula, whereas extraction methods of target features covers waveform simulation, analytical formula, waveform decomposition and deconvolution, and waveform metrics. Finally, it outlines strengths and limitations, as well as future directions for target extraction using full-waveform LiDAR in earth observation and its potential applications in space surveillance.

2 Methods of laser echo waveform simulation and processing

2.1 Waveform simulation

To extract target features, it is essential to comprehensively investigate the properties of laser echo waveform. Acquiring experimental echo waveform data across diverse scenarios remains challenging, therefore, it is necessary to advance the development of echo waveform simulation methods.

Prior to the launch of ICESat-1 (GLAS), to design waveform processing algorithms and evaluate their effectiveness, NASA developed a series of airborne full-waveform LiDAR systems, such as the Laser Vegetation Imaging Sensor (LVIS) and the Scanning Lidar Imager of Canopies by Echo Recovery, along with waveform simulation methods^{39,40}. In the field of waveform simulation, Blair and Hofton proposed a semi-empirical model³⁹. This model assumes uniform

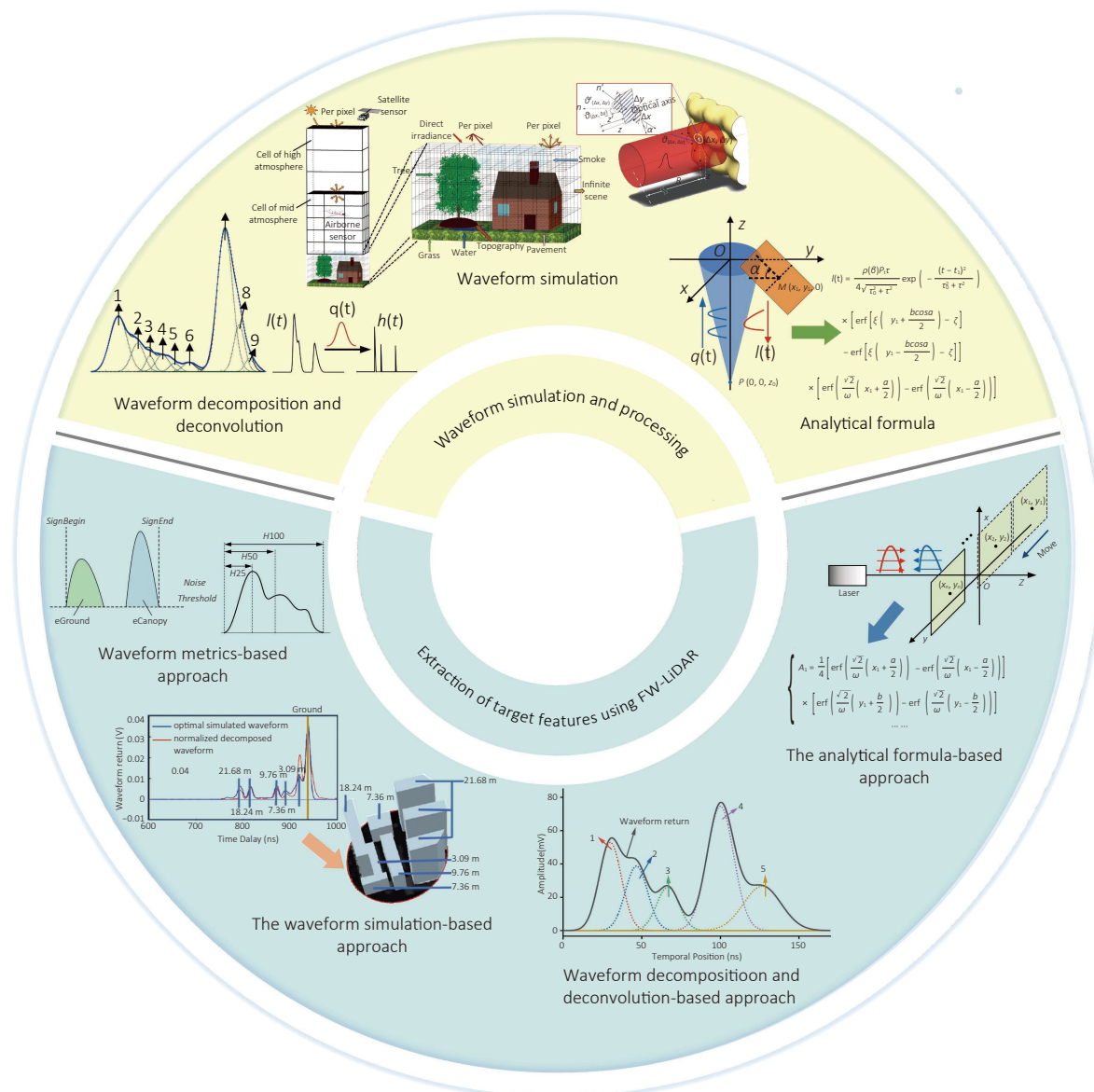


Fig. 1 | Sketch of waveform simulation and processing, and extraction of target feature using full-waveform LiDAR (FW-LiDAR). The former contains waveform simulation^{19,35}, waveform decomposition and deconvolution³⁶, and analytical formula²¹. The later includes waveform simulation³⁷, analytical formula³⁸, waveform decomposition and deconvolution, and waveform metrics. Figure reproduced with permission from: ref.¹⁹, Elsevier Ltd; ref.²¹, Taylor & Francis; ref.³⁵, under a Creative Commons Attribution License; ref.³⁶, IEEE; ref.³⁷, IEEE; ref.³⁸, under a Creative Commons Attribution License.

reflectance across all surfaces and represents the impulse response function as a superposition of normally distributed intensities in both the vertical (characterized by pulse width) and horizontal (characterized by footprint size) dimensions. A "pseudowaveform" is then generated by convolving this impulse response with the transmitted pulse. By incorporating instrument noise into this framework, Hancock et al. developed the GEDI simulator, which was used to calibrate echo waveform analysis algorithms for the GEDI mission⁴⁰. The performance of GEDI simulator was tested at several sites with different types of forests, canopy cover, and ground slopes. The simulated echo waveform, waveform

metrics such as relative height (RH) (for example, RH25, RH50, and RH80), and ground-finding errors estimated using Gaussian, inflection, and maximum methods, all agree excellently with those of LVIS. GEDI simulator has become the common simulation method of full-waveform LiDAR^{41–44}. Sun and Ranson et al. introduced a 3D LiDAR echo simulation model that treats the scene as a turbid medium voxel grid, with simplified assumptions including homogeneous surface types, uniform canopy structures, evenly distributed leaves, and only single scattering events⁴⁵. Since these methods rely on data collected from terrestrial or airborne small-footprint LiDAR systems to represent the

scene, they are referred to as data-driven semi-empirical models⁴⁶.

The assumption that ground object emissivity and single scattering are identical is a simplification of the properties of ground object scenarios and the transmission patterns of pulsed lasers. To accurately describe ground object scenarios and laser transmission patterns, researchers have developed various analytical radiative transfer models based on the concept of geometrical optical or Monte Carlo Ray Tracing^{35,47–54}. This model divides the scene into a series of grids, employs the radiative transfer equation to statistically track the radiative transfer process of photons within each grid, and obtains the modulated echo waveform of the ground object scenario. The model eliminates assumptions in scene modeling and precisely describes the complex interaction mechanisms between ground objects and lasers, including multiple scattering, thereby achieving accurate simulation of echo waveforms. Ni-Meister et al. constructed a hybrid model combining geometric optics and radiative transfer theory to characterize the influence of discrete canopy structures on environmental radiation⁴⁷. The approach models the surface as a slope with constant reflectance and represents vegetation canopies as single or double-layer ellipsoidal structures to simulate echo waveforms. In recent years, developed methods for simulating pulsed laser echo waveforms include Discrete Anisotropic Radiative Transfer (DART, including DART-RC and DART-Lux), Analytical Clumped Two-Stream (ACTS), Digital Imaging and Remote Sensing Image Generation (DIRSIG), time-dependent radiosity-based radiative transfer model, time-dependent stochastic radiative transfer theory, Spatially explicit 3-D lidar waveform model, FLIGHT, RAYTRAN, FLIES, LITE, LIBRAT, DELiS, and LESS^{21,35,48–54}.

Radiative transfer model accounts for the propagation of pulsed laser radiation through the atmosphere and multiple scattering within the target, making them suitable for objects such as vegetation where both scattering and refraction occur. For hard targets like ships or aircraft, where single reflection dominates and refraction is negligible, a common approach is to discretize the target into a set of micro-facets^{18,19}. Each facet is assigned properties such as position, tilt angle (angle between the normal direction of facet and the propagation direction of incident pulse), and reflectivity based on the target's 3D structure, attitude, and surface optical characteristics. Each facet produced an echo waveform that retains a Gaussian shape, with the peak intensity calculated using the LiDAR equation and taking into account the spatial energy distribution of incident pulse, and the time delay determined by the distance between the lidar system and the facet. The overall echo waveform modulated by the target is then obtained by summing the contributions from all illuminated micro-facets. Using this method, Li et al. studied the influence of side length and incident pulse width on the modulated echo waveform from a square, as well as the effects of base radius, tilt angle between the cone axis and the laser propagation direction, and incident pulse width on the echo from a conical target, discussing the

underlying physical mechanisms of modulation¹⁸. Hao et al. demonstrated how atmospheric turbulence and the tilt angle between the laser direction and the axis of four shapes, namely, a plane, a cone, a sphere, and an aspherical surface, affect the modulated echo waveform¹⁹. They quantified the relationship between beam radius, propagation distance, and turbulence intensity, as well as the dependence of echo pulse broadening on tilt angle using two methodologies: an analytical approach and a phase screen simulation.

The waveform simulation model enables the generation of LiDAR waveforms for any landscape and LiDAR sensor configuration. It can be applied to pre-launch calibration, on-orbit validation, data calibration, waveform data performance evaluation, target feature estimation model development, and land surface data product generation for both airborne and spaceborne full-waveform LiDAR systems^{35,39–57}. Based on these simulated models, comprehensive studies have been conducted to investigate how modulated echo waveforms depend on various parameters. These include structural, biophysical, and biochemical parameters of targets, such as vegetation type, height, crown dimensions, fractional cover, LAI, foliage chlorophyll content, surface topography, urban features, and water bodies. Other influencing factors encompass incident laser characteristics, detection angles, and additional system parameters. Recently, Tan et al. explored the capability of full-waveform LiDAR in capturing tree height composition, demonstrating that waveforms can reasonably reflect structural changes in even-aged forest canopies⁴¹. Liu et al. developed and validated a forest canopy structural complexity (CSC) estimation model, contributing to the global mapping of canopy structural complexity⁴². Cardenas-Martinez revealed the impact of geolocation correction on the inversion of vegetation structural parameters and elucidated the mechanism behind geolocation influences in sparse mountain forests⁴⁴. Leite et al. computed vegetation structural metrics using simulated waveforms and established relationships with field-measured fuel load components through a random forest (SF) model, supporting fuel load distribution mapping in the Cerrado region using on-orbit GEDI data⁵⁵. Zhao et al. simulated hyperspectral full-waveform data for various combinations of damaged areas and stress levels, assessing the potential of hyperspectral full-waveform LiDAR in monitoring forest pests and diseases⁵⁶. Mkaouar developed a novel method for assessing and correcting geolocation errors in full-waveform satellite laser altimetry data, thereby improving geolocation accuracy⁵⁷.

The key metrics for waveform simulation methods are simulation and accuracy speed. In terms of simulation speed, Li et al. developed a canopy radiative transfer model based on small-footprint LiDAR point cloud data, which physically describes the laser propagation process through the canopy⁴⁶. Compared to the GEDI simulator, their approach increased the average R^2 by approximately 2% to 5% and reduced the average RMSE by about 0.5% to 3%. Wang et al. compared the performance of DART-RC,

DART-Lux, and GEDI simulator across 48 turbid medium scenes, 32 RAMI (Radiation Transfer Model Intercomparison) scenes, and a forest site in the Manzano Mountains, New Mexico, USA, as shown in Table S1⁴³. The simulation accuracy of the first order-scattering models DART-RC and DART-Lux was comparable to that of the GEDI simulator, while the simulation accuracy of the multiple-scattering models DART-RC and DART-Lux outperformed GEDI simulator in sparse canopies.

The DART model produces more accurate simulated echo waveforms because it not only interprets the 3D scene structure but also tracks the transport and multiple scattering processes of over one million photons within the scene. However, this high fidelity comes at the cost of significantly increased computational time: generating a single echo waveform in DART-RC requires about 22 minutes, whereas the GEDI simulator generally completes a simulation in just 2 seconds⁴³. To reduce simulation time, Winiwarter et al. simplified the interaction model between beam and scene to a computationally feasible yet physically realistic level⁵². They developed HELIOS++, which strikes a balance between physical accuracy and computational complexity. Although its 3D scene preprocessing stage takes approximately 10 times longer than DART's, its beam simulation phase is 314 000 times faster. Luo et al. developed a multi-platform LiDAR simulation model based on the large-scale remote sensing and image simulation framework, LESS⁵³. In abstract scenes, the waveform relative error of their model was less than 1% compared to the DART LiDAR module. In real-world scenarios, R^2 ranged from 0.9108 to 0.9984, with an average of 0.9689. Moreover, their approach achieved a simulation speed more than three times faster than DART, with the single-ray mode further reducing simulation time by over 28 times.

2.2 Analytical formula

As mentioned above, both the facet computation method and the radiative transfer equation require the construction of a target scene and the calculation of either the laser-facet interaction or photon transport within a grid to simulate echo waveforms. Whenever target parameters, such as size, orientation, or position, that influence the echo waveform are altered, the model must be reconstructed. On one hand, as shown in Fig. S2, even for a simple square target, it must be discretized into 20,000 facets and calculate the individual echo waveform to form the correct overall echo waveform. This entire process takes over one second. For meter-scale, complex 3D targets, which require discretization into hundreds of thousands or even more facets along with occlusion judgment, the computation time can extend to several tens of minutes. On the other hand, neither the facet-based approach nor the radiative transfer equation establishes a direct relationship between target characteristics and waveform features, which hampers the inversion of target

characteristic parameters. Therefore, it is imperative to first characterize the target features and subsequently develop an analytical expression for the echo waveform that incorporates these properties.

The echo waveform analytic method is based upon the ideal that the modulated echo waveform is the convolution between the transmitted waveform and the target's impulse response function¹⁶. The expression of the target's impulse response function is

$$h(t) = \rho(\alpha) g(x, y) \delta(t - 2z(x, y)/c), \quad (1)$$

where ρ is the surface optical properties, α is the tilt angle between the normal direction of planar target and the propagation direction of incident pulse, $g(x, y)$ describes the spatial distribution of incident pulse, $z(x, y)$ refers to the distance between the target and LiDAR system, and $\delta(t)$ is the Dirac function, which distinguishes the echo waveforms modulated by target surfaces located at different distances relative to the LiDAR system. In general, the temporal and spatial distributions of incident pulse both obey Gaussian function

$$\begin{cases} p(t) = \exp\left(-\frac{4t^2}{\log 2 \times \tau^2}\right) \\ g(x, y) = \frac{2P_t}{\pi\omega^2} \exp\left(-2\frac{x^2 + y^2}{\omega^2}\right) \end{cases} \quad (2)$$

To clearly reveal the modulation principles of target characteristics on the echo waveform, parameters unrelated to the target's modulation mechanism, such as antenna gain and atmospheric transmittance, are neglected, leading to a general analytical expression for the target-modulated echo waveform

$$I(t) \propto \frac{2P_t \rho(\alpha)}{\pi\omega^2} \iint \exp\left(-2\frac{x^2 + y^2}{\omega^2}\right) \exp\left[-\frac{4(t - 2z(x, y)/c)^2}{\log 2 \times \tau^2}\right] dx dy. \quad (3)$$

In Eq.(5), target characteristics such as 3D structure, size, position, and orientation influence the modulated echo waveform by affecting $z(x, y)$. The 3D structure governs the expression form of $z(x, y)$, while the size, position, and orientation determine its relevant parameters. As a result, the analytical expressions of the modulated echo waveforms differ across targets of various shapes.

Steinvall computed the laser radar cross-section and impulse response function for four common basic targets: cones, spheres, paraboloids, and cylinders¹⁶. He evaluated the relationship between these two parameters and the angular dependence of the bidirectional reflectance distribution function across different target shapes, and further investigated how the laser radar interface varies with reflective parameters and the ratio of target-to-beam size. The results indicate that understanding the coupling effect between target geometry and reflective properties is essential for predicting both the waveform characteristics and amplitude intensity of LiDAR echoes. If one parameter is known, the

other can be inversely derived¹⁶. However, without accurate angular reflectance data, it is challenging to perform deconvolution for unknown target geometries. Johnson derived an analytical expression for the modulated echo waveform from an infinitely large (i.e., extended) planar target¹⁷. Through simulations, he demonstrated how the echo waveform varies with the tilt angle (the angle between the surface normal and the incident pulse). The results show that target tilt causes broadening of the echo pulse. When the target is several kilometers away and tilted more than 45 degrees, the increased pulse width significantly degrades ranging accuracy.

Compared to infinite planar targets, finite (non-extended) targets require consideration of attributes such as shape and position, making the modulation mechanism more complex. Xu et al. derived analytical expressions for the modulated waveforms from planar targets, cones, and cylinders²⁰. Using simulation data, Xu et al. conducted a detailed investigation into the effects of various parameters on echo waveforms, including, tilt angle for both Lambertian and non-Lambertian extended planar targets, side length of Lambertian and non-Lambertian rectangular targets, base radius of non-extended cones and half-cone angle of extended cones, and base radius and distance for extended Lambertian cylinders, etc. Subsequently, Xu experimentally validated the simulation results and proposed a hybrid ranging method based on a back-propagation neural network to minimize ranging errors induced by target surface properties.

Xu et al. focused on circumferential detection scenarios using pulsed laser, which exhibits the several characteristics. The first is the short detection range, which ranges from several to tens of meters, and the laser footprint is sub-meter (with a divergence on the order of 10 mrad). The second is target geometry, which is basic shapes such as rectangles, extended planes, cones, and cylinders. The third is target position, the center of rectangles, the vertex of cones, and the axis of cylinders are all assumed to be aligned with the center of the laser spot.

In contrast, for remote sensing spanning tens to hundreds of kilometers, the laser footprint expands to several or tens of meters, often fully illuminating the target. In such cases, the influence of target characteristics, such as shape, size, position, and orientation, on the echo waveform must be thoroughly accounted for. Hu et al. derived analytical expressions for echo waveforms modulated by basic planar shapes including, rectangles, circles, and equilateral triangles, as shown in Eqs.(4), (5), and (6)²¹

$$l(t) = \frac{\tau\rho(\beta)}{4\sqrt{\tau^2 + \tau_0^2}} \exp\left[-\frac{(t-t_1)^2}{\tau^2 + \tau_0^2}\right] \left\{ \operatorname{erf}\left[\frac{\sqrt{2}}{\omega}\left(x_1 + \frac{a}{2}\right)\right] - \operatorname{erf}\left[\frac{\sqrt{2}}{\omega}\left(x_1 - \frac{a}{2}\right)\right] \right\} \left\{ \operatorname{erf}\left[\xi\left(y_1 + \frac{b\cos\alpha}{2}\right) + \zeta\right] - \operatorname{erf}\left[\xi\left(y_1 - \frac{b\cos\alpha}{2}\right) - \zeta\right] \right\}, \quad (4)$$

$$l(t) = \frac{\rho(\beta)}{8\omega} \sqrt{\frac{\pi}{2}} \exp\left[-\frac{(t-t_1)^2}{\tau^2 + \tau_0^2}\right] \int_{y_1-R\cos\alpha}^{y_1+R\cos\alpha} \left\{ \operatorname{erf}\left[\frac{\sqrt{2}}{\omega}\right. \right. \\ \left. \left. (x_1 + p)\right] - \operatorname{erf}\left[\frac{\sqrt{2}}{\omega}(x_1 - p)\right] \right\} \exp[-(\xi y - \zeta)^2] dy, \quad (5)$$

$$l(t) = \frac{\rho(\beta)}{\pi\omega} \sqrt{\frac{\pi}{2}} \exp\left[-\frac{(t-t_1)^2}{\tau^2 + \tau_0^2}\right] \int_{y_1-\sqrt{3}d\cos\alpha/3}^{y_1+\sqrt{3}d\cos\alpha/6} \\ \exp[-(\xi y - \zeta)^2] \left\{ \operatorname{erf}\left[\frac{\sqrt{2}}{\omega}(x_1 + q)\right] - \operatorname{erf}\left[\frac{\sqrt{2}}{\omega}(x_1 - q)\right] \right\} dy, \quad (6)$$

where

$$\begin{cases} t_1 = \frac{2(z_0 - y_1 \tan\alpha)}{c} & \tau_0 = \frac{\sqrt{2}\omega \tan\alpha}{c} \\ \xi = \sqrt{\frac{1}{\tau^2} + \frac{1}{\tau_0^2}} \frac{2\tan\alpha}{c} & \zeta = \frac{t-t_1}{\sqrt{\tau^2 + \tau_0^2}} \frac{\tau_0}{\tau} \\ p = \sqrt{R^2 - (1 + \tan^2\alpha)(y-y_1)^2} & q = \frac{\sqrt{3}}{3} \left(\frac{y-y_1}{\cos\alpha} + \frac{\sqrt{3}}{3}d \right) \end{cases}, \quad (7)$$

where (x_1, y_1, z_1) is the center position of target, and (b, a) , R , and d are the size of rectangle, circle, and equivalent triangle, respectively.

Eqs.(4), (5), and (6) establishes a relationship between the properties of the modulated echo waveform and the characteristics, including shape, size, position, and tilt angle, of four typical types of two-dimensional (2D) planar targets. Based on simulations using these equations, Hu et al. systematically investigated the influence of target properties on the echo waveform^{21,22}. Under vertical illumination (i.e., non-sloped conditions), the echo retains a Gaussian form with unchanged pulse width, and time delay is determined by the distance between the laser receiver and the target. The peak intensity is sensitive to target attributes such as shape, size, center position, and reflectivity. Hu et al. demonstrated the feasibility of distinguishing target shapes by analyzing the specific dependence of peak intensity on the target's center position. The results indicate that when the target size exceeds 0.407 times the laser footprint size, target shapes can be effectively discriminated²². Conversely, when the target is smaller than this threshold, it approximates a point target, and the echo peak intensity loses sensitivity to variations in shape.

As the tilt angle increases, the echo waveform progressively deviates from a Gaussian shape, characterized by a decrease in peak intensity and a rise in pulse width²¹. At a tilt angle of 80° with the target center aligned to the beam center, the modulated echo peak position remains unchanged for rectangular and circular targets. With increasing target size, the waveforms undergo a sequential transition: starting as Gaussian, then flattening into a non-Gaussian form, and eventually restoring a Gaussian-like bulge. In contrast, for

equilateral triangular targets, the peak position shifts, and the waveform transitions directly from Gaussian to non-Gaussian without recovery. When the target center is offset from the beam center, the modulated echoes from rectangular and circular targets not only deviate from the Gaussian form but also lose left-right symmetry. The direction of skewness (right or left) depends on both the target's position and its tilt orientation.

Recently, Hu et al. derived analytical expressions for the modulated echo waveforms from four typical 3D targets, namely, a cuboid, hexagonal prism, cone, and ellipsoid⁵⁸. They conducted a detailed investigation into how target shape, size, orientation, and position influence the echo waveform. From the perspective of target geometry and the spatial energy distribution of the incident pulse, they revealed the underlying modulation mechanisms. This study provides a foundation for analyzing echo modulation effects from satellite bodies, European-style building roofs, and forest canopy structures.

Wu et al. also reported that as target size increases, the modulated echo waveform transitions from Gaussian to a non-Gaussian form with a flattened top¹⁸. Conversely, with increasing pulse width, which is equivalent to a reduction in effective size, the waveform reverts from a flattened non-Gaussian shape back to Gaussian. However, Wu et al. focused solely on rectangular targets and did not account for the influence of target position. In contrast, simulation results from Xu et al. indicated that for a rectangular target centered in the laser footprint, the modulated echo retains a Gaussian shape even with increasing tilt angle and target size, with only the peak intensity varying²⁰. This discrepancy in waveform behavior stems from differences in simulation parameters: in Xu et al.'s setup, the beam radius was 0.1 m, and at a 70° tilt, the maximum pulse broadening was 1.3 ns. The increase in echo pulse is significantly less than the 6 ns transmitted pulse width, and results in negligible echo waveform broadening. In Hu et al.'s study, however, the beam radius was 0.5 m, and at the same tilt angle, the pulse broadening reached 6.5 ns, far exceeding the 1 ns transmitted pulse width and leading to pronounced waveform broadening.

2.3 Waveform decomposition and deconvolution

2.3.1 Gaussian decomposition

The above two methods focus on generating the echo waveform and does not involve waveform processing. The full-waveform LiDAR technology evolved from conventional laser ranging captures the complete echo signal, which consists of multiple sub-waveforms modulated by different components of complex composite targets or by distinct individual targets within the laser footprint. The waveform decomposition method aims to separate the composite echo into a series of such sub-waveforms, thereby isolating the

contributions from each component or target. By analyzing properties, including the number, width, peak position, peak amplitude, and skewness of these sub-waveforms, it is possible to infer quantitative information, such as number, 3D structure, and biophysical or biochemical properties, about the targets within footprint and ultimately enabling the reconstruction of the underlying scene.

Figure S3 reveals the typical flowchart of waveform decomposition. The key processes include mathematical model selection, parameter estimation and optimization, etc. Blair and Hofton pioneered the Gaussian decomposition approach, which assumes that the echo signal modulated by each individual target within the laser footprint follows a Gaussian form⁵⁹

$$I(t) = n(t) + \sum_{m=1}^N A_m \exp \left[-\frac{(t - t_m)^2}{2\sigma_m^2} \right], \quad (8)$$

where n is the noise, N refers to the number of sub-waveforms, and A_m , t_m , and σ_m denote the peak intensity, peak position, and pulse width, respectively of decomposed component. The number of Gaussian components is determined based on the number of inflection points. Initial estimates of the half-width and peak positions are derived from consecutive inflection points, while the initial amplitude estimates are obtained using a nonnegative least square method. Finally, all initial parameter estimates are optimized via the Levenburg-Marquardt (LM) method. Using the proposed algorithm, Blair and Hofton successfully decomposed a simulated waveform into twelve Gaussian components and applied the LM method to refine three of them. After optimization, the errors in the peak position, peak intensity, and pulse width of these three components were all within 2.6%.

Owing to the relatively straightforward mathematical process and the well-definite physical meaning of parameters such as pulse width and peak position in the decomposed component, Gaussian decomposition has become the most mature and widely used method for echo waveform decomposition^{60,61}. NASA developed a waveform decomposition technique capable of resolving up to six Gaussian components for processing GLAS data, and Duncanson et al advanced the methodology to accommodate an unlimited number of Gaussian components⁶².

2.3.2 Wiener deconvolution

The modulated echo waveform from a target is the convolution of the target's impulse response waveform and the transmitted pulse. Even after decomposition, the resulting waveform components still contain contributions from the transceiver system. To remove the system effects and accurately retrieve 3D structural and biophysical and biochemical information of the targets, the waveform deconvolution methods has been proposed to recover the target's impulse response waveform

$$l(t) = n(t) + p(t) \otimes h(t), \quad (9)$$

where $p(t)$ and $h(t)$ are the temporal distribution of the incident pulse and the impulse response of target, respectively.

Jutzi and Stilla pioneered the idea of waveform deconvolution⁶³. In their experiment, a pulsed laser was used to illuminate both a single planar surface and two parallel surfaces separated by 0.15 m, both at a distance of 100 m. The echo waveforms were processed using the Wiener filtering algorithm for deconvolution, and the impulse response of each surface, which is theoretically a Dirac function, was approximated via Gaussian fitting. The results showed a standard deviation of 6.2 mm for the single plane, and a separation of 0.149 m and a separation standard deviation of 5.9 mm for two planes. Although the transmitted pulse had a width of 5 ns and the echo time difference between the two surfaces was only 1 ns, the Wiener filter-based deconvolution successfully resolved the superimposed waveforms. This outcome significantly improved the ranging accuracy and demonstrated the potential of this approach in discriminating overlapping responses from multiple surfaces.

2.3.3 Target response waveform (TRW)

Qiu et al. established a mathematical expression for the TRW of urban in definite integral form based on the lidar equation and the working principle of laser altimeters⁶⁴. Using a rectangular plane as an example, they simulated and analyzed the influences of target properties, including region area, center geolocation, and surface slope, on the TRW morphology and characteristic parameters such as signal energy, time delay, pulse width, and skewness. The simulation results indicate that, except when the target center coincides with the laser footprint center, the TRW generally exhibits left-skewed or right-skewed distributions. The signal energy increases with larger region areas and decreases with greater distance between the target center and the laser footprint center, while being attenuated by the cosine of the surface slope. The root-mean-square pulse width is primarily influenced by the tangent of the surface slope. Therefore, the influence of different target properties on the characteristic parameters of TRW are fundamentally distinct. Consequently, the TRW should not be simplistically treated as Gaussian, and its skewness parameters should not be ignored.

The TRW study by Qiu et al. provides reliable prior knowledge for range correction, significantly improving ranging accuracy and serving as an important theoretical foundation for understanding earth's topography. However, their research was limited to rectangular shapes and failed to establish analytical expressions for TRW extending from typical 2D targets to complex 3D composite structures. Moreover, it did not systematically examine the influence mechanisms of target characteristics such as 3D geometry, dimensions, orientation, position, and surface optical properties on TRW.

The conclusions of Qiu et al. regarding the influence of target characteristics on TRW are closely aligned with those of Hu et al. on how target characteristics affect echo waveforms (as opposed to TRW). This similarity is notable even though Hu et al. incorporated a variety of typical 2D shapes in their analysis. The reason for this alignment lies in the clever configuration of parameters in Hu et al.'s study: the target size was set to more than 1.5 times the ranging accuracy corresponding to the pulse width, and the footprint radius was approximately 3.3 times that value. Under such conditions, the influence of system parameters is minimized, making the echo waveform highly sensitive to target shape, position, and surface slope, potentially deviating from a Gaussian form. If the pulse width or beam size is large (i.e., the target size is much smaller than the footprint and ranging resolution corresponding to the pulse width), the target approximates a point target. In this case, the echo waveform remains Gaussian, and its shape becomes independent of target parameters such as shape, size, location, and tilt angle.

2.3.4 Improvement and comparison of waveform decomposition and deconvolution algorithms

The Gaussian decomposition method is sensitive to the initial values of sub-waveforms. Deviations in initial values would lead to convergence to local minima or even failure to converge. To address this issue, Wu et al. proposed a method for estimating initial parameters. By optimizing the fitting parameters, they achieved a decomposition distance error of less than 0.015 meters for simulated waveforms and less than 0.1 meters for echo signals in field tests⁶⁵. Further advancing the methodology, Xie et al. introduced a genetic algorithm-based decomposition approach. This method utilizes probabilistic transition rules instead of deterministic ones, eliminating the need for accurate initial parameters and exhibiting global optimization capabilities. The value of R^2 improved from 0.931 for a single waveform to 0.997 for three waveforms⁶⁶. Xu et al. adopted a modified differential evolution algorithm to replace the traditional LM algorithm. This modification substantially enhanced the accuracy of waveform decomposition, outperforming not only the LM method but also conventional differential evolution, genetic algorithms, and particle swarm optimization⁶⁷.

Echo signals tend to weaken and become susceptible to being masked by noise or other overlapping echoes when the target is located at the edge of the footprint, has a small size, or exhibits low reflectivity. Moreover, when two targets are separated by a distance smaller than the range resolution corresponding to the transmitted pulse width, the modulated echoes would be overlap, forming a pseudo-Gaussian return. Conventional Gaussian decomposition often struggles to detect and separate such hidden or superimposed waveforms effectively. Additionally, the Gaussian decomposition method is valid only under specific conditions: the incident pulse itself must be Gaussian, and interaction with

a single target should alter only the peak position and width of the echo, not its fundamental waveform. In practice, however, transmitted pulses are not always Gaussian, they may exhibit right-skewed, heavy-tail, or negative-tail shapes. Furthermore, previous simulation and experimental results have shown that echoes modulated by typical 2D and 3D targets are strongly influenced by properties such as target shape, size, and orientation. As a result, the echo waveform often deviates significantly from a Gaussian form and cannot be accurately decomposed using conventional Gaussian-based methods.

To address I) highly overlapping pseudo-Gaussian waveforms caused by closely spaced targets, II) weak echoes prone to being masked by noise, III) transmitted pulses of diverse shapes, and IV) echo waveforms that deviate from the transmitted pulse shape after target modulation, a variety of waveform decomposition methods beyond the conventional Gaussian model have been developed. These include the generalized Gaussian, adaptive Gaussian, generalized normal Gaussian, log-normal Gaussian, generalized log-normal Gaussian, linearly approximated iterative Gaussian decomposition, Weibull, B-spline, log-normal distribution, Burr, skew-normal distribution, piecewise exponential functions, Nakagami, and multiscale wavelet analysis, among others. Similarly, to improve processing performance and leverage the benefits of deconvolution, numerous waveform deconvolution techniques have also been proposed. These encompass the Richardson-Lucy (RL) algorithm, Gold's algorithm, B-spline deconvolution, skew-normal modeling, the Expectation-Maximization algorithm, blind deconvolution, sparsity-based regularization, the Tikhonov regularization filter, and Bayesian deconvolution, among others. Table S2 summarize the performance evaluation and comparison of these waveform decomposition and deconvolution methods in detail. Here we focus on some representative studies.

Zhou et al. proposed the Gaussian inflection point selection (GIPS) method⁶⁸. This approach uses the number of inflection points near the peak of the echo signal to determine the component's half-width location. An appropriate half-width value is then selected to iteratively decompose the echo signal. A global LM algorithm is employed to optimize the Gaussian model for fitting the echo waveform. In experiments using LVIS echo waveforms from the southwestern region of Issaquena, Mississippi, the GIPS method achieved an average coefficient of determination of 0.9799, significantly outperforming Gaussian inflection point matching (0.6860) and iterative decomposition (0.9385). These results demonstrate that GIPS can more effectively identify hidden echo signals compared to other methods. Zhang et al. introduced a decomposition method combining hidden peak detection and an adaptive genetic algorithm (HAGA)⁶⁹. This method uses a hidden peak detection algorithm to accurately extract Gaussian components and provide robust initial parameters. An adaptive genetic algorithm integrated

with the LM algorithm is then applied to refine the parameters and achieve high-precision waveform inversion. Experimental validation showed that the HAGA method outperformed conventional techniques in both waveform decomposition accuracy and tree height extraction. Specifically, it attained an average R^2 of 0.955 across more than 2 000 laser footprints, compared to 0.916 for the optimized decomposition method and 0.862 for a standard genetic algorithm.

Zhang et al. proposed an integrated approach combining waveform deconvolution and decomposition: the response waveform of the ground object scene is first obtained through deconvolution, and then waveform features are extracted via Gaussian fitting⁷⁰. By integrating Wiener filtering, Tikhonov regularization filtering, RL, and blind deconvolution algorithms with Gaussian decomposition, four hybrid processing methods, namely Wnr-Gauss, Reg-Gauss, Lucy-Gauss, and Blind-Gauss, were developed. In simulated urban areas, these four methods detected 8.1%, 6.5%, 8.2%, and 7.7% more peaks, respectively, than standard Gaussian decomposition. In simulated forest areas, the corresponding increases in the number of detected peaks were 20.8%, 18.1%, 15.8%, and 17.2%, respectively.

To systematically evaluate the performance of various waveform decomposition methods, comparative studies have been conducted from different perspectives. Zhang et al. reviewed the mathematical models, underlying assumptions, and advantages/disadvantages of multiple methods, including Gaussian, generalized Gaussian, log-normal distribution, B-spline, and wavelet decomposition, though without providing explicit comparative results of the analyses²³. Liu et al. advanced the field by proposing a B-spline progressive decomposition method suitable for low signal-to-noise ratio (SNR) conditions⁷¹. They further conducted a detailed comparison of five methods, namely, Gaussian decomposition, Gaussian RL deconvolution, B-spline RL deconvolution, Gaussian progressive decomposition, and B-spline progressive decomposition, across five metrics, target detection ability, robustness against noise, ranging trueness, ranging repeatability, and time efficiency. Their findings indicated that only the B-spline progressive decomposition method demonstrated superior performance across all five metrics. Furthermore, focusing on terrain-specific applicability, Zhou et al. provided a detailed comparison of five methods, Gaussian decomposition, adaptive Gaussian decomposition, Weibull decomposition, RL deconvolution, and Gold deconvolution, in processing echoes from four classic terrain types: forest, glacier, lake, and residential areas⁷². Their results revealed that Gaussian decomposition is fast with high component accuracy but struggles to resolve overlapping peaks. Adaptive Gaussian decomposition and Weibull deconvolution show greater advantage in detecting weak and overlapping signals, making them suitable for densely vegetated areas, albeit with issues of overfitting and higher false detection rates, respectively. RL deconvolution performs best in sparsely vegetated areas but converges

slowly. In contrast, Gold deconvolution exhibits the strongest comprehensive capability in handling complex waveforms involving multiple, weak, and overlapping signals, maintaining very low false detection rates across all four terrain types while also being computationally efficient.

2.3.5 Waveform decomposition of multi- and hyperspectral full-waveform LiDAR signal

Multispectral/hyperspectral LiDAR captures both 3D structural information and material properties of targets by emitting and receiving echo waveform signals across tens to hundreds of wavelengths⁷³. Currently, optical satellites such as the Landsat series and China's Gaofen series are generally equipped with multispectral cameras^{2,8,9}, with satellites like Gaofen-5 even adopting full-spectrum imaging instruments^{3,7}. These have been widely used in surface ecological monitoring, resource exploration, disaster prevention and mitigation, among other fields^{2,3,7–9}. Laboratory research of waveform decomposition techniques for multispectral/hyperspectral LiDAR have gradually overcome key challenges including weak/overlapping component missed detection, multi-channel data consistency, long-range signal attenuation, pseudo-single-pulse identification for extremely close targets, and component mismatch for natural targets^{74–79}. This progress marks a critical transition from multi-wavelength collaborative decomposition to large-scale hyperspectral applications, demonstrating considerable potential in effective decomposition of overlapping echoes and improvement of target ranging accuracy.

Utilizing the differential responses of different wavelengths to the same target for complementary detection, Song et al. proposed a multispectral waveform decomposition method⁷⁴. Based on the physical principle that echo center positions (reflecting target position) are wavelength-independent, while amplitude and full width at half maximum vary with wavelength (related to target reflectance, etc.), the innovation lies in the multi-wavelength component optimization strategy: if the peak distance of a waveform component across multiple wavelength channels exceeds one-sixth of the pulse width, it is considered indicative of weak or overlapping components, and a new component is added to avoid missed detection in single-wavelength analysis. To validate the method, simulated data and laboratory-measured data were used. Results showed that the relative neighbor distance error (RNDE) of multispectral waveform decomposition was only 0.0100–0.0610, significantly better than the 0.0566–0.2833 of single-wavelength methods. It could distinguish targets spaced 20 cm apart (single-wavelength waveform decomposition only resolved 30 cm spacing), with the R^2 stable between 0.89 and 0.92. However, the logic of energy-sorting matching relied upon by this method is prone to component mismatch in natural scenes due to reflectance fluctuations, and increasing spectral channels as well as uneven energy distribution

across channels over long distances exacerbate this issue.

As the number of LiDAR spectral channels increased from several to hundreds, and detection scenarios expanded from the laboratory to outdoor long-range settings, two upgraded methods emerged addressing hundred-channel data consistency and long-range signal attenuation, respectively. Xia et al. proposed a hundred-channel hyperspectral decomposition method targeting issues such as false detection, missed detection, and over-detection in large-scale data from a 101-channel (550–1050 nm, spectral resolution 5 nm, pulse width 4 ns) hyperspectral LiDAR⁷⁵. They designed a single-channel waveform decomposition (SCWD) and multi-channel mutual correction and decomposition (MCMCD) process: SCWD stage used a skewed normal distribution model to fit potentially asymmetric waveforms and initialized parameters via a hierarchical stripping strategy, and MCMCD stage statistically analyzed distance differences of adjacent targets across all channels. In a camouflage net occlusion experiment (distances: hyperspectral LiDAR-camouflage net-target were 3 m and 0.45 m, respectively), MCMCD improved average decomposition accuracy by 20.1%, reduced the RNDE between adjacent targets from 0.1253 to 0.0037, and decreased distance dispersion by 95.36%, achieving stable decomposition of hundred-channel-scale hyperspectral waveforms for the first time.

Addressing severe signal attenuation and uneven multi-channel energy distribution in long-range (>500 m) detection, Wang et al. proposed a multi-channel interconnected waveform decomposition (MIWD) method and extended⁷⁶. Its core innovations include multi-wavelength weighted accumulation and hyperspectral waveform parameter extraction: the former stage dynamically selected 20 optimal channels using a multi-echo quality metric and assigned weights based on noise levels to suppress interference from high-noise channels, and the later stage fixed the center positions of multi-channel components after time synchronization calibration. In a 518 m outdoor long-range experiment, MIWD improved ranging accuracy by more than twofold with the ability to distinguish the target spaced 43 cm (single-wavelength waveform decomposition 60 cm), and controlled the RNDE within 0.0042–0.1958. This extended hyperspectral LiDAR waveform decomposition from short-range laboratory scenarios to practical outdoor long-range applications for the first time.

Focusing on the pseudo-single-pulse challenge formed by extremely closely spaced targets, which is difficult to distinguish from single-target broadened pulses using traditional methods, Xia et al. proposed a range resolution enhancement method⁷⁷. The core breakthrough lies in judging pulse type based on spectral characteristics: the geometric center positions of pseudo-single-pulses exhibit clustered asymmetric distribution with wavelength, whereas those of single-target broadened pulses show random distribution. Building on this, initial component parameters were dynamically set according to the change ratio of broadening width

to skewness coefficient, followed by multi-channel mutual optimization to correct abnormal channels. Validations with simulated data and laboratory-measured data showed that this method improved the ranging resolution from 60 cm to 5 cm under a 4 ns pulse width, breaking the traditional pulse width limitation on resolution. The relative error of average distance was controlled within 0.003–0.064, with R^2 between 0.88 and 0.96, providing key technical support for extremely close target detection.

To further advance towards complex natural targets and extreme-resolution scenarios, hyperspectral waveform decomposition method has developed two representative methods. Bai et al. proposed the Rclonte and Rclonte-M methods^{78,79}. Addressing the failure of the energy-sorting matching logic of multispectral waveform decomposition for natural targets (owing to drastic reflectance fluctuations with wavelength), they adopted a center-position-sorting matching logic, leveraging the invariance of echo center position order along the laser path for multiple natural targets: the relative spatial order of the same set of targets remains fixed across different wavelengths. Both methods start by sorting the target center positions obtained from decomposition at each wavelength in descending order, inferring that center positions with the same order likely belong to the same real target, thereby compensating for missed or false detection parameters (center position, full width at half maximum, amplitude) in single-wavelength decomposition. This addresses the difficulty traditional single-wavelength decomposition faces with weak responses and overlapping echoes from natural targets. The method first employs a two-stage initialization, coarse estimation (peak and inflection point detection) + fine estimation (supplementing weak components in single-peak intervals), to reduce initial parameter deviation. Subsequently, in the re-optimization stage, it matches multi-wavelength components according to position order based on the rule that the center position order of the same target remains unchanged across wavelengths, effectively compensating for weak or overlapping components missed in single-wavelength analysis. Validation via simulated data and laboratory-measured data showed that Rclonte achieved an average R^2 of 0.9356 (multispectral waveform decomposition only 0.7196), a root mean square error (RMSE) of 0.2889 mV (close to system noise level), a RNDE of 0.026–0.085, and extracted spectral curves with similarity to measured reflectance between 0.8186 and 0.9917. This realized accurate decomposition of hyperspectral waveforms for natural targets, while also obtaining relatively accurate spatial and spectral information.

Addressing the trade-off between processing efficiency and accuracy for large-scale measured data, Shi et al. achieved efficient processing of kilometer-scale massive hyperspectral LiDAR waveform data⁸⁰. The core of their approach is a processing framework of hierarchical waveform feature screening - parallelized parameter optimiza-

tion - adaptive error correction: first, fast filtering of invalid waveforms based on features such as waveform amplitude and rising edge slope; then, parallel fitting of multi-component waveforms using a differential evolution algorithm; finally, adaptive error correction of decomposition results based on spectral consistency criteria. The method was validated on large-scale measured data covering urban building clusters and mountainous forests. Experimental results showed that under an 8 ns pulse width, the ranging resolution reached 0.3 m, processing efficiency for massive data improved 15-fold compared to traditional methods, and decomposition error was controlled within 3%. Moreover, the method is compatible with multi-platform hyperspectral LiDAR systems, effectively addressing the industry challenge of balancing high accuracy with high efficiency.

2.3.6 Artificial intelligence in waveform decomposition and deconvolution

The rapid advancement of artificial intelligence has propelled progress in waveform decomposition and deconvolution. Liu et al. developed an automatic peak count-driven classifier and extractor based on machine learning⁸¹. The results from processing GLAS waveforms showed that five classifiers, namely, support vector machine (SVM), RF, gcForest, kernel extreme learning machine, and deep learning model of convolutional neural network (CNN) all can accurately eliminate invalid waveforms and peaks, distinguish waveforms with different numbers of peaks, and provide significant assistance in selecting suitable waveforms for characteristic studies. The performance of all five classifiers closely matched the official GLAH14 records, as evidenced by the overall accuracies of 90.65%, 92.69%, 92.45%, 91.95%, and 93.05%, respectively.

To address the effect of noise, Liu et al. developed Gaussian decomposition approaches enhanced by two deep learning architectures: a Dense Connection Networks (FDCN) for waveforms with SNR above 24 dB, and a Deep Residual Network (FDRN) tailored for low SNR conditions⁸². These methods achieve substantially smaller errors on both echo peak position and amplitude estimation compared to conventional empirical mode decomposition. For example, the successful ratios of FDCN, RDRN and EMD-soft are 0.99, 0.99 and 0.63 for 30 dB, 0.93, 0.89, and 0.24 for 24 dB, 0.40, 0.72, and 0.04 for 18 dB, and 0.26, 0.68, and 0 for 12 dB, respectively. Furthermore, they proposed an attention-based full-waveform decomposition network that operates without predefined parameters⁸³. This end-to-end framework demonstrates strong performance, with decomposed echo components attaining an average R^2 of 0.9799, and a classification accuracy of 96% using the first module.

For improving the distance resolution, Xu et al. proposed a waveform decomposition method based on the sparrow search Algorithm and long short-term memory neural network⁸⁴. When the decomposition success rate for ground

distance exceeded 70%, the minimum decomposition distance of this method was 0.1 m smaller than that of the inflection point (IP) and RL deconvolution methods, and 0.2 m smaller than that of the multi-Gaussian decomposition method. When the peak distance decomposition success rate exceeded 70%, the minimum decomposition distance of this method was 0.25 m smaller than that of the IP method, and 0.1 m smaller than that of the RL and MGD methods.

To propose a novel waveform signal detection model, Ji et al. integrated an improved extreme learning machine algorithm, leapfrog sliding window full-waveform segmentation, and a self-annealing particle swarm optimization algorithm⁸⁵. This model enables waveform classification and segmentation without complex feature extraction or iterative processes. Airborne experimental results demonstrated that when combined with the RL deconvolution algorithm, this method achieves higher detection efficiency and accuracy compared to generalized Gaussian decomposition and the standalone RL deconvolution algorithm.

3 Extraction of target features using full waveform LiDAR

3.1 The waveform simulation-based approach for target feature inversion

The method for acquiring target features via waveform simulation involves matching experimentally measured echoes with simulated echoes. When the correlation coefficient between the two exceeds a predefined threshold, the target features from the simulated scene are assigned to the measured scenario. Waveform simulation methodologies primarily include lookup table approaches, optimization algorithms, and iterative numerical optimization techniques^{37,51,86–88}.

In the lookup table method, Koetz and Sun et al. constructed an echo waveform database comprising 100 000 combinations of five ecological parameters: maximum tree height, crown length, crown width, LAI, and fractional cover (fcover)⁸⁶. By comparing LVIS echo waveforms from the Swiss National Park with this simulated waveform database, they obtained R^2 of 0.88, 0.23, 0.96, and 0.71 for fractional cover, LAI, tree height, and crown length, respectively, with corresponding RMSE of 0.082, 0.85, 0.58, and 0.61. Subsequently, they expanded the database to include 100 000 combinations of seven ecological parameters: foliage chlorophyll content (Cab), foliage water content (Cw), foliage dry matter, LAI, fcover, maximum tree height (Tree_z), and vertical crown extent (C_ext)⁸⁷. They compared the waveform database with LVIS echo waveforms from ten areas with varying soil drainage conditions, from poorly drained to well-drained, at the International Paper's Northern Experimental Forests site near Howland, Maine, USA. The results yielded mean R^2 values of 0.97, 0.97, and 0.92 for

Tree_z, C_ext, and fcover, respectively, with RMSE values of 1.46, 1.72, and 0.08. Further comparison using LVIS data alongside a waveform database integrated with imaging spectroscopy resulted in mean R^2 values of 0.80, 0.80, and 0.76 for LAI, Cab, and Cw, with RMSE values of 1.15, 5.92, and 0.0058.

For maize (crop density: 9 plants/m², interrow spacing: 0.755 m) and wheat (crop density: 250 plants/m², interrow spacing: 0.17 m), Hmida et al. developed an echo waveform database based on three ecological parameters: height, LAI, and soil reflectance⁸⁸. Using echo waveforms acquired by a Leica airborne full-waveform LiDAR (ASL50-II) over sites in Lamothe, France, and Norunda, Sweden, they compared these measurements with the database to estimate ecological parameters for maize and wheat. The RMSE values for crop height were 0.07 m and 0.04 m for maize and wheat, respectively, while the RMSE values for LAI were 0.17 and 0.5. Bye et al. used the FLIGHT model to generate a database of 107 100 echo waveforms based on seven ecological parameters: LAI, maximum height to first branch (Hmax), minimum height to first branch (Hmin), fraction of ground covered by vegetation (Fc), ground slope (Sy), crown horizontal radius (Exy), and crown vertical radius (Ez)⁵¹. By comparing GLAS data with the database, they retrieved ecological parameters for different regions and investigated the impact of solution set size on inversion accuracy. The results showed that maximum R^2 values for Hmax, Fc, and Sy in the Forest of Dean, Gloucestershire, UK, were 0.74, 0.52, and 0.57, respectively, under different solution sets. For the southern boreal forest in Saskatchewan, Canada, maximum R^2 values for Hmax and Fc were 0.07 and 0.34, respectively, while corresponding values for Norunda, Sweden, were 0.28 and 0.13.

Based on prior knowledge provided by QuickBird optical imagery, Yang et al. employed a 3D geometric-optical and radiative transfer model to simulate echo waveforms for various building height combinations³⁷. The simulated waveforms were compared with GLAS-observed echo waveforms from a building area in Zhangye City, Gansu Province. When the Pearson correlation coefficient was maximized, R^2 between the simulated building heights (430 instances) and the actual building heights reached 0.971.

This method faces two constraints. The first is the prior knowledge. The mapping relationship between echo waveforms and target types (e.g., buildings, vegetation) or target attributes (e.g., 3D structure, biophysical and biochemical properties) has not been established. Currently, the target types cannot be inferred from waveform analysis and must be obtained through prior knowledge. Additionally, for each new target parameter added to the waveform database, the total number of waveform combinations increases by one to two orders of magnitude, similarly extending the database construction time by one to two orders of magnitude. To make database construction feasible, it is also essential to reduce the number of target parameters through prior

knowledge.

The second is that, this method has to involve the inherent trade-off between the efficiency of constructing an echo waveform database and the accuracy of inverting target characteristics. This comes from that the inversion accuracy for target parameters improves with a smaller parameter step size, a larger number of echo waveforms, and higher simulation fidelity. Previous studies have shown that DART achieves higher simulation accuracy for modulated echo waveforms in sparse forest scenarios, but the simulation time for a single waveform is 22 minutes⁴³. In this case, constructing a database of 1 00 000 waveforms using the DART model would require over 15 00 days (approximately 4.2 years) of continuous computation. When the simulation time per individual waveform cannot be reduced, it is difficult to reduce the step size and increase the number of waveforms necessitates a reduction in the simulation time per individual waveform.

Owing to the limitation prior knowledge and simulation time, waveform simulation has been less frequently employed in recent years to retrieve the properties of targets.

3.2 The analytical formula-based approach for target feature inversion

Compared to microfacet-based computation and radiative transfer models, the analytical echo waveform formula establishes a rigorous mathematical mapping between echo waveform characteristics and target properties through concise equations. By directly fitting experimentally measured echo waveforms using the analytical formula, multidimensional target characteristics can be retrieved.

Hu et al. conducted preliminary work in this area³⁸. When a pulsed laser vertically illuminates a 2D planar target, the echo waveform maintains a Gaussian shape with unchanged pulse width, while the peak intensity reflects target attributes such as shape, size, and location. Based on their derived analytical formula for the modulated echo waveform from vertical illumination on planar targets, Hu et al. obtained an explicit formula for the modulated echo peak intensity. Taking a rectangular target as an example, the echo peak intensity A is given by

$$A = D \left\{ \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (x + a) \right] - \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (x - a) \right] \right\} \left\{ \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (y + b) \right] - \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (y - b) \right] \right\}, \quad (10)$$

where D represents the combined effect of target reflectivity and other factors on the echo peak intensity, (x, y) and (a, b) denote the target's position and size, respectively, and ω refers to the beam radius. Generally, Eq.(10) is insufficient to simultaneously determine the position and size, owing to the "one-to-many" mapping relationship between the echo peak intensity and the various target features. To address this

challenge, Hu et al. proposed a multi-position redundant observation strategy. This method constructs a system of equations for peak intensity by combining echo data acquired from different viewing geometries

$$\begin{cases} A_1 = D \left\{ \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (x_1 + a) \right] - \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (x_1 - a) \right] \right\} \\ \quad \left\{ \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (y_1 + b) \right] - \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (y_1 - b) \right] \right\} \\ A_2 = D \left\{ \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (x_2 + a) \right] - \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (x_2 - a) \right] \right\} \\ \quad \left\{ \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (y_2 + b) \right] - \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (y_2 - b) \right] \right\} \\ \dots\dots\dots \\ A_n = D \left\{ \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (x_n + a) \right] - \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (x_n - a) \right] \right\} \\ \quad \left\{ \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (y_n + b) \right] - \operatorname{erf} \left[\frac{\sqrt{2}}{\omega} (y_n - b) \right] \right\} \end{cases}. \quad (11)$$

Assuming that the positional constraint of the target is

$$\begin{cases} x_1 = x_1, y_1 = y_1 \\ x_2 = x_1 + q_x, y_2 = y_1 + q_y \\ \dots\dots\dots \\ x_n = x_1 + (n-1)q_x, y_n = y_1 + (n-1)q_y \end{cases}. \quad (12)$$

Where (x_1, y_1) represents the initial position, and (q_x, q_y) denotes the moving step size and direction. Therefore, the unknown parameters for a 2D planar target may include size, initial position, moving step size and direction, and beam radius, with a maximum of seven unknowns, while the number of available equations (n) is unlimited. Eqs. (11) and (12) transform the mathematically unsolvable "one-to-many" mapping between echo peak intensity and target characteristics in Eq.(10) into a solvable equation system with multiple variables, enabling precise retrieval of the target's lateral characteristics within the laser footprint.

Hu et al. focused on three scenarios of target characteristics: only the target size unknown; both target size and initial position unknown; and all parameters, including size, initial position, step size, and beam radius unknown. Using simulated data, they systematically investigated the effects of the number of unknowns, noise, movement direction, target shape, and target size (with target areas corresponding to 0.917, 0.637, 0.407, 0.229, 0.102, and 0.025 times the footprint area) on the retrieved target size. The results showed that, when only the target size was unknown and when all target parameters were unknown, I) the average size errors across all noise levels, movement directions, target shapes, and target sizes were 0.56% and 4.3%, respectively, II) under various noise conditions, movement directions, and target sizes, the average errors for rectangles, squares, circles, and equilateral triangles were 1.48% and 9.84%, 0.23% and 2.8%, 0.23% and 1.23%, and 0.32% and 3.32%, respectively, III) as

target area increased from 0.025 to 0.917 times the footprint area, the average size errors under different noise levels, movement directions, and target shapes were 0.03% and 3.25%, 0.13% and 3.45%, 0.14% and 3.06%, 0.18% and 1.53%, 0.33% and 3.59%, and 1.03% and 7.56%, respectively, IV) in noise-free scenarios with only target size unknown, the average size error across all shapes, sizes, and movement directions was 0.31%. When considering only squares, circles, and equilateral triangles, the error was as low as 0.04%. In this case, the lateral sizing errors were all below the ICESat-2's longitudinal ranging accuracy of 0.3 mm⁸⁹.

Subsequently, Hu et al. developed a full-waveform LiDAR platform. A rectangle (7.06 cm long, 3.53 cm wide), a square (5.0 cm side length), a circle (2.82 cm radius), and an equilateral triangle (7.60 cm side length) were translated through a laser footprint region with a radius of approximately 5.0 cm in steps of 0.3 cm. Through jointly inverting the echo peak intensities, the retrieved sizes were determined to be 7.23 cm, 5.17 cm, 2.79 cm, and 7.49 cm, corresponding to errors of only 2.41%, 3.47%, 0.89%, and 1.42%, respectively. The absolute error in retrieved size did not exceed 0.17 m, with a relative error not exceeding 3.47% (3.4% relative to the spot radius). These experimental results confirm the capability for super-resolution, high-precision acquisition of the target's transverse characteristics within the scale of the laser footprint.

In the aforementioned studies, laser pulses vertically illuminated the targets, and the target shapes were assumed to be known a priori. Recently, for inclined 2D planar targets, Hu et al. proposed a CNN based algorithm incorporating attention mechanisms and feature engineering to classify target shapes⁹⁰. The algorithm extracts parameters such as the normalized peak intensity and peak position of the echo. It was trained on 576 echo waveforms simulated under varied noise levels, movement directions, sizes, and tilt angles for rectangles (with different aspect ratios), circles, and isosceles triangles (with different apex angles), and tested on 72 additional waveforms. This approach achieved a target shape classification accuracy of 82.87% under high-noise conditions. By selecting the corresponding analytical echo waveform formula, the average errors in retrieved target size and tilt angle across various shapes, sizes, tilt angles, noise levels, and movement directions were 13.31% and 6.42°, respectively.

While Hu's work demonstrates the significant potential of analytical formulas in acquiring target characteristics, several key breakthroughs are still required for this method to process actual echo data and retrieve features of terrestrial and aerial targets. First, mathematical description of target shape. It is fundamental of the waveform analytical method to accurately describe the properties, including shape, position, orientation, and surface optical properties of targets. While the properties of regular shapes are easily defined, describing the characteristics of irregular objects, such as forest canopies and building roofs, or streamlined

aerial targets like aircraft, remains a great challenge.

Second, target shape discrimination. Different target shapes correspond to distinct mathematical descriptions and echo waveform equations; therefore, selecting the correct analytical formula is a prerequisite for applying this method. Previous research indicates that when the area of a 2D planar target is less than 40% of the laser footprint area, identifying the target shape solely by analyzing the variation curve of echo peak intensity versus target position becomes challenging. In such cases, two solutions are typically viable: constructing a more universal general waveform equation or relying on prior knowledge about the target shape.

Third, simplifying assumptions in theoretical models. The derivation of current echo waveform equations is primarily based on two key assumptions: considering only single backscattering and treating targets as Lambertian surfaces. These assumptions face challenges in complex scenarios. In vegetation canopies with relatively sparse leaves, the clumping effects within and between crowns become significant, necessitating the inclusion of multiple scattering. Simultaneously, the optical properties of leaf surfaces are typically described using a Lambertian-Beckmann model. In contrast, for spatial or spaceborne targets, while the single scattering assumption generally holds, their surface optical properties require characterization by the complex bidirectional reflectance distribution function. Therefore, deriving a universal echo waveform equation that comprehensively accounts for canopy clumping effects, multiple scattering, and non-Lambertian surface optical characteristics remains a theoretical challenge.

Fourth, separation of echoes from complex targets. When a laser footprint covers multiple targets simultaneously, it becomes necessary to separate the modulated echo contributions from each individual target. Existing waveform decomposition methods can be applied if the transmitted and received echo waveforms share the same shape. However, new decomposition techniques are required when the echo waveform becomes distorted and deviates from the transmitted waveform due to modulation by target properties. Since the degree of this deviation depends strongly on attributes such as target shape, effective solutions must rely on acquiring prior knowledge of the target shape or employing echo waveform equations capable of describing general modulation relationships.

Fifth, Dimensional matching strategy. The number of parameters obtainable from a single-pulse echo waveform is generally fewer than the number of target characteristics to be inverted. This underdetermined mapping between target features and waveform metrics renders the retrieval of target features from echoes an ill-posed inversion problem. Consequently, it is essential to develop effective regularization methods. Feasible approaches include physical constraint strategies, such as leveraging the continuity of target motion (changing target position) or the spatial distribution of laser energy (varying laser footprint size), as well as a multi-pulse

joint observation strategy employing overlapping fields of view of an array detector. The essence of these methods is to transform the underdetermined inversion problem of a single observation into an overdetermined problem solvable through multiple observations, by introducing additional constraints or increasing independent measurements.

3.3 Extraction of target features with waveform decomposition and deconvolution

Recently, numerous target features have been extracted by full-waveform LiDAR in terms of waveform decomposition and deconvolution. Figure 2 exhibits some typical examples, and the main text discusses more detailed content.

3.3.1 Estimation of canopy and building heights

The most direct yet typical application of waveform decomposition and deconvolution is to evaluate the canopy and building heights, which holds significant scientific and application value. Forest canopy height is a fundamental ecosystem attribute that directly influences forest carbon storage patterns, ecosystem functioning, biodiversity conservation, and their response to climate change. Precise estimation of building height, on the other hand, is crucial for urban planning, disaster assessment, and sustainable development. It serves as the basis for calculating building shadow cycles, managing urban energy consumption, and researching regional climate and human-environment interactions, also playing a key role in urban aesthetics and renewal. Together, they underscore the importance of accu-

rate 3D structural quantification for understanding natural ecosystems and supporting the sustainable development of human society.

In waveform decomposition and deconvolution methods, the first and last components of the echo signals are generally considered to represent the canopy (or building) top and the ground, respectively, and subsequently the interval of peak intensity are the height of canopy or building

$$h = (t_{\text{last}} - t_{\text{first}}) h_{\text{bin}} , \quad (13)$$

where t_{last} and t_{first} are the peak positions of the last and first decomposed components, and h_{bin} refers to the length of one waveform time bin. In Blair et al.'s pioneer work, by processing LVIS waveform over Sequoia National Forest in California, they obtained the averaged canopy height of 68.52 m, 60.78 m, 47.08 m and 35.19 m above the ground⁵⁶. In a study on building height estimation in Zhangye, Gansu Province, Yang et al. reported that the Gaussian decomposition method (GLA14) achieved R^2 of 0.486 and rRMSE of 37.6% when comparing heights of 74 buildings with reference data from a Leica ALS70 laser scanner system³⁷. By applying a wavelet-transform-enhanced Gaussian decomposition approach, the R^2 increased to 0.903, while the rRMSE decreased to 18.1%. In a separate study focusing on mixed coniferous-broadleaf canopy height in the Harvard Forest region (Petersham, MA, USA), Chen et al. proposed the biosection approximation decomposition (BAD)⁹³. When comparing GEDI-derived canopy heights with field measurements from a National Ecological Observatory Network (NEON) Optech Gemini system, the

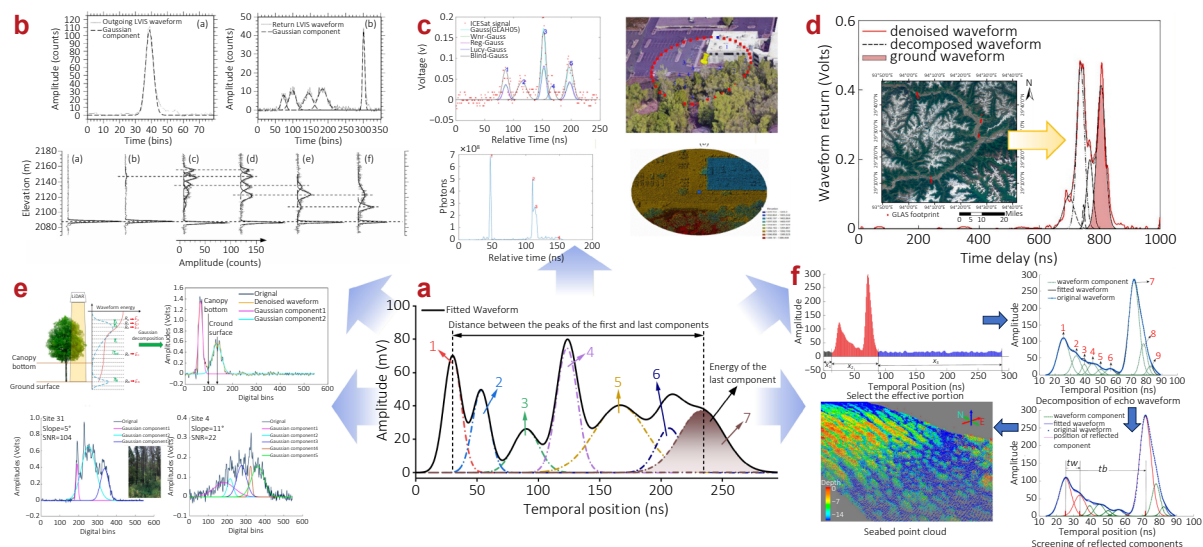


Fig. 2 | Typical examples of target feature extraction in terms of decomposition and deconvolution. (a) Illustration of waveform decomposition, (b) Gaussian decomposition for canopy height⁵⁹, (c) combined deconvolution and Gaussian decomposition approach for building height⁷⁰, (d) Gaussian decomposition for leaf index area estimation⁹¹, (e) Gaussian decomposition for canopy clumping index⁹², (f) GHPD for shallow laser bathymetry³⁶. Figure reproduced with permission from: ref.³⁶, IEEE; ref.⁵⁹, Elsevier Ltd; ref.⁷⁰, IEEE; ref.⁹¹, under a Creative Commons AttributionLicense; ref.⁹², under a Creative Commons AttributionLicense.

R^2 values for BAD, Gaussian, wavelet, and deconvolution methods were 0.67, 0.37, 0.25, and 0.23, respectively, with the corresponding RMSE were 3.33 m, 5.66 m, 7.49 m, and 8.94 m. BAD also achieved the highest mean accuracy of 84.57%, exceeding the comparative methods by 10.74%, 21.49%, and 28.93%, respectively.

In another investigation, Zhang et al. used the proposed combined deconvolution and Gaussian decomposition approach to process GLAS echo waveforms from several residential areas in Utah, USA, for building height retrieval⁷⁰. The results were compared with reference heights obtained from a Leica ALS 70 point clouds. As shown in Table S3, the conventional Gaussian decomposition yielded the lowest accuracy, whereas the Blind-Gauss method showed strong agreement. This confirms the superior performance of the blind deconvolution-Gauss hybrid method in range retrieval across both vegetated and urban landscapes.

3.3.2 Estimation of LAI

LAI is a key parameter that characterizes vegetation canopy structure and directly quantifies core ecological processes such as photosynthesis, respiration, and canopy interception. It serves as a fundamental metric for understanding the exchange of matter and energy between terrestrial ecosystems and the atmosphere. Consequently, LAI has been designated as a crucial Essential Climate Variable by global climate change research organizations. Passive optical remote sensing has been widely used for LAI retrieval, but it is limited by saturation effects⁹⁴. LiDAR technology can effectively mitigate this issue through direct measurement of vegetation vertical structure. Wang et al. systematically reviewed LAI retrieval methods under different LiDAR detection mechanisms, including terrestrial point clouds, airborne scanning, and spaceborne full-waveform systems, and evaluated the advantages and limitations of gap-based methods, contact-based methods, biophysical regression methods, and model inversion methods⁹¹. The study also provided a detailed analysis of the impacts of clumping effect, footprint size, height threshold and sampling size, occlusion, topography, types of return, voxel size, and woody material on LAI estimation. It highlighted that the clumping effect and woody material represent common challenges across all LiDAR systems.

When retrieving LAI using full-waveform LiDAR, the approach shares certain similarities with the estimation of forest canopy height or building height: the last echo is generally identified as the ground return, while the other echoes are attributed to the forest or vegetation canopy. The key difference lies in the fact that the retrieval of LAI relies on echo energy rather than the distance between peaks. Denoting the energy of the last echo as R_g and the total energy of the other echoes as R_v , the gap probability within the laser footprint can be expressed as

$$P_{\text{gap}}(0) = \frac{\frac{\rho_v R_g}{\rho_g}}{R_v + \frac{\rho_v R_g}{\rho_g}}, \quad (14)$$

where ρ_v and ρ_g represent the reflectivity of the canopy and the ground, respectively. Their ratio generally ranges between 1.0 and 2.0, with 1.5 commonly taken as the typical value. Under the assumption of a homogeneous scene within the laser footprint, the LAI calculation model can be derived by combining the BeerLambert law and the gap probability as follows

$$LAI_e = -\frac{1}{k} \ln P_{\text{gap}}(0), \quad (15)$$

where k represents the projection coefficient, which depends on the leaf angle distribution and the direction of the illuminating beam. To correct for the clumping effect caused by the spatial heterogeneity of canopy gaps, Yang et al. developed a physical method based on the gap fraction model⁹¹

$$LAI_{e-f_{\text{cover}}} = -\frac{1}{k f_{\text{cover}}} \ln \frac{P_{\text{gap}}(0) - 1 + f_{\text{cover}}}{f_{\text{cover}}}, \quad (16)$$

where f_{cover} denotes the fraction of crown cover within the laser footprint, which can be derived from Landsat-5. Yang et al. processed GLAS data for retrieving the LAI at 47 birch forest sites in Linzhi Country, Tibet, and compared the results with those obtained using the Tracing Radiation and Architecture of Canopies instrument. The results showed that LAI_e and $LAI_{e-f_{\text{cover}}}$ were 0.77 and 0.83, respectively. Compared to the more than 30% underestimation observed when directly using Eq.(15), Eq.(16) significantly improved the accuracy of LAI retrieval.

In Yang et al.'s study, an apparent underestimation (bias = $-0.49 \text{ m}^2/\text{m}^2$) was observed between GLAS-derived LAI and field-measured LAI due to the assumption that leaves are randomly distributed in crown-covered regions. To correct for within-crown clumping caused by non-random canopy organization, Jiang et al. proposed a method that compensates for occlusion effects, where lower canopy layers are obscured by upper layers during LiDAR sampling⁹⁵. By deriving relative path length distributions from occlusion-corrected waveforms to characterize canopy heterogeneity, their approach effectively corrected the within-crown clumping effects. The computational model for relative path length distribution is expressed as

$$p(l_r) = \frac{\hat{p}(l_r)}{\int_0^1 \hat{p}(l_r) d(l_r)}, \quad (17)$$

where $l_r = h/h_{\text{max}}$ denotes the relative path length, h and h_{max} are defined as the ratio of the actual path length to the maximum path length, $\hat{p}(l_r)$ represents the probability that l_r falls within the interval $[l_r, l_r + d(l_r)]$, where $d(l_r) = 0.025$ and $\int_0^1 p(l_r) d(l_r) = 1$. The LAI_PATH metric can be retrieved by simultaneously correcting for both between-crown and within-crown clumping effects through the

following relationship

$$P_{\text{crown}} = \frac{P_{\text{gap}}(0) - 1 + f_{\text{cover}}}{f_{\text{cover}}} = \int_0^1 \exp(-k(\rho h_{\text{max}}) l_r) p(l_r) d(l_r), \quad (18)$$

$$LAI_{\text{PATH}} = f_{\text{cover}} \int_0^1 (\rho h_{\text{max}}) l_r p(l_r) d(l_r), \quad (19)$$

where ρ is the leaf area density. With known P_{crown} , f_{cover} , $p(l_r)$, and k , the ρh_{max} can be solved from Eq.(18), then the LAI_{PATH} can be calculated using Eq.(19). The within-crown clumping correction in LAI retrieval can be realized with known gap probability and $p(l_r)$.

Based on simulated data, Jiang et al. systematically investigated the error characteristics of three LAI retrieval models under different scenarios. The study included cylindrical, spherical, and conical vegetation, as well as conical vegetation with varying configurations. The analysis focused on how model errors varied with leaf area density. The findings revealed that: I) Clumping effects depend on forest canopy structure. From cylindrical to spherical to conical shapes, the error of LAI_e decreased, whereas that of LAI_{e_fcover} increased. The LAI_{PATH} error for conical vegetation was higher than that for spherical vegetation. II) Clumping effects are related to crown dimensions. As crown length increased from 4 m to 8 m and 14 m, errors in LAI_e and LAI_{e_fcover} increased, though the change in LAI_{e_fcover} was not significant. III) Crown dimensions and leaf area density jointly influenced error levels. As leaf area density increased from 0.25 to 1.5, LAI errors under all three models generally rose. Across all scenarios, LAI_{PATH} errors were typically remaining around 10%, with a maximum error below 20%. In contrast, the maximum errors for LAI_e (which corrects no clumping effects) and LAI_{e_fcover} (which corrects only between-crown gaps) reached 69% and 47%, respectively. Neglecting between-crown clumping could lead to an underestimation of up to 46.7%, highlighting the importance of correcting horizontal canopy clumping to mitigate LAI underestimation. When applied to 17 GLAS waveforms from the Linzhi region of Tibet, LAI_{PATH} achieved a R^2 of 0.74 and an RMSE of 1.18 against TRAC-measured LAI, further validating the reliability of the method.

Wang et al. proposed an alternative method for correcting both between-crown and within-crown clumping effects⁹⁶

$$\begin{cases} \Omega_{\text{footprint}} = LAI_e / LAI_t \\ \Omega_{\text{cover}} = LAI_e / LAI_{e_fcover} \\ \Omega_{\text{crown}} = LAI_{e_fcover} / LAI_t \end{cases}, \quad (20)$$

where $\Omega_{\text{footprint}}$, Ω_{cover} , and Ω_{crown} quantify the relative errors introduced by the footprint-level clumping effect, between-crown clumping, and within-crown clumping, respectively. Wang et al. applied Eq.(20) to process GEDI data from 22 sites in the United States representing 12 temperate forest types, as well as from the Fenghuang Mountain area in Xuzhou, Jiangsu Province, China. They employed NEON discrete-return lidar data to conduct direct and indirect

evaluations of the estimated LAI. They also systematically assessed the impacts of various factors on LAI estimation accuracy, including differences in canopy gap fraction, sensor system parameters, canopy structural traits, topography, and soil properties. Without between-crown clumping correction, the GEDI LAI_e exhibited systematic underestimation in temperate forests (bias: $-0.56 \text{ m}^2/\text{m}^2$), with a mean absolute error (MAE) and RMSE of $0.70 \text{ m}^2/\text{m}^2$ and $0.89 \text{ m}^2/\text{m}^2$ respectively. Dense broadleaf and mixed forests generally showed higher underestimation in GEDI LAI_e than sparse coniferous forests. Factor analysis revealed that multiple factors collectively accounted for 52% of the variance in GEDI LAI_e errors, with differences in canopy gap fraction contributing the most (relative importance: 49.82%), followed by canopy characteristics (23.20%) and soil properties (16.18%). Canopy gap fraction was identified as a key mediating variable, and other factors influenced GEDI LAI_e errors primarily through their effects on this parameter. In the Fenghuang Mountain area, before between-crown clumping correction, the comparison between GEDI LAI_e and digital hemispherical photography (DHP)-derived LAI_t showed a bias, MAE, RMSE, %Bias, and %RMSE of $-1.05 \text{ m}^2/\text{m}^2$, $1.05 \text{ m}^2/\text{m}^2$, $1.12 \text{ m}^2/\text{m}^2$, -0.56 , and 0.6 , respectively. After applying the between-crown clumping correction Ω_{cover} , the bias, MAE, and RMSE all decreased by $0.05 \text{ m}^2/\text{m}^2$. This moderate but consistent improvement indicates that within-crown clumping was the dominant source of error compared to between-crown clumping, accounting for approximately 93% of the relative error. These findings underscore that clumping effect correction is essential for improving the accuracy of GEDI LAI estimation, particularly the correction of within-crown clumping in temperate coniferous forests.

To address the underestimation issue of GEDI's global LAI products in dense forests, Liang et al. explored the use of Digital Elevation Models (DEMs) as auxiliary data to separate ground and canopy signals in the echo waveforms⁹⁷. The total echo waveform and the ground return energy are denoted as

$$A = \sum_{i=1}^{n-1} A_i \exp\left[-\frac{(t-t_i)^2}{2\sigma_i^2}\right] + A_n \exp\left[-\frac{(t-t_n)^2}{2\sigma_n^2}\right] + \varepsilon, \quad (21)$$

$$R_g = 2 \int_{t_m}^{t_{DEM}} A_n \exp\left[-\frac{(t-t_{DEM})^2}{2\sigma_n^2}\right] dt. \quad (22)$$

The definitions of relevant parameters can be found in the corresponding references. Liang et al. first derived LAI by processing GEDI data at two sites in Fujian Province. After DEM correction, the R^2 improved from -0.55 to 0.81 , and RMSE decreased from 3.94 to 1.43 , demonstrating that calibrating ground return energy using DEM significantly alleviates LAI underestimation. To evaluate the method's applicability across different regions and forest types, the study was extended to the entire Fujian Province and the Conti-

mental United States (CONUS). The estimated GEDI footprint LAI was compared with three widely used global LAI products for consistency assessment. Results indicated that the underestimation of GEDI footprint LAI under different parameter settings ranged from 6% to 55% in Fujian and from 4% to 29% in CONUS. When using the recommended parameter settings of the GEDI L2B LAI product, underestimation rates reached 15% in Fujian and 13% in CONUS. After incorporating DEM to separate ground and canopy signals, LAI estimation improved notably in Fujian: R^2 increased from -0.024 to 0.040 , and RMSE dropped from 4.75 to 2.52 . In CONUS, R^2 changed from -0.229 to 0.157 , and RMSE declined from 5.24 to 1.96 . The substantial reduction in underestimation confirms that DEM can effectively serve as auxiliary data to mitigate severe LAI underestimation in dense forest environments.

3.3.3 Estimation of other biophysical and biochemical parameters of forests and vegetation

Full-waveform LiDAR can retrieve various forest and vegetation biochemical parameters, including canopy clumping index, canopy cover, among others.

To accurately characterize both between-crown and within-crown clumping effects and enable detailed interpretation of canopy structural attributes, Cui et al. developed a semi-physical model for estimating the canopy clumping index⁹². The method makes full use of the comprehensive vertical canopy profile information. By converting the original waveforms into forest gap distribution and random gap probability, and integrating radiative transfer theory with a modified Beer-Lambert model, the approach achieves robust retrieval of the clumping index

$$\Omega_E = \frac{\ln [P_{\text{gap}}(0)]}{\gamma \ln [P_r]} = \frac{\ln [P_{\text{gap}}(0)]}{\gamma \ln \left[\prod_{i=1}^{l_{\text{bottom}}} T_i \right]} = \frac{\ln [P_{\text{gap}}(0)]}{\gamma \ln \left[\prod_{i=1}^{l_{\text{bottom}}} (E_{i+1}/E_i) \right]}, \quad (23)$$

where l_{bottom} denotes the canopy base height, T_i represents the gap fraction between two successive recording layers, E_i refers to the pulse energy intensity profile at a resolution of 0.15 m, and γ indicates the needle-to-shoot area ratio obtained from literature review. Cui et al. tested the algorithm in the Saihanba National Forest Park and in Heilongjiang Province, China. They found that reliable clumping index estimation was achievable only when GLAS waveforms exhibited a high SNR (e.g., > 65) and were located in gently sloped terrain (e.g., $< 12^\circ$). Under such high-quality data conditions, the GLAS-based clumping index estimates showed strong agreement with TRAC CI ($R^2 = 0.72$, RMSE = 0.07) but were weakly correlated with MODIS CI (V005 MODIS CI: $R^2 = 0.09$, RMSE = 0.15 , bias = 0.07 ; V006 MODIS CI: $R^2 = 0.26$, RMSE = 0.12).

Canopy cover quantifies the proportion of the ground surface vertically occupied by canopy materials. As a key structural parameter, it is widely used in ecological and

climate models and plays an essential role in national forest inventories, land surface process modeling, forest dynamic monitoring, and LAI estimation models. While Yang et al. utilized optical remote sensing to assist in estimating canopy cover, its ability to detect subtle variations is often limited due to signal saturation and algorithmic uncertainties. To address this, Tang et al. employed GLAS data to calculate footprint-level canopy cover fraction as the ratio of vegetation return energy to the weighted total waveform energy⁹⁸. Comparing the results in the African tropical rainforest of Rabi, Gabon with LVIS-derived canopy cover and MODIS VCF products showed that the R^2 , and RMSE were 0.27 , and 21.6% , respectively. When integrating data from multiple regions across Arkansas, California, Maryland, Maine, Costa Rica, the agreement improved, with an R^2 of 0.57 , and RMSE of 16.87% . A global comparison between MODIS VCF and ICESat footprint-level canopy cover yielded an R^2 of 0.29 , and an RMSE of 26.28 . When stratified by forest type, R^2 values for evergreen needleleaf forest, evergreen broadleaf forest, deciduous needleleaf forest, mixed forest, shrubland, and savanna were 0.02 , 0.2 , 0.05 , 0.12 , 0.02 , and 0.1 , with corresponding biases of 25.66% , -2.81% , 7.4% , -7.5% , 14.96% , and -5.6% . These findings demonstrate that LiDAR-based improvements in canopy cover estimation can support comprehensive analysis of fine-scale forest structural changes in extremely dense canopy landscapes, providing unique insights into biophysical stratification and vertical canopy dynamics in densely vegetated tropical regions.

Li proposed that forest canopy cover can be derived as one minus the gap probability calculated using Eq.(14)⁵⁴. To validate the universality of this method across different forest types and evaluate the impact of geolocation bias, Li computed GEDI canopy cover for 41 forest areas. Table S4 exhibited the statistical comparison of the CC estimated from GEDI and other data (DHP, ALS, and simulated) across different forest types and waveform processings. The geolocation correction reduced RMSE by 0.05 – 0.15 and increased R^2 by 0.13 – 0.33 . After correction, R^2 values for all forest types exceeded 0.80 (average 0.85), and RMSE remained around 0.10 (average 0.10). Furthermore, the R^2 and RMSE between GEDI CC and the Global Forest Cover Change CC product were 0.567 and 0.203 before correction, improving to 0.722 and 0.157 after correction. Geolocation correction had the least impact on mangrove forests and the greatest impact on deciduous broadleaf forests.

Although the GEDI L2B product already provides vertical canopy cover profile data; however, their accuracy and uncertainties remain unquantified. Recently, Li et al. adopted the canopy cover calculation model proposed by Ni-Meister⁹⁹

$$CC = \frac{R_v(z)}{R_v(0)} \frac{1}{1 + \frac{\rho_v R_g}{\rho_g R_v(0)}}, \quad (24)$$

Where $R_v(z)$ and $R_v(0)$ represent the return energy from

the canopy top to height z and the total return energy from the canopy, respectively. They processed a total of 1 572 GEDI footprints, distributed across coniferous forest (317), deciduous broadleaf forest (182), evergreen broadleaf forest (323), mangrove forest (58), mixed forest (397), rainforest (135), and urban forest (160). Overall, GEDI CC showed good agreement with CC estimates from DHP, ALS, and simulated waveforms, with R^2 of 0.65, 0.71, and 0.71, respectively. Compared to ALS-based CC estimates, coniferous forests exhibited the highest correlation, while shrublands showed the lowest correlation but the smallest bias. As tree height increased from ground level to 35 meters, the MAE of GEDI CC decreased from 0.15 to 0.09. Among waveform processing algorithms, A2 and A6 yielded the best performance for total CC, achieving the highest R^2 values (≥ 0.6) and the lowest RMSE (≤ 0.23). The study highlights that employing customized waveform processing algorithms tailored to different vegetation conditions, adopting more realistic ρ_v/ρ_g ratios, and developing vegetation-specific waveform analysis methods can significantly improve the accuracy of CC estimation.

3.3.4 Artificial intelligence-assisted extraction of biophysical and biochemical properties

In above studies, researchers have established analytical inversion models for vegetation ecological parameters by integrating parameters including canopy return energy, and ground return energy within the theoretical framework of the Beer-Lambert law and gap probability models. However, due to factors such as spatial heterogeneity in vegetation structure and terrain effects, it remains challenging to develop analytical expressions between ecological parameters and waveform decomposition/deconvolution parameters based solely on theoretical derivation or empirical relationships. In this context, artificial intelligence demonstrates distinctive advantages: it can autonomously learn effective nonlinear mapping relationships from high-dimensional waveform features without requiring predefined functional forms. Through end-to-end feature learning and pattern recognition mechanisms, it effectively overcomes the reliance of traditional physical models on prior assumptions, thereby providing an efficient solution for vegetation parameter estimation in complex ecological environments.

AGB serves as a key indicator for assessing the impacts of global climate change, and its accurate estimation is crucial for quantifying terrestrial carbon storage and evaluating the global carbon cycle. Zhuang et al. pioneered the use of Gaussian decomposition-derived waveform metrics for biomass estimation¹⁰⁰. They defined the Gaussian Area Index

$$\text{GAI} = \int_b^a A \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] dx, \quad (25)$$

where x refers to the time bin of the waveform, and the start (a) and end (b) of the signal were determined by locating the

initial and final time bins where the waveform amplitude exceeded the mean noise level. Utilizing Gaussian centroid height information, Zhuang et al. innovatively developed two categories of Gaussian decomposition (GD) metrics: I) Percentile height GAI metrics, representing the cumulative GAI values of Gaussian centroid subsets within specific height percentile ranges, including GAIPer25, GAIPer50, GAIPer75, and GAIPer100; II) Height-weighted GAI metrics, calculated by weighting the waveform GAI based on Gaussian centroid heights, including GAHSum, GAIEHSum, GAIRatio, and GAIERatio. Analyzing LVIS waveform data with biomass ranging from 15.4 to 364.3 Mg/ha, the study incorporated eight newly developed GAI-based GD metrics and five previously unused GD-derived metrics into a RF regression algorithm to construct AGB estimation models. These were systematically evaluated against models using 12 non-GD metrics. Compared to non-GD metrics, GD metrics achieved significantly better prediction accuracy for biomass estimation in deciduous plots (19% lower RMSE, 25% higher R^2) and marginally better accuracy in coniferous plots (4% lower RMSE, 6% higher R^2). When combining GD and non-GD metrics, the performance for deciduous forests and coniferous forests was comparable to that of the GD-only models, indicating that incorporating non-GD metrics did not enhance the accuracy of the random forest-based biomass estimation. The superior performance of GD metrics can be attributed to their effective integration of both waveform amplitude and height information from Gaussian function parameters.

Luo et al. employed small-footprint full-waveform LiDAR data to construct voxelized waveforms at a $0.5 \text{ m} \times 0.5 \text{ m}$ resolution, which were then aggregated to generate pseudo-large-footprint waveforms at a $20 \text{ m} \times 20 \text{ m}$ quadrat scale³⁶. From the Gaussian-decomposed voxelized waveforms and pseudo-large-footprint waveforms, 14 waveform metrics were derived based on height and energy characteristics, including ground return ratio, height-to-median ratio, height of median energy, and front slope angle. Additionally, 14 LiDAR point cloud metrics were extracted, encompassing height percentiles, maximum height, mean height, standard deviation of height, and skewness of height. The study was conducted in plantation forests located in Huailai County, Hebei Province, covering young, middle-aged, and near-mature forest stands. Using a RF regression algorithm to develop AGB estimation models, the results showed that the models based on discrete point clouds, voxelized waveforms, and pseudo-large-footprint waveforms achieved R^2 values of 0.80, 0.81, and 0.84, respectively, and RMSEs of 16.00, 15.88, and 14.60 Mg/ha. The study confirms that using pseudo-large-footprint waveform metrics can achieve superior estimation accuracy for AGB.

3.3.5 Laser bathymetry with waveform decomposition

Airborne lidar bathymetry systems, known for their high

acquisition efficiency and high point cloud density, has been widely used in constructing DEMs for both terrestrial and seabed surfaces. In principle, laser bathymetry shares the same fundamental approach as laser ranging and the retrieval of forest or building heights: it involves separating the water surface signal from the bottom signal, with the peak distance indicating the water depth. However, differences in refractive index lead to relatively weak water surface echoes, and the laser beam attenuates rapidly in water. As a result, in shallow waters, the bottom and surface echoes tend to overlap significantly, making them difficult to distinguish; in deep waters, the bottom return is extremely weak and often undetectable. These issues pose considerable challenges for the practical application of laser bathymetry, leading to various efforts in addressing the challenge of decomposing airborne lidar bathymetry waveforms in very shallow waters.

Guo et al. proposed a Gaussian half-wavelength progressive decomposition (GHPD) method¹⁰¹. This approach progressively decomposes selected waveform segments using half-wavelength Gaussian functions and establishes an effective component selection mechanism to accurately retrieve reflection parameters, thereby mitigating unreasonable decomposition and peak shift caused by waveform superposition. In tests using simulated waveforms, the propagation time errors for CGD and GHPD were 0.42% and -6.62%, respectively. Two areas were selected to evaluate the method's performance in bathymetric measurements. Area A located in the northwestern part of an island with a fine sand bottom, and had an average water depth of approximately 6.5 m. Area B was a reef site where coral reefs and debris formed a complex and variable seabed topography, with a maximum depth of nearly 20 m. Comparisons between water depths derived from GHPD and those from the commercial Intelligent Waveform Digitizer software showed strong agreement: In Area A, for seven depth intervals from 2–4 m to 14–16 m, the average R^2 and RMSE were 0.935 and 0.050 m, respectively, with a waveform decomposition success rate of 87.718%. In Area B, the average R^2 and RMSE are 0.985 and 0.048 m, with a success rate of 90.298%. These results demonstrate the effectiveness of GHPD in full-waveform processing and information extraction.

To determine initial Gaussian decomposition parameters for water surface and bottom reflections, Fang et al. developed an innovative method that integrates a signal resolution enhancement model with fractional differential mathematics¹⁰². When applied to 503 echo waveforms collected from the Yanzhi Island area of the Western Xisha Archipelago in the South China Sea, across four depth intervals of 0–0.5 m, 0.5–1.0 m, 1.0–1.5 m, and 1.5–2.0 m, the overall average RMSE was 0.6244 ns, the average R^2 was 0.9585, and the average denoised signal correlation was 0.9950. Notably, the RMSE for water surface position identification was 0.62 ns, which is smaller than the system's 1 ns sampling frequency, corresponding to a ranging accuracy of

7.0 cm.

Qi et al. developed an optimized method integrating RL deconvolution (RLD) with curve fitting¹⁰³. This approach first applies RLD to obtain the system impulse response function, compress the width of water surface and seabed waveforms, and determine initial Gaussian parameters. The LM algorithm is then employed to iteratively optimize these parameters for precise extraction of water surface and bottom echo peak positions. When tested on an Optech Aquarius ALB dataset (pulse width: 8.3 ns) from Yanzhi Island, the seabed point recall rates in shallow waters were 72.3% for the Gaussian decomposition method based on peak detection (GDPD), 87.5% for the method using Grünwald-Letnikov fractional differentiation (GDFD), and 90.2% for the proposed method. For measured waveforms at 0.5 m and 1 m water depths, the GDPD method failed to separate water surface and bottom signals effectively. The GDFD method yielded depths of 1.17 m and 0.84 m after separation, while the proposed method achieved 0.31 m and 0.85 m. Further analysis of 10 000 simulated waveforms from the Wa-LiD model showed that in the 0.4–0.9 m, 0.9–1.4 m, and 1.4–2.0 m depth intervals, the average errors for GDFD and the proposed method were 2.48 cm vs. 0.75 cm, 2.14 cm vs. 2.21 cm, and 4.28 cm vs. 4.20 cm, respectively, with corresponding RMSE values of 3.36 cm vs. 0.83 cm, 2.53 cm vs. 2.27 cm, and 4.31 cm vs. 4.24 cm. These results demonstrate that while the new method performs comparably to GDFD in waters deeper than 0.9 m, it significantly outperforms GDFD in extremely shallow waters (< 0.9 m), effectively addressing the technical gap in bathymetric surveying for depths less than 1 meter.

Wu et al. proposed a multi-channel waveform fusion bathymetry (MWFB) method to achieve high-precision and reliable nearshore water depth detection¹⁰⁴. This approach employs adaptive multi-channel waveform fusion to suppress system and random noise, reconstructs energy-waveform curves using waveform curvature, and performs iterative decomposition based on curvature and energy features to improve decomposition accuracy and component reliability. Furthermore, it jointly detects water surface and bottom peaks by integrating multi-channel waveform curvature, thereby accurately capturing peak time delays for high-precision water depth retrieval. The MWFB method and conventional GD were applied to waveform data at three sites: Ganquan Island (max depth ~28 m), Lingyang Reef (max depth ~9 m), and Dong Island (max depth ~16 m). Results showed that MWFB achieved R^2 of 0.99, 0.90, and 0.99 with RMSEs of 0.41 m, 0.62 m, and 0.22 m at the three sites, respectively, while the conventional GD yielded 0.98, 0.84, and 0.99 with RMSEs of 0.64 m, 0.80 m, and 0.27 m. When focusing on shallow waters (≤ 3 m depth), MWFB achieved correlations of 0.89, 0.72, and 0.95 with RMSEs of 0.26 m, 0.41 m, and 0.10 m, compared to 0.76, 0.53, and 0.74 with RMSEs of 0.51 m, 0.31 m, and 0.33 m for the conventional method. These findings demonstrate that,

compared to traditional GD, MWFB enhances the robustness of water column and substrate echo detection through multi-channel information complementarity, improving both bathymetric accuracy and reliability, with particularly pronounced advantages in shallow water regions.

3.3.6 Laser bathymetry with artificial intelligence

The advancement of deep learning technologies has effectively mitigated the challenges of waveform overlap in shallow water environments. These methods avoid the intricate parameter and threshold determination required in conventional airborne lidar bathymetry approaches. This not only simplifies the data processing workflow but also enhances bathymetric accuracy, particularly in shallow waters.

Ji et al. proposed a novel full-waveform signal detection (FWSD) model for single-wavelength bathymetric LiDAR⁸². This method integrates a convolutional neural network, the model leverages local characteristics, global structure, and contextual information of waveforms. It enables automatic classification and segmentation of isolated, terrestrial, over-saturated, water surface, and bathymetric waveforms, thereby eliminating the need for complex feature extraction and iterative procedures in traditional methods. Simulation results showed that the overall classification accuracy of the waveform-characteristic-based RF approach, recurrent neural network, long short-term memory network, and the FWSD method reached 96.7%, 93.4%, 94.5%, and 96.9%, respectively. The FWSD model achieved the highest classification accuracy without relying on empirical threshold selection. In airborne full-waveform segment recognition conducted in the Wuzhizhou Island area of Haitang Bay, northeast of Sanya City, Hainan Province, the average RMSE values of the LM-GGM, RLD, and FWSD algorithms were 0.154 m, 0.137 m, and 0.137 m, respectively, with the FWSD model exhibiting an average bias of 0.09 m. These results demonstrate that the FWSD model can accurately classify terrestrial and non-terrestrial waveforms from raw data without auxiliary information, showing strong adaptability for shallow water waveform processing.

Chen et al. proposed a deep-learning bathymetry fusing waveform features and spatial-angular field features (DBWSF)¹⁰⁵. This approach constructs energy curves using waveform curvature to enhance feature representation and employs Gram angular difference fields to convert time-series waveforms into 2D images, thereby increasing feature dimensions and information content. A dual-path neural network incorporating an attention mechanism is constructed to accurately separate water surface and seabed waveforms, achieving simultaneous waveform decomposition and nearshore bathymetry. The multi-channel waveform fusion method, multilayer perceptron approach, and the DBWSF method were applied to analyze waveforms at Ganquan Island, Lingyang Reef, and Dong Island. Comparative results showed that R^2 of the three methods at Ganquan

Island were 0.987, 0.981, and 0.996, at Lingyang Reef were 0.893, 0.885, and 0.951, and at Dong Island reached 0.995, 0.983, and 0.998, respectively. The averaged RMSE at three sites of the DBWSF method was reduced by 0.33 m, 0.27 m, and 0.04 m compared to the multi-channel waveform fusion method, and by 0.47 m, 0.40 m, and 0.25 m compared to the multilayer perceptron approach, demonstrating a significant improvement in nearshore bathymetric and seafloor topographic mapping capabilities.

Guo et al. proposed a pregeneration–recognition method for detecting weak seafloor echoes in full-waveform airborne LiDAR bathymetry¹⁰⁶. This method used the backpropagation neural network model, which was trained on data from confirmed surface-seabed laser shots, to recognize true seafloor points from the pre-generated candidates and optimize the results. The proposed method was applied to analyze four data swaths collected by the Optech Aquarius ALB system near Wuzhizhou Island and Ganquan Island in the South China Sea. Results demonstrated that the method increased the number of detected seafloor points by 195.9% near Wuzhizhou Island and 40.1% near Ganquan Island compared to the system's standard processor, outperforming the RL deconvolution method.

Li et al. proposed a deep-learning assisted exponential waveform decomposition method for bathymetric LiDAR¹⁰⁷. This approach leverages ResNet-18 architecture to obtain high-quality initial estimates for the exponential waveform decomposition from raw input waveforms, eliminating the need for hand-crafted features or prior knowledge. Two networks with different sensitivities to weak signals were trained and used to initialize the parameters, followed by a least squares optimization to achieve the final waveform decomposition. The method was applied to real waveforms acquired by a RIEGL VQ-840-G system in the bay area of Toulon, a Mediterranean coastal city in France. Results demonstrated that the deep learning-assisted estimations were less influenced by intermediate peaks from in-water particles, producing a cleaner point cloud with fewer isolated points below the water surface compared to the conventional exponential decomposition.

3.3.7 Other applications of waveform decomposition and deconvolution

The utility of waveform decomposition and deconvolution extends far beyond the aforementioned examples, finding relevant applications in a number of other domains, including tree species and land cover classification, and forest pest and disease, etc.

Forest inventory requires detailed information on tree species composition. Tree species classification based on full-waveform features typically relies on symmetric Gaussian decomposition to extract parameters such as echo amplitude, pulse energy, and pulse width. However, the symmetry of echo waveforms in forested areas can be

affected by multiple scattering and the spatial distribution of leaves, among other factors. To address this, Bruggisser et al. introduced a decomposition method based on the skew normal distribution¹⁰⁸. This approach achieves comparable decomposition accuracy and computational cost to Gaussian decomposition, while also extracting high-order statistical moments—namely, skewness and kurtosis. These parameters are independent of echo energy and pulse width, providing additional dimensions for species discrimination. Using a RIEGL LMS-Q680i scanner to survey seven tree species (969 individual trees) in a temperate mixed forest, the study demonstrated that interspecies differences were primarily reflected in the energy and amplitude of the first echo, as well as the skewness of all echoes. Among 21 species pairs, the mean energy of the first echo showed discriminative power in 15 pairs, making it the most distinguishing feature. The mean amplitude of the first echo and the mean skewness of all echoes within individual tree crowns each exhibited differentiation in 14 species pairs. Compared to using echo amplitude alone, SND decomposition improved the Cohen's κ for conifer-broadleaf classification from 0.389 to 0.611 and increased overall accuracy from 89.76% to 93.45%. For classifying three dominant species within the test site, SND raised Cohen's κ from 0.202 to 0.256 and overall accuracy from 55.95% to 59.74%. Subsequently, Xu et al. extended the application of SND decomposition to the analysis of satellite models¹⁰⁹. By examining the modulated echo waveforms from the satellite body (cuboid) and solar wing inclined planes, and integrating analytical models, simulated waveforms, and experimental data, they achieved precise retrieval of solar wing dimensions and tilt angles.

To explore the potential of full-waveform LiDAR in land cover classification, particularly in remote mountainous areas, Liu et al. performed Gaussian decomposition on 1 367 GLAS waveforms (269 from flat terrain, 217 from buildings, 213 from terraces, 440 from forests, and 228 from mountains)¹¹⁰. They extracted a set of waveform metrics, such as the number of Gaussian components, waveform pulse width, global noise, relative timing between peaks, and waveform start and end points. Six classification algorithms, RF, Linear SVM, Radial Basis Function SVM, Logistic Regression, K-Nearest Neighbors, and Naive Bayes, were employed for terrain type identification. Results indicated that all methods achieved perfect classification for flat terrain (producer's and user's accuracy both 100%). RF performed best for the other four land cover types, with producer's and user's accuracies as follows: buildings (98.14%, 100.00%), terraces (97.24%, 95.49%), forests (98.64%, 96.03%), and mountains (94.64%, 100.00%). Naive Bayes model performs the poorest. Logistic Regression outperformed the Radial Basis Function SVM, Linear SVM, and K-Nearest Neighbors. These results successfully validate the feasibility of classifying land types at the laser footprint level using full-waveform features.

Accurate monitoring of forest pest and disease distur-

bances is crucial for forest management and the implementation of effective forest quality improvement measures. Leveraging the penetrating capability of laser pulses, LiDAR technology offers a potential solution for monitoring the health conditions of sub-canopy vegetation, though the influence of branch occlusion must be considered. To this end, Zhao et al. integrated 3D radiative transfer modeling with hyperspectral full-waveform LiDAR to systematically evaluate the potential of airborne hyperspectral LiDAR (AHSL) in monitoring forest pest and disease damage⁵⁶. For entirely damaged canopies, HI (hyperspectral imaging) achieved an overall accuracy and Kappa coefficient of 57.02% and 41.86%, respectively, while AHSL reached 89.45% and 85.92%. In upper canopy damage scenarios, HI results were 35.31% and 12.32%, compared to 85.37% and 80.41% for AHSL. For middle canopy damage, HI scored 34.43% and 11.86%, while AHSL attained 77.56% and 69.90%. In lower canopy damage cases, HI reached 33.99% and 11.15%, whereas AHSL achieved 65.95% and 54.58%. These results demonstrate that AHSL significantly outperforms HI in identifying pest and disease stress levels, with overall accuracy ranging from 65.95% to 89.45%, compared to 33.99% to 57.02% for HI. HI performed best only in fully damaged canopy scenarios and worst in lower canopy damage. In contrast, AHSL exhibited consistent performance across all canopy layers, with the lowest accuracy still observed in lower canopy damage (OA: 65.95%), yet substantially higher than HI. This study confirms that AHSL is a highly promising tool for forest monitoring, capable of effectively detecting structural and spectral changes in the middle and lower canopy, and holds significant potential for early monitoring of forest pests and diseases.

Tree height composition, which describes the relative abundance of trees across different height strata, is a key structural characteristic in community ecology. To explore height composition information derived from echo waveform features, Tan et al. conducted a study using simulated waveforms of fir, ash, and mango trees, along with real GEDI waveforms from the ABBY region⁴¹. They defined a set of trunk and crown architecture metrics, including the trunk-to-height ratio, crown-to-height ratio, upper-to-lower ratio, and the Gamma parameter. Simulated GEDI waveforms reasonably reflected variations in even-aged canopy structures and showed acceptable capability in identifying different height layers in multi-storied forests. The height of local maximum (HoI) intensity showed R^2 of 0.95, 0.88, and 0.95 with the trunk-to-height ratio, crown-to-height ratio, and upper-to-lower ratio, respectively, while increases in Gamma had no effect on HoI. For ash trees, total height was 12.3 times the HoI, and for mango trees, it was 1.15 times the HoI. In a three-layer canopy, the correlation between HoI and actual height was 0.39 for the top layer, 0.16 for the middle layer, and 0.43 for the bottom layer. This study demonstrates the considerable potential of GEDI waveforms in detecting vertical height composition and within-

canopy structure, providing essential guidance for future retrieval of tree height composition in more complex forest environments using actual GEDI observations.

To address the uncertainty in the backscattering cross-section and the unseparated reflectance in hyperspectral LiDAR, Ran et al. proposed a multi-echo reflectance correction method based on neighboring single-echo reflectance¹¹¹. By establishing a relational model between the initial multi-echo reflectance and adjacent single-echo reflectance, the method calculates stable and reliable reflectance correction coefficients. After correction, the R^2 between simulated multi-echo reflectance and spectrometer measurements exceeded 0.95, with RMSE below 0.042. For actual measured multi-echo reflectance, the R^2 with spectrometer values improved to above 0.56, while RMSE was reduced to below 0.15. The high-accuracy reconstruction of previously underestimated reflectance values lays a foundation for forest attribute assessment, enabling the vertical characterization of multi-target features using hyperspectral lidar reflectance profiles.

3.3.8 Limitation of waveform decomposition and deconvolution in retrieving target characteristics

The mathematical procedure for waveform decomposition is relatively straightforward, while the primary advantage of waveform deconvolution lies in its independence from prior knowledge or initial value estimation. Regarding target feature acquisition, both methods are capable of extracting a wealth of ecological parameter information from echoes, with the notable strength that their information resolution is independent of detection range. However, the intricate interaction between the pulsed laser and terrestrial or spatial targets fundamentally limits the capability for fine-grained target information retrieval based on waveform decomposition and deconvolution. This limitation is primarily manifested in three aspects.

First, the extraction of weak and overlapping echoes under wide pulse conditions. Limited by factors such as detector dynamic range, bandwidth, and environmental noise, the pulse width of spaceborne full-waveform LiDAR systems (e.g., GLAS, GEDI, Gaofen-7) is typically at the nanosecond level (5 ns, 15.6 ns, and 4–8 ns, respectively), corresponding to ranging accuracies between 0.6 and 2.25 meters. Under such wide pulses, the distortion of the echo waveform caused by modulation from a single object is often subtle, complicating deconvolution. Furthermore, when echoes from multiple closely spaced objects highly overlap, or when a weak echo from one target is masked by a strong adjacent echo, accurately determining the number of components during waveform decomposition becomes challenging.

Second, the selection of analytical function models for waveform decomposition and target impulse response function models for waveform deconvolution. The echo wave-

form is complexly modulated by target shape, size, position, and attitude. This necessitates using different analytical functions to decompose echoes from different targets, or even from targets of the same shape but different sizes (whose echoes may exhibit Gaussian or non-Gaussian forms). However, the modulation mechanism of how the target alters the echo waveform is not yet fully understood and remains a research focus in the field.

Third, the accurate interpretation of the physical meaning of parameters from individual waveform components. Currently, the peak locations of the first and last waveform components are commonly used to represent canopy or building height. The energy of the last echo combined with other echoes, alongside ground and vegetation reflectance, is used to invert ecological parameters like LAI, canopy cover, and clumping index. However, under narrow pulses, target modulation easily causes the echo to deviate from a Gaussian form. Although the 15.6 ns pulse width of GEDI ensures most echoes remain approximately Gaussian, the center position of its canopy/roof echo component still systematically deviates from the actual height. Moreover, correcting for within- and between-crown clumping effects lacks a solid theoretical foundation.

3.4 Target characteristic inversion method based on waveform metrics

The waveform metrics approach extracts characteristic parameters of echo waveforms and establishing relationships between these parameters and the properties of targets. This methodology has been endorsed by official institutions such as NASA. Commonly used waveform metrics include waveform quantile height, waveform height index, waveform energy index, among others. The forest canopy height and AGB accounts for the proportion of applications at 31% and 23%, respectively¹¹². While waveform simulation, decomposition, and deconvolution can also estimate canopy height and AGB, these methods often involve more complex procedures. This section focuses primarily on models for estimating forest canopy height, AGB, as well as other ecological parameters using waveform features, as shown in Fig. 3.

3.4.1 Height retrieval of forest canopy height based on waveform length

In terms of echo delay, the scattering echo from the top of the canopy exhibits the minimum delay, while the ground return shows the maximum delay. The time interval between these two echoes, i.e., the waveform length, corresponds to the canopy height. This principle forms the theoretical basis for estimating forest canopy height using waveform decomposition and deconvolution. However, the actual echo waveform is influenced by multiple factors, including pulse width, footprint size, terrain parameters, and forest type. To improve the estimation of forest canopy

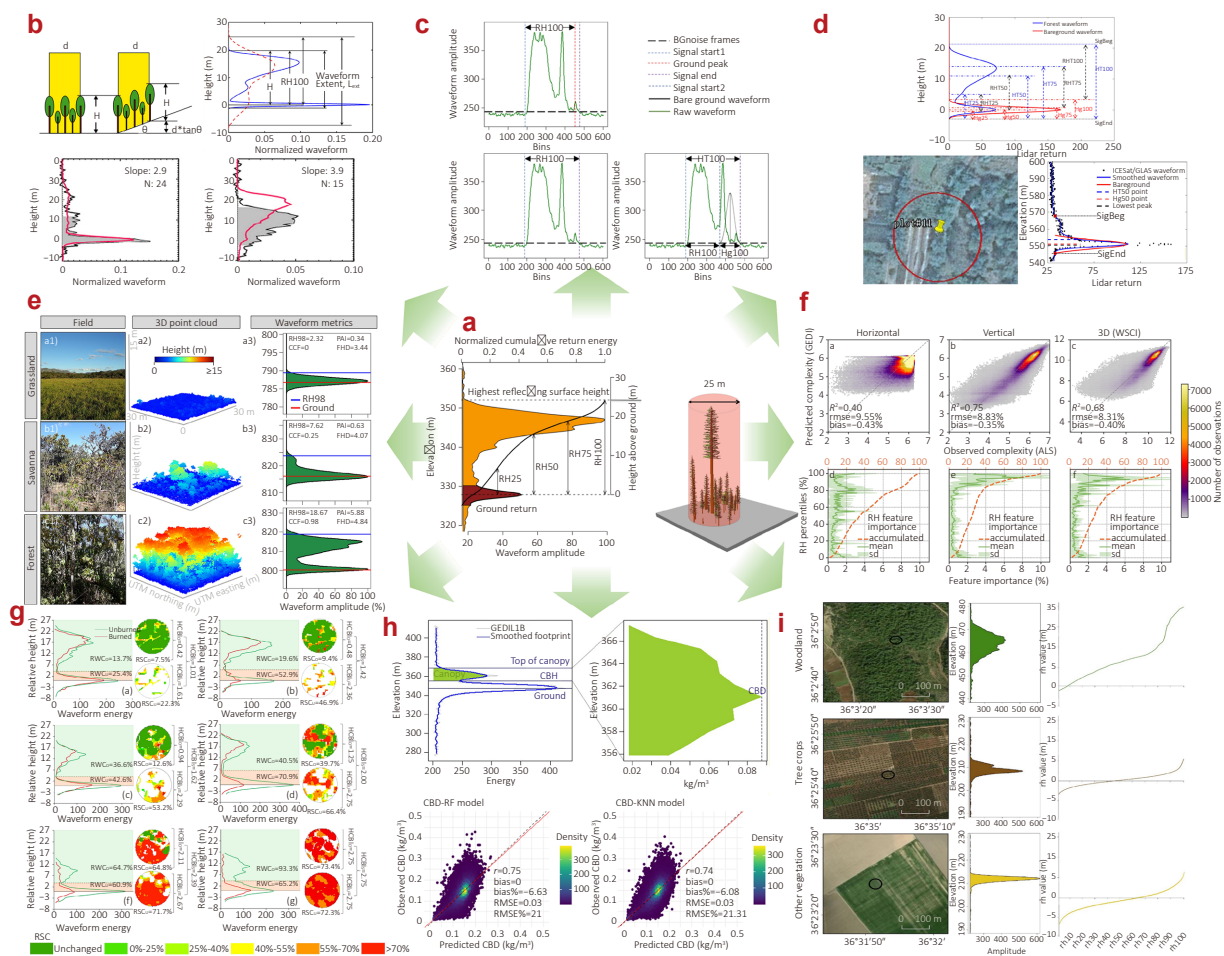


Fig. 3 | Typical examples of target feature extraction in terms of waveform metrics. (a) from official website of GEDI, (b) canopy height estimation using waveform length¹¹⁵, (c) optimization of TECIS LiDAR waveform metrics for forest height estimation¹¹⁷, (d) slope-adaptive waveform metric for AGB evaluation¹¹⁸, (e) fuel load distribution mapping⁵⁵, (f) 3D canopy complexity map¹¹⁹, (g) post-fire vegetation recovery patterns and ecosystem resilience¹²⁰, (h) canopy fuel load and canopy bulk density¹²¹, (i) tree crops classification¹²². Figure reproduced with permission from: ref.⁵⁵, Elsevier Ltd; ref.¹¹⁵, Elsevier Ltd; ref.¹¹⁷, Elsevier Ltd; ref.¹¹⁸, under a Creative Commons Attribution License; ref.¹¹⁹, Elsevier Ltd; ref.¹²⁰, under a Creative Commons Attribution License; ref.¹²¹, under a Creative Commons Attribution License; ref.¹²², under a Creative Commons Attribution License.

height, Lefsky et al. first proposed a terrain-corrected canopy height estimation model¹¹³, which was later refined by incorporating a waveform leading-edge correction factor¹¹⁴. Subsequently, Lee et al. developed a model that integrated corrections for terrain, footprint size, and reflected pulse width¹¹⁵. Xing et al. introduced an additional correction for surface roughness¹¹⁶. The key expressions of these models are presented below:

$$H_{\max} = b_0 (w - b_1 g), \quad (26)$$

$$H_{\max} = b_0 (w - b_1 g + b_2 l), \quad (27)$$

$$H_{\max} = w - d \tan \theta - c \times FWHM/2, \quad (28)$$

$$H_{\max} = b_0 (\ln w - b_1 g) + b_2, \quad (29)$$

where w denotes the waveform length, g represents the

terrain correction factor and can be derived from a DEM, l refers to the leading edge length of the waveform, which is defined as the distance from the start of the effective signal to the peak position of the first Gaussian component, d is the footprint radius, θ is the terrain slope angle, $FWHM$ is the transmitted pulse width, b is an adjustment coefficient, and c denotes the speed of light.

Furthermore, Du et al. defined the canopy height as the distance between the waveform signal start position and the ground return position. Using full-waveform LiDAR data from the Gaofen-7 satellite, they retrieved forest canopy heights in the Pu'er region of Yunnan Province under terrain slopes less than 20° ³³. Table S5 summarizes the study areas, regional forest characteristics, and estimation accuracy (in terms of R^2 and RMSE) of the forest canopy height using the above models applied to GLAS and Gaofen-7 echo waveforms.

3.4.2 Forest canopy and building height retrieval and optimization based on RH metrics

Although forest canopy height estimation models based on waveform length have incorporated corrections for topographic effect and transmitted pulse, their structures remain relatively complex. In practice, the effects of terrain slope, footprint size, and pulse width on canopy height retrieval can be effectively separated through refined waveform processing and analysis. Simard et al. introduced the concept of RH metrics¹²³, which pioneered the systematic use of relative height metrics for characterizing forest canopy and building height. GEDI further provides refined waveform processing capabilities, including RH1 to RH100 metrics, echo data derived from six combinations of smoothing widths and signal start/end thresholds, as well as elev_lowestmode and elev_highestreturn, etc. These data resources establish a robust foundation for high-precision retrieval of ecological parameters such as forest canopy height and building height.

To evaluate the performance of different RH metrics in retrieving forest canopy and building heights and identify the optimal metric, multiple researchers have conducted systematic studies. Potapov et al. estimated forest canopy heights in the United States, Mexico, Congo, and Australia using RH75, RH90, RH95, and RH100¹³; Liu et al. characterized canopy heights across five heavily forested regions in the U.S. with metrics including RH90, RH92, RH94, RH96, RH98, and RH100¹²⁴. Han et al. employed eight consecutive metrics from RH85 to RH100 to retrieve canopy height in the Penobscot Experimental Forest in Maine, USA¹²⁵. Utilizing data from 11 NEON sites across the United States, Zhu et al. conducted a systematic evaluation of the forest canopy height retrieval performance for a series of RH metrics, namely RH80, RH85, RH90, RH95, RH98, and RH100, under four distinct observation mechanisms: powerful beams at nighttime, coverage beams at nighttime, powerful beams at daytime, and coverage beams at daytime²⁹. Min et al. selected RH90, RH95, RH98, and RH100 to invert forest canopy heights in four areas, ABBY, UKFS, SCBI, and OSBS¹²⁶; Chen et al. extracted canopy heights in the Harvard Forest area using all metrics from RH90 to RH100 with a step size of 1⁹³. Liu et al. employed RH75, RH80, RH85, RH90, RH95, and RH100 to characterize forest canopy height across eight distinct vegetation divisions, including cold-temperature needleleaf forest¹²⁷. For building height estimation, Ma et al. utilized RH50, RH75, RH90, RH95, and RH100, and calculate mean RH (i.e., Mean RH) of all footprints within it to estimate building heights in the Yangtze River Delta region¹²⁸. As shown in Table S6, R^2 and RMSE values between these RH metrics and reference canopy/building heights indicate that RH metrics exhibits a consistently strong correlation with canopy height. However, the retrieval accuracy is significantly influenced by specific ground scenarios, and no universally optimal RH metric exists.

Furthermore, Liu et al. found that under nighttime powerful beam conditions, the R^2 for GEDI in retrieving forest canopy height increased from near 0 to approximately 0.9 as the RH metric increased from RH1 to RH100²⁷. Chen et al. elucidated the impact of RH on building height retrieval accuracy in six cities, Kitchener, Washington D.C., London, Brno, Hong Kong, and Christchurch³⁴. They observed that the MAE, RMSE, and relative RMSE (rRMSE) all decreased approximately linearly as the RH percentile increased from 50 to 94, while accuracy remained stable from RH94 to RH99, with RMSE ranging from 4.40 m to 4.55 m and MAE from 1.66 m to 1.90 m. Additionally, using the RH96 metric for building height estimation resulted in a lower rRMSE for flat-roof buildings, specifically 1.86% and 4.74% lower than for pitched-roof and curved-roof buildings, respectively.

To evaluate the capability of different RH metrics derived from full-waveform data in characterizing the maximum forest canopy height, Chen et al. proposed a model for estimating the maximum forest canopy height¹⁴

$$CH_y = RH_y - RH_x, \quad (30)$$

where RH_y and RH_x represent the top reference metric and the base reference metric for canopy height, respectively. y can take values such as 100, 95, 90, 85, 80, etc., while x may be set to values like 0, 5, 10, 15, 20, etc., or alternatively be assigned as the RH value corresponding to the last peak in the full waveform. Using full-waveform data from TECIS over the Quebec region of Canada, the study systematically compared RMSE between forest canopy heights derived from various RH base combinations and seven reference parameters from the canopy height model product within the laser footprint, namely CH100, CH95, CH90, CH85, CH80, CH75, and CHaver. The results indicated that using RH0 as the base significantly overestimated canopy height, whereas employing RH10 or higher led to underestimation. In contrast, using RH5 as the base considerably outperformed both the last peak position from waveform decomposition and other RH bases, with less sensitivity to variations in terrain slope. When RH5 was used as the base in combination with an appropriate background noise threshold, the forest canopy heights estimated from RH100, RH98, RH95, RH90, and RH85 closely matched the corresponding reference canopy height metrics. Moreover, the retrieval accuracy of canopy height from full-waveform data improved progressively as the selected RH metric decreased. With RH5 as the base, the RMSE for retrieving different percentile forest canopy heights ranged between 3.58 m and 4.23 m, and the MAE varied from 2.52 m to 3.21 m.

3.4.3 Study on factors influencing forest canopy/building height retrieval using RH metrics

Factors influencing the retrieval accuracy of forest canopy and building heights include waveform pre-processing

methods (e.g., signal thresholding), topographic relief, detection mechanisms, vegetation structural characteristics, and building attributes (such as height and roof shape).

Regarding signal threshold selection, Liang et al. conducted a systematic analysis of the effects of signal threshold settings and waveform metric calculation methods on canopy height retrieval accuracy, using the TECIS LiDAR waveform metrics framework in forested regions of Northern China¹¹⁷. The study proceeded through three phases: I) Four signal truncation and noise calculation grouping settings were combined with six multiples of noise standard deviation to form 24 waveform threshold combinations; II) Three categories of waveform metrics were proposed: rH (waveform-derived ground and signal start/end positions), rh (relative energy height based on signal start and end positions), TrH (simulated terrain combined with signal start/end positions), along with the waveform signal leading/trailing edge length (LE1, TE1), and the leading/trailing edge length of the waveform signal; III) Seven waveform metric combinations, (rH100), (rH100, TE1, LE1), (rH100, TE2, LE2), (rh100), (rh100, TE1, LE1), (rh100, TE2, LE2), and (Trh100), were employed to construct regression models. The results demonstrated that stricter background noise thresholds or more conservative signal truncation settings both reduced the retrieved canopy height values. The rH metrics showed the best response to maximum canopy height. Furthermore, models integrating RH metrics with leading and trailing edge lengths achieved optimal performance, with the best model yielding a correlation coefficient R^2 of 0.86 under the strictest signal truncation condition. In high-slope areas, parameters from the slope-adaptive metrics algorithm outperformed traditional waveform length metrics.

To address the influence of topographic relief on observation accuracy, Tang et al. proposed a correction method integrating the GEDI Simulator and the Shuttle Radar Topography Mission DEM¹²⁹. This approach generated simulated bare ground echo waveforms to correct the ground elevation based on Gaussian components from GEDI L1B waveform decomposition, and to adjust the canopy top elevation using the signal start position from GEDI L2A data. The study utilized 1 090 GEDI echo samples (331 from Site 1 in the western United States, dominated by evergreen forests, and 759 from Site 2 in the south-eastern United States, covering deciduous, evergreen, and mixed forests). The results showed that for maximum canopy height in the L2A product, the RMSE and R^2 were 7.601 m and 0.526, respectively. After correction, these metrics improved to 6.538 m and 0.548, respectively. The RMSE for ground elevation, canopy top elevation, and maximum canopy height improved by 9.2%, 15%, and 14%, respectively, compared to the L2A product. Wang et al. proposed a geolocation error correction method for forest footprints in plain areas¹³⁰. Assisted by WorldCover's tree cover data, the method was applied to five plain regions,

namely Lixin County, Lujiang County, Nanqiao District (plain areas only), Jin'an District, and Yongqiao District. After correction, R^2 of GEDI RH100-based forest canopy height retrieval improved from 0.93 to 0.96, mean error reduced from 2.15 m to 1.68 m, and RMSE dropped from 3.09 m to 2.27 m.

Cho et al. proposed an improved terrain correction method, which begins by calculating the residual between the GEDI-derived ground elevation and the synchronized airborne laser scanning digital terrain model data¹³¹. The residual is then used to systematically correct the GEDI canopy height within each footprint, effectively eliminating biases introduced by topographic variations

$$h_p^{\text{surface-adjusted}} = h_p + (h_{\text{lowest_mode_elevation}} - h_{\text{ALS-based_terrain_height}}), \quad (31)$$

where h_p represents the GEDI RH. This value is then systematically calibrated using a site-level additive offset calculated via Theil-Sen regression with synchronized airborne laser scanning canopy height model data

$$\hat{h}_p^{\text{surface-adjusted}} = h_p^{\text{surface-adjusted}} + \alpha_0, \quad (32)$$

where α_0 denotes a site-level regression intercept coefficient determined by regressing the study site $h_p^{\text{surface-adjusted}}$ values (as the response variable) against the coincident $h_{\text{ALS-based_terrain_height}}$ values (as the predictor variable). Taking $p = 95$, the study analyzed GEDI data from three sites: tropical evergreen lowland forest in the Democratic Republic of the Congo (Mai Ndombe Province, MNDP), temperate pine-broadleaf mixed forest in Alabama, USA (Talladega National Forest, TALL), and temperate mixed forest in Maryland, USA (Smithsonian Environmental Research Center, SERC). Under the condition of beam sensitivity ≥ 0.9 , the accuracy was compared across three stages: initial GEDI h95 versus ALS-based canopy height reference, GEDI h95 after outlier removal versus ALS reference, and $h_p^{\text{surface-adjusted}}$ versus ALS reference. In terms of RMSE, the MNDP site decreased from 7.59 m to 7.52 m and 5.29 m; TALL decreased from 6.28 m to 5.59 m and 4.71 m; SERC decreased from 9.69 m to 6.73 m and 4.37 m. After continued outlier removal, the RMSE values for MNDP, TALL, and SERC further improved to 3.37 m, 3.53 m, and 4.27 m, respectively. R^2 for MNDP, TALL, and SERC were 0.67, 0.70, and 0.61, respectively.

Regarding detection mechanisms and forestry characteristics, Li et al. conducted the first baseline validation of the RH98 metric across multiple savanna vegetation types using 22 813 GEDI footprints from 11 sites³⁰. They also identified the influence of various factors on estimation error, including algorithm setting groups, beam type, diurnal acquisition mode, beam sensitivity, and vegetation phenology. In terms of algorithm settings, the errors for setting groups SGs 4–6, designed for dense forests, were 0.341, 0.129, and 0.361, significantly higher than those for SGs 1–3, which were

0.479, 0.448, and 0.455, respectively. For beam type, diurnal mode, and beam sensitivity, the coverage beam achieved 92% sensitivity during the day and 93% at night, while the power beam achieved approximately 96% sensitivity for both day and night acquisitions. The results indicate that acquisition time and beam type had no significant effect on the bias or RMSE of RH98 observations, a conclusion that contrasts with findings from Liu et al. in U.S. forested regions²⁷, but consistent with the finding of Chen³⁴ and Li³¹. Vegetation phenology significantly affected RH98 observation errors: when savanna vegetation with $RH98sim < 15$ m was in the leaf-on season, RH98 accuracy was higher ($R^2 = 0.61$, and $RMSE = 1.48$ m), while during the leaf-off season, accuracy decreased noticeably ($R^2 = 0.43$, and $RMSE = 2.27$ m).

Focusing on 53 967 770 conifer-dominated plantation trees in Aichi Prefecture, Japan, Li et al. employed RH98 as a proxy for forest canopy height to systematically examine the effects of beam sensitivity, terrain slope, and forest structural parameters, including height-to-diameter ratio, canopy-to-length ratio, and stem density on the accuracy of GEDI canopy height retrieval³¹. The study found that beam sensitivity considerably influenced retrieval accuracy: differences between airborne lidar and GEDI RH98 values were 6.8 m, 1.49 m, and -0.27 m in the sensitivity ranges of 0.90–0.95, 0.95–0.98, and 0.98–1.0, respectively. After geographic location correction, the RMSE of GEDI RH98 decreased from 10.18 m to 6.10 m. Across eight slope intervals from 0° – 5° to 35° – 40° , the RMSE values increase monotonically from 2.51 m to 10.35 m. For stem density grouped into eight classes from 0–10 to 70–300 stems, the RMSE values decrease from 10.03 m and stabilized at about 6 m. When height-to-diameter ratio ranged from 30%–50% to 110%–300%, RMSE values were 8.95 m, 6.22 m, 5.68 m, 6.11 m, 6.03 m, 6.52 m, 7.87 m, and 3.20 m; for canopy-to-length ratio across eight intervals from 0–10% to 70%–80%, RMSE values nearly held at around 6 m. These results indicate that the accuracy of GEDI RH98 retrieval is primarily influenced by beam sensitivity, geographic location, terrain slope, and canopy-to-length ratio, while its association with height-to-diameter ratio and stem density is relatively weak.

Li et al. further conducted a quantitative evaluation of the accuracy of GEDI waveform data under six different processing algorithms for forest canopy height retrieval and performed a systematic error analysis³². By examining how RMSE varies with parameters, they identified the number of peaks and canopy cover as the most significant error factors. Subsequently, they developed an integrated error correction model using the RF algorithm, which simultaneously accounts for multiple error sources including stand density and topographic effects. After correction, the R^2 between RH100-derived canopy height and NEON reference canopy height across 19 forest sites in the United States improved from 0.75 to 0.91, while the RMSE decreased from 5.60 m to 4.19 m, achieving a substantial improvement in forest

canopy height retrieval accuracy.

Based on refined processing of GEDI data, including the exclusion of geolocation errors, Chen et al. selected RH96 as the metric for building height and analyzed buildings in multiple cities: Kitchener (87 footprints), Washington D.C. (153 footprints), London (10143 footprints), Brno (393 footprints), Hong Kong (523 footprints), and Christchurch (119 footprints)³⁴. Their study systematically revealed the influence of building height, slope, detection mechanism, and roof type on the accuracy of building height retrieval. The results showed that GEDI achieved the highest accuracy for mid- and high-rise buildings (RMSE approximately 2–6 m, rRMSE about 15%–25%), whereas its performance was lower for low-rise buildings (< 5 m). Sensitivity and slope were identified as the most critical factors affecting accuracy: data with sensitivity above 0.95 had an rRMSE that was 4.66% lower than data with sensitivity above 0.90. Slope had a significant negative impact on accuracy. Within the lower slope range (0° to 4°), both RMSE and rRMSE exhibited some fluctuation; however, as slope increased from 8° to 28.60° , both RMSE and rRMSE showed substantial increases. Roof type had a moderate influence: flat roofs slightly outperformed sloped and domed roofs, with a R^2 of 0.96 (0.11 and 0.15 higher, respectively) and an rRMSE of 15.54% (1.86% and 4.74% lower, respectively). Beam type and acquisition time had no significant effect: compared to the coverage beam, the power beam reduced MAE and RMSE by 0.29 m and 0.25 m, respectively, but increased rRMSE by 1.59%; the rRMSE values for daytime and nighttime acquisitions were 28.22% and 28.36%, respectively, differing by only 0.14%.

3.4.4 Comparative analysis and integration of RH metrics for forest and building height retrieval

Evaluating the accuracy and applicability of different methods for characterizing forest canopy and building height is essential for their practical implementation. To systematically compare the performance of waveform decomposition and waveform metric-based methods in forest canopy height retrieval, Zhou et al. proposed a new approach: the canopy top elevation was determined from the starting interval of the RL deconvolved waveform, while the ground elevation was derived with the centroid elevation of the ground return by combining Shuttle Radar Topography Mission data and the deconvolved waveform¹³². The difference between these two elevations was taken as the canopy height. Results from GEDI waveform processing in rugged terrain at Mount Lukens, Colorado, a region dominated by coniferous forest, showed that R^2 and RMSE values for the RH100 method, GD, the simulated ground echo-assisted correction method, and the proposed method were 0.615 and 7.220 m, 0.575 and 8.257 m, 0.617 and 7.523 m, and 0.666 and 5.813 m, respectively. Compared to GEDI RH100, the new method improved the ME, MAE, and RMSE by 3.011 m, 1.330 m,

and 1.406 m. Further analysis across three slope ranges, 20°–25°, 25°–30°, and > 30°, revealed that compared to GEDI RH100, the proposed method reduced the RMSE by 0.10 m, 1.14 m, and 3.26 m, respectively, confirming its strong suitability for canopy height retrieval in rugged mountainous areas. It should be noted that the new method incorporated SRTM-based topographic correction, whereas GEDI RH100 does not inherently include such correction. Moreover, among GEDI RH metrics, RH100 does not necessarily represent the highest retrieval accuracy.

To conduct a detailed comparison of the accuracy and influencing factors of terrain and canopy height retrieval between GEDI and ICESat-2, Liu et al. used data from 40 sites across the contiguous United States, Alaska, and Hawaii, which cover diverse eco-climatic zones and vegetation types²⁷. For terrain retrieval, both ICESat-2 and GEDI showed perfect correlation with NEON reference elevation data. In mid- to low-latitude regions, ICESat-2 achieved an RMSE of 2.24 m and a bias of −0.20 m, outperforming GEDI, which had an RMSE of 4.03 m and a bias of 0.97 m. The study identified steep slopes (>30°) as the primary challenge for both missions, and tall (>20 m), dense (canopy cover >90%) forest ecosystems were also found to reduce terrain estimation accuracy. In forest canopy height retrieval, under full-time all-beam conditions, ICESat-2 and GEDI achieved R^2 of 0.61 and 0.82, with overall RMSE values of 7.21 m and 5.02 m, respectively. When using only nighttime data from powerful beams, their R^2 improved to 0.89 and 0.90, and overall RMSE decreased to 3.93 m and 3.56 m, respectively. This indicates that ICESat-2, with its photon-counting mechanism, is more sensitive to natural light and beam intensity. The RMSE values for ICESat-2 and GEDI remained stable at 4–5 m and 3–4 m, respectively. Overall, GEDI demonstrated superior forest canopy height retrieval accuracy over ICESat-2 under nearly all topographic conditions, forest types, and coverage scenarios.

Despite that GEDI and ICESat-2 are good at canopy height retrieval and topographic mapping, respectively, the potential and synergistic value of integrating both datasets to derive consistent and reliable forest canopy height products have not been fully explored. To address this gap, Zhu et al. proposed a calibration algorithm designed to enhance the accuracy of GEDI forest canopy height data²⁹

$$H_{\text{corr}} = \frac{H \cos \gamma}{\cos \eta} - \sqrt{a^2 \cos^2 (\beta - \varphi) + b^2 \sin^2 (\beta - \varphi)} \frac{\sin \gamma}{\cos \eta}, \quad (33)$$

where a and b are the semimajor and semiminor axes of the footprint, respectively, θ is the intersection angle of the azimuth of the laser pointing direction β and the orientation of the satellite LiDAR footprint φ , η is the terrain slope, and γ is the intersection angle between the laser pointing direction and the vertical terrain direction. Liu et al. utilized 11 sites from the NEON as study areas, employing RH95 (for GEDI) and RH100 (for ICESat-2) as forest canopy

height metrics. They analyzed the influence of forest coverage, slope, site, and forest type on the discrepancies between the retrievals from the two systems. The analysis revealed that the difference between GEDI-derived and ICESat-2-derived heights increased with forest coverage, while no clear pattern was observed across different forest types, slope ranges, or sites. Subsequently, the study established consistency models to predict GEDI heights from ICESat-2 metrics using stepwise regression (SR), RF, and a combined SR-RF algorithm. The results showed that the combined SR-RF model performed best, achieving an average R^2 of 0.86 and an RMSE of 2.56 meters on the ICESat-2 ATL08 data, which represented an improvement of 0.37 in R^2 and a reduction of 2.71 meters in RMSE compared to baseline models. By incorporating training data covering all forest types and sites, a universal consistency model was developed, demonstrating the potential of jointly using ICESat-2 and GEDI data to acquire a consistent forest height dataset.

Kaya et al. conducted a direct comparison of the performance of GEDI and ICESat-2 in building height retrieval and developed two integrated inversion methods¹³³. They used the interquartile range (i.e., the difference between the third and first quartiles) to characterize building height. For 10 698 buildings in New York, this method achieved a R^2 , and RMSE, and MAE of 0.62, and 11.31 m, respectively; for 15 959 buildings in Los Angeles, the corresponding values were 0.18, and 8.27 m. When using ICESat-2 data, the R^2 , and RMSE for 41 451 buildings in New York were 0.54, and 8.06 m, respectively. For 69 094 buildings in Los Angeles, the R^2 , and RMSE were 0.28, and 6.17 m. These results indicate that ICESat-2 outperforms GEDI in building height retrieval accuracy, with comparable R^2 but significantly lower RMSE. By integrating GEDI and ICESat-2 data, the R^2 , and RMSE for 50 654 buildings in New York were 0.48, and 8.28 m, respectively, with 46% of buildings having errors below 5 m and 7% below 1 m. For 84 045 buildings in Los Angeles, the corresponding metrics were 0.18, and 6.42 m, with 67% of errors below 5 m and 7% below 1 m. The integrated approach improved retrieval accuracy compared to using GEDI alone, while achieving performance levels comparable to those of ICESat-2 used independently.

3.4.5 Analytical retrieval of AGB from waveform features

Previous studies have confirmed a strong correlation between full-waveform LiDAR metrics and AGB^{134,135}. In 2005, building upon the forest canopy height model, Lefsky et al. proposed a linear regression model between AGB and the square of canopy height¹¹³. In the Santarem sites, the R^2 between the estimated AGB and field measurements was 0.73, with an RMSE of 58.3 Mg/ha. In recent years, with the launch and operation of the GEDI mission, which includes AGB estimation as one of its primary objectives, a wide range of AGB retrieval models based on full-waveform LiDAR metrics data have been developed.

The GEDI mission directly provides footprint-level AGBD products. Duncanson et al. used a data-driven approach to detail the development process of the model suite for this product, including variable selection, data filtering and transformation, model tiering strategies, and model selection procedures²⁸. They also summarized the forest types (boreal, temperate, tropical, and pantropical), LiDAR types, acquisition dates, geographic extents, modeling approaches, predictors for the best AGBD model, accuracy of the best model, plot sizes, and number of plots associated with these models. Modeling approaches included nonlinear regression, k-nearest neighbors, functional regression, partial least squares regression, stepwise multiple regression, support vector regression, k-MSN, ordinary least squares, power-law models, jackknife regression, maximum entropy, and random forest. Predictors encompassed RH metrics, density metrics, waveform metrics, density decile metrics, canopy height metrics, canopy cover, LAI, canopy volume, crown area, NDVI metrics, height of median energy, median height of CHM, hyperspectral bands, gap fraction, basal area, wood density, and canopy energy. RH predictors included RH10 to RH98 as individual metrics and in combined forms (for example, $RH10 * RH50$, and $RH60 * RH98$, etc), while data transformations mainly involved $\sqrt{\text{sqrt-sqrt}}$, original-log , and original-sqrt .

Leveraging this database, Duncanson et al. expanded the geographical scope of previous full-waveform LiDAR studies and stratified the globe into four strata based on plant functional types and six geographic regions. Using a parametric ordinary least squares model with simulated RH metrics as predictors, they developed a complete footprint-level AGBD prediction model by selecting the best predictor combinations and data transformations for each GEDI geographic stratum. Key findings included: I) Model selection often prioritized combinations of RH98, RH90, RH50, and RH10; incorporating lower RH metrics (e.g., RH10) did not significantly improve performance. II) Enforcing the inclusion of RH98 in all models was crucial and did not degrade performance; the best models remained parsimonious, typically containing only 1–3 predictors. III) Stratifying by geographic region (PFT and geographic zone) significantly outperformed a global unstratified model. IV) Most optimal models employed square-root transformations of field AGBD and/or height metrics. Model performance varied considerably across geographic strata, with the poorest performance in regions with sparse training data and/or high AGBD values.

Recently, Kellner et al. elaborated on the algorithmic foundation for developing the GEDI L4A footprint-level AGBD data product¹³⁶. Using various RH metric combinations and 13 linear models applied to 8 587 high-quality simulated GEDI waveforms from 21 countries worldwide, they predicted AGBD through 32 plant functional type and geographic region combinations, resulting in the most geographically comprehensive dataset to date for GEDI

model development using lidar remote sensing. Table S7 displayed the performance of these models as well as other AGB inversion models in detail.

Although the GEDI L4A product employs 13 estimation models, it exhibits limited variability in RH metric combinations: all models include the combination of RH98 and RH100. Beyond that, five models use RH50 and RH100, two use RH60 and RH100, and four use RH70 and RH100. In terms of vegetation categories and geographical coverage, the model system does not encompass Asian continental forests, tropical dry deciduous broadleaf forests and savannas, or northern Australian evergreen broadleaf forests, and also lacks representation of typical vegetation types such as deciduous coniferous forests, bamboo forests, and mixed forests. Furthermore, the models do not yet incorporate the influence of environmental factors such as topography. Therefore, the two aforementioned review studies do not cover all LiDAR-based AGB estimation models or all typical geographic regions. Several other representative studies will be introduced below.

To investigate the relationships between discrete-return LiDAR point cloud metrics, full-waveform LiDAR waveform metrics, and their combinations with AGB, Nie et al. performed Gaussian decomposition on echo waveforms in the Dayekou coniferous forest area of Zhangye, Gansu Province¹³⁷. As shown in Table S7, a series of airborne discrete-return LiDAR-derived metrics (e.g., $H25_{DR}$, H_{SD}), full-waveform LiDAR-derived metrics (e.g., $H25_{FW}$, $H100_{WF}$), and structural parameters integrating height metrics with canopy cover (e.g., $H100_{DR} * CC_{DR}$, $H100_{WF} * CC_{FW}$) were derived to evaluate the effect of nonlinearly fusing discrete-return and full-waveform metrics on AGB estimation. The results indicated that canopy cover from discrete returns was insensitive to AGB (R^2 : 0.30–0.41). In contrast, discrete-return height metrics and their derived structural parameters showed better AGB prediction capability, with $H75_{DR}$ achieving the highest R^2 (0.702) and the lowest RMSE (14.711 Mg/ha). Canopy cover estimated from full-waveform data exhibited a moderate correlation with AGB ($R^2 = 0.635$, RMSE = 16.322 Mg/ha), while its structural parameters were strongly correlated with AGB, particularly $H75_{FW} * CC_{FW}$ ($R^2 = 0.760$, RMSE = 13.205 Mg/ha). Further analysis revealed that the synergistic use of discrete-return and full-waveform metrics provided superior AGB estimation: $H75_{DR} * CC_{FW}$ achieved a correlation coefficient of 0.815 and an RMSE of 11.985 Mg/ha, outperforming any individual metric from a single data source.

To mitigate the impact of terrain slope on the estimation of AGB and improve the retrieval accuracy in mountainous forest areas, Wang et al. proposed a slope-adaptive waveform metric calculation method, referred to as PIPE (Plane Inclined and Points of signal Ending)¹¹⁸. The method begins by simulating the echo from bare ground with a known slope, which is aligned with the signal end point of the measured forest waveform to establish a reference. Energy

quartile points of both the forest waveform and the simulated bare-ground echo, denoted as HT25, HT50, HT75, and HT100 for the forest, and Hg25, Hg50, Hg75, and Hg100 for the bare ground, are then extracted. The terrain-adaptive waveform metric RHT is derived by $RHT = HT - Hg$. In parallel, Gaussian decomposition method (GD, RH25, RH50, RH75, and RH100) and waveform parameter method (wfExt, trail10, and lead10) were employed to develop linear regression models for AGB estimation. Model performance was evaluated using simulated LiDAR waveforms from temperate mixed forests and GLAS data from the Xishuangbanna region (tropical seasonal rainforest and tropical mountain rainforest). The results indicated that the estimation accuracy of both the GD and WP methods was sensitive to slope variations, particularly under larger footprint sizes: with a 25 m footprint, for slope ranges of 0° – 20° and 20° – 40° , the GD method yielded R^2 of 0.83 and 0.73, with RMSEs of 23.63 Mg/ha and 30.45 Mg/ha, respectively. The WP method achieved R^2 of 0.85 and 0.76, with RMSEs of 22.48 Mg/ha and 28.64 Mg/ha. With a 70 m footprint, the GD method's R^2 decreased to 0.76 and 0.26, with RMSEs rising to 28.55 Mg/ha and 50.07 Mg/ha, while the WP method yielded values of 0.78 and 0.42, with RMSEs of 27.34 Mg/ha and 44.07 Mg/ha. In contrast, the PIPE method effectively reduced the influence of both terrain slope and footprint size on AGB retrieval accuracy: with a 70 m footprint, it achieved R^2 of 0.93 across both the 0° – 20° and 20° – 40° slope ranges, with RMSEs of 15.13 Mg/ha and 15.31 Mg/ha, respectively. And across the full slope range of 0° – 40° , the R^2 were 0.88 and 0.93 for the 25 m and 70 m footprints, with corresponding RMSEs of 19.79 Mg/ha and 15.38 Mg/ha.

Furthermore, Ni et al. introduced a significant refinement to the waveform energy metrics by extending the original energy quartile points to decile points (10%, 20%, ..., 100%), allowing for a more granular characterization of the waveform's energy structure¹³⁸. Recognizing the potential uncertainties introduced by geolocation inaccuracies and the determination of waveform signal start and end points, the study integrated these factors into its methodology and provided revised definitions accordingly

$$\begin{cases} RHT_i = HT_i - Hg_i = E_i - E_e - Hg_i \\ RH = E_i - E_l \\ extent = E_{0.9} - E_e \\ trail10 = E_{0.1} - E_e \quad lead10 = E_s - E_{0.9} \end{cases} \quad (34)$$

where E_s and E_e represent the elevations of the signal starting and ending points, respectively, and E_l is the elevation of the lowest mode provided in the GEDI L2A products. Analysis of GEDI data from the Temperate Needleleaf Forest in the Cibola National Forest on the Sandia Mountains showed that after geolocation correction, the R^2 (RMSE) values for the GD, PIPE, and WP methods improved from 0.48 (19.7 Mg/ha), 0.56 (18.1 Mg/ha), and 0.44 (19.7 Mg/ha) to 0.69 (14.3 Mg/ha), 0.78 (12.1 Mg/ha), and 0.65 (15.5 Mg/ha), respectively. Differences in AGB estimation were observed

due to varying signal start and end point definitions: Group a4, which used the maximum threshold for identifying signal endpoints, resulted in the shortest effective waveform and the poorest accuracy for the GD and PIPE metrics, with R^2 (RMSE) values of 0.68 (15.4 Mg/ha) and 0.73 (13.8 Mg/ha) under the mean shift of optimized geolocations. In contrast, Group a5, which used the minimum threshold, resulted in the longest effective waveform and the poorest accuracy for the WP metric, with an R^2 (RMSE) of 0.61 (16.5 Mg/ha) under the mean shift of optimized geolocations. After correcting the geolocation offset, using SRTM terrain slope data, all three models (GD, PIPE, and WP) performed well in estimating forest AGB in areas with slopes $\leq 15^\circ$, with R^2 values of 0.82, 0.83, and 0.77, and RMSE values of 11.4, 11.1, and 12.6 Mg/ha, respectively. In areas with slopes $> 15^\circ$, the PIPE method demonstrated a clear advantage. When ALS terrain slope data was used, the R^2 of the PIPE model could be further improved to 0.74 (RMSE = 13.4 Mg/ha). For all 744 waveforms, the PIPE model achieved the best accuracy when using ALS terrain slope data ($R^2 = 0.8$, RMSE = 11.7 Mg/ha). These results demonstrate the significant potential of slope-adaptive waveform metrics for estimating AGB in mountainous forests.

To overcome the regional limitations of the empirical relationships between AGB and full-waveform LiDAR height metrics, Ni-Meister et al. started with individual tree biomass, aggregated the biomass of all trees within a laser footprint, and developed an estimation model for AGB based on large-footprint lidar waveform data¹³⁹

$$\begin{cases} AGBD = k_{WF} BI_{WF} \\ k_{WF} = \frac{\pi}{40} d \frac{b^{-\frac{2}{p}} F \rho}{a P_{gap} G y F_a} \\ BI_{WF} = \int_{z_0}^{z_n} \frac{dP(0, z)}{dz} z \left(1 + \frac{2}{p}\right)^{-\infty} dz \end{cases}, \quad (35)$$

where BI_{WF} refers to the waveform-based biomass index, and k_{WF} constitutes the ratio between the foliage profile-based coefficient and the factor of averaged canopy gap probability. The definitive descriptions of all parameters incorporated are available in the cited Reference. Collectively, Eq.(34) constructs a geographical-context-independent model for estimating AGB, which is dependent solely on attributes perceived from the return waveform. This model is derived from an allometric equation for diameter at breast height, crown volume, and tree height, as weighted by the foliage profile, and is initiated from waveform-perceived tree height and crown distribution characteristics. For model validation, an analysis was performed utilizing waveforms from temperate deciduous/coniferous forests in the northeastern United States and montane coniferous forests in the Sierra National Forest of California. The performance of the proposed model was compared with that of linear regression models for AGB developed using three distinct

height metrics: RH75, RH50, and the combined set (RH100, RH75, RH50). In the Northeastern study area, the performance of the proposed AGBD algorithm was comparable to that of the models based on the three RH metrics, yielding an R^2 of 0.81 and an RMSE of 26.86 Mg/ha, against R^2 of 0.82–0.84 and RMSEs of 24.88–26.45 Mg/ha for the benchmark models. Within the California region and the combined CA + NE dataset, the proposed AGBD algorithm demonstrated superior performance, with R^2 of 0.88–0.90 and RMSEs of 66.41–107.79 Mg/ha. This markedly exceeded the results obtained from the height-index-based AGBD models, which produced R^2 of 0.55–0.74 and RMSEs of 107.54–176.91 Mg/ha. Consequently, the proposed model effectively resolves the issue of underestimation inherent in the application of height metrics within high-biomass regions.

While investigating the influence of forest structural parameters and topographic parameters on the accuracy of RH98-derived canopy height in artificial coniferous forests in Aichi Prefecture, Japan, Li et al. also examined the impact of these parameters on the accuracy of the GEDI AGBD product³¹. They found that the RMSE, ME, and MAE of AGBD initially increased and then decreased with increasing terrain slope, tree count, and height-to-diameter (HD) ratios. As the crown length ratio (CLR) increased, the RMSE, ME, and MAE first decreased and then increased, reaching 124.33 Mg/ha, 54.13 Mg/ha, and 99.43 Mg/ha in the 30%–40% CLR range. Furthermore, Li et al. employed the recommended EBT_Coarse and DBT_Coarse models, using RH60 and RH98 as variables to compute AGBD. The results showed that for evergreen broadleaf forests, evergreen needleleaf forests, and deciduous broadleaf forests, the GEDI AGBD product demonstrated only moderate accuracy, indicating a pressing need to develop a region-specific GEDI AGBD model for Japan.

3.4.6 AI-Assisted footprint-level target retrieval from waveform features

Full-waveform LiDAR does not directly retrieve target characteristics; instead, it establishes relational models between waveform features and parameters such as surface reflectance, vegetation indices, LAI, fractional cover, tree height, and canopy height. These models are then used to develop estimation models for parameters like AGB, thereby obtaining rich ecological information of the earth. As illustrated earlier in this paper and shown in Table S7, the relational models linking waveform features to ecological parameters are highly complex. Their accuracy and adaptability are jointly influenced by detection conditions, terrain conditions, and target characteristics, posing significant challenges to the precise acquisition of target features in surface scenes.

Similar to waveform decomposition and the subsequent extraction of target features, artificial intelligence technolo-

gies in the field of target feature retrieval from waveform characteristics avoid the need for constructing explicit physical models and enable the simultaneous retrieval of multiple ecological parameters of targets. Kellner suggested that machine learning could potentially identify or construct features that exhibit a linear relationship with footprint-level AGBD¹³⁶. If successful, this would facilitate the application of machine learning within a hybrid inference framework. Furthermore, it could simplify or reduce the number of GEDI04_A models used for global prediction, or lead to an adjustment in the stratification strategy of GEDI04_A.

To quantify fuel loads over extensive areas, support integrated fire management plans in fire-prone regions, and protect carbon stocks, biodiversity, and ecosystem functions, Leite et al. employed a RF algorithm to develop machine learning models for estimating fuel loads in Brazil's Cerrado biome (grassland, savanna, and forest vegetation types, etc.)⁵⁵. Using GEDI waveform metrics simulated from UAV LiDAR 3D point clouds as input data, namely RH98, plant area index (PAI), canopy cover fraction, and foliage height diversity (FHD), the study assessed the capability of GEDI data for large-scale multi-layer fuel load estimation in a tropical savanna ecosystem. For surface fuels, herbaceous fuels, combined surface and herbaceous fuels, shrub fuels, tree fuels, and total fuel loads, the model performed best in predicting woody fuels and total fuel load ($R^2 = 0.88$ and 0.71 , respectively), and showed low to moderate accuracy (R^2 ranging from 0.15 to 0.46) but remained reliable for the lower fuel strata. This confirms the potential advantage of using GEDI data to improve the estimation of vegetation properties in structurally complex ecosystems.

To extend the application of GEDI for estimating AGBD in African savannas, Li et al. analyzed multiple sites with terrain slopes of less than 5.2° , encompassing variations in annual rainfall, vegetation type, height, and canopy cover¹⁴⁰. They developed generalized linear models (GLM) using individual or combined GEDI waveform metrics, including RH50, RH75, RH98, canopy cover, FHD, and PAI, to estimate AGBD. A comparison of R^2 values indicated that the GLM incorporating RH98, RH50, and FHD demonstrated the best performance. A further comparison was conducted among the GEDI AGBD product, the GLM estimation model, and a RF model (using RH50, RH75, RH90, RH98, sensitivity, FHD, and canopy cover as inputs). The results showed that both the GLM and RF models significantly outperformed the GEDI AGBD product. The RF model achieved the highest R^2 (0.71) and the lowest %RMSE (53.3%). In contrast, the GLM had the lowest relative mean systematic deviation (RMSD) of 9.2% , which was only half that of the RF model across all biomass intervals. The GEDI product exhibited significant systematic underestimation, with an RMSD as high as -37% . These findings underscore the importance and advantage of local biomass model calibration for unlocking the full potential of GEDI metrics in AGBD estimation.

To quantify the complex relationship between forest canopy height and AGB, Zhang et al. utilized ICESat-2 ATL08 data, fused ATL03/ATL08 data, and the GEDI RH98 to derive canopy height¹⁴¹. They evaluated the most suitable dataset for canopy height estimation using RF and Monte Carlo uncertainty analysis. The accuracy of canopy height features in AGB estimation was compared using both linear regression and RF models, and the contribution of these features across varying biomass gradients was quantified using a semivariogram function. The study area was the San Shan region in Ningxia Province, China, which includes the Helan Mountain, Liupan Mountain, and Luoshan National Nature Reserves. In terms of canopy height estimation accuracy, the R^2 for ATL08, GEDI RH98, and fused ATL03/ATL08 data were 0.28, 0.32, and 0.58, respectively, with RMSE values of 4.76 m, 4.89 m, and 3.44 m. This indicates that the fused ATL03/08 data achieved the highest accuracy, while the ATL08 data performed the worst. Incorporating canopy height features led to a relative improvement of 13.89% and 10.34% in R^2 , and a reduction of 4.07% and 6.62% in RMSE for the linear regression model and RF, respectively.

Although the GEDI AGBD product does not include bamboo forests according to Kellner's summary, bamboo, characterized by rapid growth and significant carbon sequestration potential, has become a critical research focus in forest remote sensing for biomass estimation. To address this, Yang et al. employed kriging interpolation to convert GEDI footprint variables into raster data to estimate the AGB of Dragon Bamboo in Xiping County, Yunnan, China¹⁴². Based on RF feature importance ranking, five key variables were selected from 40 GEDI features. These variables served as input for three artificial intelligence algorithms: RF regression, Support Vector Regression, and a stacking ensemble learning algorithm with K-Nearest Neighbors as the meta-learner. The results demonstrated that the Stacking model performed best, achieving an R^2 of 0.93, and RMSE of 7.04 Mg/ha, outperforming both the single RF regression ($R^2 = 0.86$, RMSE = 8.21 Mg/ha) and Support Vector Regression models ($R^2 = 0.50$, RMSE = 15.07 Mg/ha). This study contributes to a better understanding of the carbon sink potential of bamboo forests.

To reveal the spatial distribution characteristics of canopy cover at a regional scale, Huang et al. employed three types of datasets, namely active remote sensing, passive remote sensing, and multiple remote sensing¹⁴³. They selected the optimal artificial intelligence model for forest canopy cover estimation and evaluated the applicability of their proposed method across different land cover types and slope aspects. Among the three datasets (SAR, Landsat, and Multiple Remote Sensing), the maximum R^2 values of the spatial distribution of canopy cover from five AI models, Bootstrap Aggregating (Bragging), RF, Multiple Linear Regression (MLR), Long Short-Term Memory (LSTM), and Support Vector Regression (SVR) were 0.67, 0.61, and 0.43, respec-

tively. The average R^2 values were 0.636, 0.546, and 0.406, respectively. Within the Multiple Remote Sensing dataset, the R^2 values for the five models were 0.65, 0.67, 0.64, 0.58, and 0.64, with corresponding RMSE values of 0.11, 0.10, 0.12, 0.14, and 0.11. Thus, compared to active and passive remote sensing, the multiple remote sensing dataset significantly improved the inversion accuracy of canopy cover spatial distribution, with the RF model achieving the highest accuracy among all models. Under the RF model, the average values of RMSE, MAE, and STD for rainfed cropland, herbaceous cover, irrigated cropland, grassland, and closed deciduous broadleaved forest were 0.10, 0.09, and 0.10, respectively, indicating a certain influence of forest type on the estimation results. For the eight slope aspects considered, the average values of RMSE, MAE, and STD were 0.12, 0.09, and 0.12, respectively, suggesting that the influence of different aspects on the estimation accuracy of canopy cover was relatively small.

Accurately characterizing post-fire vegetation recovery trajectories and distinguishing between naturally regenerating areas and those subject to sustained human disturbance are crucial for deepening our understanding of post-disturbance forest dynamics, the drivers of variation, and the potential impacts of global change. To this end, Wu et al. integrated data from GEDI, Landsat, Sentinel-2, Shuttle SRTM-3, TerraClimate, Gridded Population of the World (GPW), and Tropical Moist Forest¹⁴⁰. They developed a machine learning model based on eXtreme Gradient Boosting (XGBoost). Focusing on protected areas in the eastern Amazon, the study analyzed the rates of greenness recovery and canopy structure recovery across all burned areas, naturally regenerated areas, and permanently converted areas.

The results showed an overall model classification accuracy of 87.90% in distinguish natural recovery areas from permanently converted areas. Within five years post-fire, most burned areas exhibited greenness recovery. Specifically, recovery was relatively higher in natural recovery areas than in permanently converted areas. Statistically significant increases in greenness recovery rates were observed for all burned areas and natural recovery areas, whereas the trend in permanently converted areas was not significant. In natural recovery areas, the corresponding absolute recovery ratios for key canopy structural parameters were as follows: RH98 (0.06 m), RH75 (0.13 m), RH50 (0.24 m), RH25 (0.34 m), PAI0-5 (0.36), PAI5-10 (0.39), PAI10-15 (0.38), PAI15-20 (0.18), PAI20-25 (0.22), PAI25-30 (0), canopy ratio (0.07), canopy cover (0.19), and FHD (0). The study revealed significant recovery of structural traits, such as RH, canopy cover, and PAI in natural recovery areas, reaching pre-disturbance levels within 20 years. However, among different forest structural traits, those related to the vertical allocation of plant material (e.g., PAI per 5-meter height layer) exhibited stronger recovery ratios compared to traits associated with canopy height (e.g., RH metrics). Furthermore, differential recovery rates were observed among RH

metrics, with the lower canopy layer (e.g., RH25, recovery ratio 0.34) recovering more substantially than the upper canopy (e.g., RH98, recovery ratio 0.06).

3.4.7 Mapping forest canopy height using waveform metrics

Utilizing sophisticated waveform analysis, full-waveform LiDAR can retrieve rich ecological information ranging from canopy height to parameters like fire recovery ratio. However, its practical utility is constrained by observational scale and spatial coverage. Footprint sizes range from sub-meter to meter scale for airborne systems to decameter scale for spaceborne systems (e.g., 70 m for GLAS and 25 m for GEDI). Notably, while GEDI achieves along-track sampling, there are gaps of approximately 600 meters between its multiple observation tracks, making it essentially a discrete strip-sampling system along the ISS (International Space Station) orbit. This means that no matter how precise the footprint-level measurements are, their direct value for practical applications like regional management or global carbon accounting is limited.

Consequently, extrapolating the precise information from discrete footprints into seamless, spatially continuous distributions has become key to unlocking its core value. Table S8 extensively showed the mapping methods and accuracy for regional and global ecological parameters. In summary, the current research frontier focuses on developing multi-source remote sensing data fusion strategies to fill the information gaps between GEDI strips and ultimately generate continuous ecological parameter products at regional to global scales. Dubayah et al. proposes a two-level hierarchical modeling framework: the first level establishes relationships between ground measurements and LiDAR footprint metrics, and the second level links these footprint metrics to wall-to-wall auxiliary data for scaling up¹⁴⁵. Qi et al. demonstrated the potential of such fusion, showing that by synergistically using Tandem-X data, GEDI data, and topographic data, they achieved forest height mapping with superior accuracy and spatial completeness compared to any single data source in the Hubbard Brook Experimental Forest¹⁴⁶. Such efforts herald that the coordinated use of a new generation of active remote sensing missions, including GEDI, ICESat-2, NISAR, and BIOMASS, is propelling ecosystem structure remote sensing into a new era of multi-platform collaboration and information fusion.

In the context of forest height mapping using GEDI data, leveraging the strong correlation between GLAS RH100 values and globally distributed auxiliary variables with higher spatial density, Simard et al. established a global RH100 model based on forest type, tree cover, elevation, and climatological data¹²³. Using GLAS data from 2005, they generated a global forest canopy height map with 1 km spatial resolution. The model achieved an R^2 of 0.84 and an RMSE of 4.07 m between modeled and measured RH100,

and an R^2 of 0.69 with an RMSE of 4.4 m when compared to measured canopy height. Model errors were associated with forest type, showing larger deviations in dense broadleaf forests such as the Amazon, where errors increased with canopy height. Lang et al. employed a probabilistic deep learning method based on ensemble learning of deep convolutional neural networks¹⁴⁷. Their approach formulates the regression of GEDI waveform canopy height (represented by RH98) as an end-to-end supervised learning task. Leveraging the ability of CNN models to learn and extract robust features with strong generalization capability, the method effectively extends results from GEDI-surveyed regions to new geographic areas not covered by GEDI. The RMSE is 2.7 m, and the mean error bias is -0.1 m. This represents a clear advancement over the initial GEDI L2A height estimates, which had an RMSE of 3.4 m and a mean error bias of -0.9 m.

Although forest canopy height can be mapped using GEDI data, it is common practice to integrate GEDI with other data sources. In terms of fusing GEDI with optical imagery, Potapov et al. developed a method to spatially extrapolate GEDI RH95-derived forest canopy height using Landsat data, successfully producing a 30 m resolution global forest canopy height map for 2019¹³. The map was generated based on footprint-level forest height data from regions including the USA, Mexico, the Democratic Republic of the Congo, and Australia. Accuracy assessment showed that the global height map achieved an RMSE of 6.6 m and an R^2 of 0.62 when compared with independent GEDI validation data. Relative to airborne lidar data, it yielded an RMSE of 9.07 m and an R^2 of 0.61. In terms of classification accuracy, for forest canopy delineation at height thresholds above 3 m, 5 m, and 10 m, the overall accuracies relative to RH95 were 83.0%, 87.8%, and 92.9%, respectively. When validated against airborne lidar forest height, the corresponding accuracies were 87.6%, 88.1%, and 90.1%.

Also in the area of fusion GEDI with optical imagery, Lang et al. proposed a probabilistic deep learning model that integrates Sentinel-2 imagery with sparse, footprint-level GEDI RH98 forest canopy height data¹⁴⁸. This approach enables the direct retrieval of top-of-canopy height from Sentinel-2 images over any region globally, significantly enhancing the spatial and temporal resolution of canopy height modeling. The study successfully produced the world's first spatially continuous, wall-to-wall top-of-canopy height map at a 10 m ground sampling distance. Compared to GEDI RH98, the model-derived heights achieved an RMSE of 6.0 m, an MAE of 4.0 m, and an ME of 1.3 m. RMSE varied between 3.1 m and 9.2 m for fourteen vegetation types, tropical moist broadleaf forests, tropical dry broadleaf forests, tropical coniferous forests, temperate broadleaf forests, temperate coniferous forests, boreal forests, tropical grasslands, temperate grasslands, flooded grasslands, montane grasslands, tundra, Mediterranean

forests, deserts, mangroves. Analysis based on this global canopy height map revealed that only 5% of the global land area is covered by trees taller than 30 m, and merely 34% of these tall forests are located within protected areas.

Liu et al. integrated GEDI RH100, ICESat-2 RH98, and Sentinel-2 imagery to develop a novel Neural network guided interpolation (NNGI) method, which was used to generate a 30 m resolution forest canopy height product for China in 2019¹¹⁹. The product was validated against three independent datasets: 1.1 million GEDI footprints, approximately 33 km² of UAV LiDAR data, and over 59 000 field plot measurements. The validation yielded R^2 of 0.55, 0.58, and 0.60, with corresponding RMSE values of 5.32 m, 4.93 m, and 4.88 m, respectively. This study demonstrates the feasibility of the NNGI method for spatially continuous forest canopy height monitoring from national to global scales by integrating multi-platform spaceborne LiDAR data and optical imagery. It provides a new technical pathway for more accurately quantifying terrestrial carbon stocks and deepening the understanding of forest ecosystem processes.

Also based on the fusion of GEDI, ICESat-2, and Sentinel data, and aiming to mitigate the adverse effects of terrain relief and vegetation cover on canopy height retrieval accuracy while overcoming the spatially heterogeneous uncertainty in forest canopy height mapping, Wang et al. proposed a novel hybrid modeling framework¹⁴⁹. Their approach integrated a suite of GEDI RH metrics, ICESat-2 canopy height metrics, and Sentinel-2 vegetation indices as input data. The study area was partitioned into multiple fusion patterns according to the spatial distribution of footprints, and corresponding canopy height models were developed for each pattern. A geospatial distance-weighted scheme was finally introduced to ensemble the outputs from these models. Validation was conducted across three representative sites: the Smithsonian Environmental Research Center (SERC, deciduous broadleaf forest), Lantau Island (evergreen broadleaf forest and savanna), and Cape Breton Highlands National Park (CBHNP, coniferous forest). For three forest types, the hybrid model achieved R^2 values of 0.85, 0.65, 0.65, and 0.90, with RMSEs of 4.15 m, 1.02 m, 1.47 m, and 0.57 m, respectively. A comparative analysis with existing products revealed that in the SERC site, the R^2 for Potapov's 30 m product, Lang's 10 m product, and the hybrid model were 0.67, 0.68, and 0.85, with RMSEs of 6.20 m, 12.39 m, and 4.15 m. In Lantau Island, the corresponding R^2 values were 0.02, 0.04, and 0.66, with RMSEs of 7.63 m, 13.10 m, and 1.22 m. In the CBHNP site, the R^2 values were 0.66, 0.55, and 0.90, with RMSEs of 3.53 m, 5.98 m, and 0.57 m. As a result, the hybrid model fully leverages the complementary strengths of multi-source remote sensing data.

To compare the performance of ICESat-2 and GEDI technologies along with various artificial intelligence algorithms in forest canopy height mapping, and to address the unique challenges posed by dense vegetation and complex structure

in tropical forest height retrieval, Liu et al. derived footprint-level canopy height using GEDI RH98 and ICESat-2 RH98¹⁵⁰. Combined with features extracted from Landsat-8, PALSAR-2, and FABDEM products, they employed the AutoGluon stacked ensemble learning algorithm to develop inversion models, generating continuous 30 m resolution canopy height maps for tropical forests in Puerto Rico under GEDI and ICESat-2 frameworks. For the ICESat-2 height map, eight machine learning methods, random forest, extra trees, neural network, XGBoost, LightGBM, CatBoost, K-Nearest neighbors, and an ensemble model, achieved R^2 values changes between 0.49 and 0.69, with RMSE values between 3.95 m and 5.24 m. For the GEDI height map, the corresponding eight models yielded R^2 values between 0.65 and 0.80, with RMSE values between 4.81 m and 5.39 m. The results indicate that the GEDI height map slightly outperforms the ICESat-2 map in overall accuracy. Among the eight machine learning methods, the ensemble model achieved marginally better accuracy than the others, and also slightly surpassed the performance of the method previously proposed by Liu et al.¹²⁷.

Liu et al. combined GEDI and SRTM data to uncover the influence of environmental drivers on canopy height¹⁵¹. Using over ten environmental variables, including GEDI RH98 and ground elevation, as inputs, their study applied a RF model to generate a spatial distribution map of canopy height across global tropical forests. The results showed a correlation coefficient of 0.75 between the RH98-derived height and the global canopy height, with an RMSE of 3.2 m. The feature importance of environmental variables in determining tropical forest canopy height at both local and regional scales was ranked in descending order as follows: ground elevation, dry season length, incoming solar radiation, temperature seasonality, wind speed, cloud frequency, soil sand percentage, precipitation, soil organic matter, and lightning strikes. Together, climate, topography, and soil properties explain approximately 75% of the spatial variation in tropical forest canopy height.

To enhance the spatial resolution of large-area forest height maps and to examine the effects of different data fusion strategies on forest canopy height retrieval, Qi et al. combined GEDI data with TanDEM-X InSAR data¹⁵². The approach utilized the vertical profile of reflectance distribution from GEDI waveforms to invert a physically based InSAR model for estimating canopy height. Subsequently, GEDI-derived canopy height and an adaptive wavenumber-based calibration model were applied to reduce errors in the InSAR-inverted height. This process yielded continuous forest height products at 25 m and 100 m resolutions for Gabon, Mexico, French Guiana, and the Amazon Basin. Accuracy assessment showed that at 25 m resolution, the R^2 values between forest canopy height from GEDI + Sentinel, GEDI + Landsat, and GEDI + TanDEM-X combinations and LVIS reference data were 0.49, 0.39, and 0.47, with corresponding RMSEs of 8.31 m, 9.29 m, and 8.48 m. At

100 m resolution, the R^2 values were 0.54, 0.44, and 0.57, with RMSEs of 7.36 m, 7.86 m, and 6.91 m. The results indicate that at 25 m resolution, GEDI + Sentinel performed similarly to GEDI + TanDEM-X, while at 100 m resolution, GEDI + TanDEM-X achieved higher accuracy. This fusion method represents the first large-scale, high-precision forest canopy height mapping based on the integration of GEDI and TanDEM-X InSAR data, laying an important foundation for AGB mapping across pantropical regions.

The aforementioned methods typically required downloading massive volumes of imagery such as Landsat and Sentinel data, and established relationships between only a single GEDI RH metric (e.g., RH95 or RH98) and surface reflectance or backscatter signals, thus overlooking the rich ecological information contained within multiple RH metrics. To address this, Xiao et al. adopted a series of RH metrics (RH0-RH100 at 10 unit intervals) and, leveraging Google Earth Engine and deep learning techniques, constructed a feature fusion model¹⁵³. This model uses a convolutional neural network module to extract features from spatially continuous remote sensing data including Landsat-8, Sentinel-1 imagery, and topographic data, while a multilayer perceptron module is designed specifically to process interpolated GEDI multi-RH metrics. Through feature fusion, accurate canopy height inversion was achieved, resulting in a 30 m spatial resolution canopy height product for China in 2020 (MRH-CHM). Within the height range of 5 to 50 meters (at 5 m intervals), the RMSE of this product was lower than that of existing forest canopy height products from Lang, Potapov, and Liu across all intervals. When validated against 6 168 244 GEDI height points including low vegetation (heights below 3 m), the R^2 values for MRH-CHM and Lang's product were 0.78 and 0.69, and with RMSEs of 5.74 m and 7.09 m, respectively. When validated against approximately 1 340 962 GEDI points with heights ≥ 3 m, the R^2 values for MRH-CHM, Lang's, Potapov's, and Liu's products were 0.44, 0.36, 0.32, and 0.18, and with RMSEs of 8.02 m, 8.47 m, 9.75 m, and 9.87 m, respectively. The results demonstrate that MRH-CHM achieves higher accuracy compared to existing mainstream products across multiple validation scenarios. This outcome provides forestry decision-makers, researchers, and stakeholders with a more reliable basis for monitoring forest health, assessing carbon stocks, and analyzing climate change.

With the release of multiple regional and global forest canopy height products, independent and systematic validation of their accuracy is crucial. To this end, using the 98th percentile height derived from airborne Riegl VQ-1560i-DW LiDAR data as reference, Li et al. systematically evaluated the impact of topography, slope, and vegetation characteristics on the accuracy of several forest canopy height products: the Forest Canopy Height map for China (FCHChina), the 2020 Global Forest Canopy Height map (GFCH2020), and the 2019 Global Forest Canopy Height

map (GFCH2019)¹⁵⁴. The accuracy of different methods across different topographical and vegetation characteristics were showed in Table S9. For these wall-to-wall products, the main conclusions are: I) Topographic effect: The accuracy of all three products was higher in hilly areas than in plains and mountainous regions. II) Inter-product bias patterns: GFCH2019 showed the highest overall accuracy but exhibited systematic underestimation in plains and hilly areas, whereas FCHChina and GFCH2020 showed significant overestimation in mountainous regions. III) Influence of slope and canopy cover: Product accuracy was significantly affected by slope and canopy cover, increasing slope led to larger errors, and areas with low canopy cover were prone to overestimation. IV) Forest type dependency: Inversion accuracy varied across forest types, with the highest accuracy for Moso bamboo forests and the lowest for Masson pine forests. This discrepancy is likely attributable to the distinct structural attributes of each forest stand.

3.4.8 Mapping AGB Using Waveform Metrics

In the field of mapping AGB using GEDI data, Huang et al. applied Kriging interpolation to extend discrete footprint data into a continuous surface¹⁵⁵. Based on the GEDI AGB product for Liaoning Province, they generated a provincial-scale biomass distribution map and systematically evaluated three existing forest AGB maps of China. The study employed direct comparison as well as two proposed refined methods, a1 and a2, to compare with the Yang, Su, and Zhang maps. The results showed Pearson R^2 for the Yang map of 0.46, 0.61, and 0.70; for the Su map of 0.61, 0.75, and 0.80; and for the Zhang map of 0.35, 0.46, and 0.69, indicating that the proposed mathematical methods significantly improved comparison accuracy. Furthermore, validation using field measurement data from the Mudanjiang Ecological Research Station yielded direct correlation values of 0.49, 0.49, and 0.39 with the three maps, respectively. After applying the a2 method, these were enhanced to 0.66, 0.65, and 0.50. Among different forest types, evergreen needleleaf forests and deciduous needleleaf forests performed similarly and both outperformed deciduous broadleaf forests, confirming that validation by forest type is more accurate than an overall assessment. Analysis across 14 prefecture-level cities in Liaoning Province revealed that cities in the forest-rich eastern region had higher R^2 with the three maps, whereas those in the forest-sparse western region showed lower R^2 . By employing Kriging interpolation, the study achieved the extension from point-based biomass information to spatially continuous regional estimation, offering a new approach to address the spatial coverage gaps in LiDAR-derived biomass data.

May et al. proposed a hierarchical full-model framework for estimating AGBD¹⁵⁶. This method integrates three nested submodels: I) Biomass model: this model establishes a linear regression relationship between simulated waveform RH95

and AGBD, II) Measurement error model: by selecting ground-detection algorithms suitable for local terrain, this model characterizes the systematic differences and error propagation between simulated waveform RH95 and GEDI measured waveform RH98, and III) Geostatistical model: this model describes the spatial distribution of RH95 provided by spaceborne GEDI observations, capturing its spatial correlation and variation. Three submodels are integrated into a unified hierarchical model, establishing a robust statistical link between the spatial pattern of GEDI observations and field-measured AGBD. Validation in Jambi Province, Sumatra, and Indonesia showed that the mean AGBD for forest, oil palm, rubber plantations, and shrubland was 199.7 Mg/ha, 106.2 Mg/ha, 121.7 Mg/ha, and 61.2 Mg/ha, respectively, with corresponding mean residuals of 0.1 Mg/ha, 23.6 Mg/ha, -5.1 Mg/ha, and -34.7 Mg/ha. Model uncertainty remained relatively low compared to predicted biomass: the median relative standard deviation was 8% at 1 km resolution and 26% at 100 m resolution.

Using the general biomass model represented by the Eq. (35), Ni-Meister et al. integrated the full vertical profile information captured by LiDAR waveforms with the fundamental allometric relationships of trees¹⁵⁷. This created a biophysically meaningful link between remotely sensed forest structure and AGB, forming a universal framework that relies solely on large-footprint, full-waveform LiDAR and is applicable across diverse forest ecosystems. The framework enabled regional-scale AGB mapping based on LVIS (airborne LiDAR) and national-scale mapping based on spaceborne GEDI. Validation across four study areas, Lope, Mondah, Mabnie, and Rabi, demonstrated that the model achieved consistent accuracy at the hectare scale compared to field measurements ($R^2 = 0.91$, RMSE = 40.38 Mg/ha). Under the same conditions, the Armston AGBD model yielded an R^2 of 0.77 and an RMSE of 77.01 Mg/ha, while the Saatchi AGBD model produced an R^2 of 0.82 and an RMSE of 54.58 Mg/ha.

Similar to forest canopy height mapping, AGB mapping is typically achieved by integrating GEDI with other remote sensing technologies. However, integrating remote sensing technologies with different spatial resolutions and varied strengths and limitations in ecological parameter retrieval requires addressing the challenge of consistently comparing or fusing estimation results from new datasets. To this end, Duncanson conducted a systematic comparative study of simulated biomass estimates from three NASA missions, GEDI, ICESat-2, and NISAR, in Sonoma County, California, USA²⁵. Using the RH-AGBD estimation model recommended by GEDI, the R^2 values for deciduous broadleaf trees, evergreen needleleaf trees, and woodland/grassland/shrub were 0.64, 0.58, and 0.61, with %RMSE values of 26.77%, 31.48%, and 34.87%, respectively. Under the same estimation model, ICESat-2 yielded R^2 values of 0.65, 0.60, and 0.69 for the corresponding vegetation types, with %RMSE values of 27.94%, 27.02%, and 29.22%. When comparing the simu-

lated results against local reference maps at each mission's native resolution (25 m, 100 m swath, and 1 ha), the RMSE values for GEDI, NISAR, and ICESat-2 were 80.2 Mg/ha, 151.55 Mg/ha, and 137.8 Mg/ha, with %RMSE values of 57%, 75%, and 89%, respectively. This work provides methodological guidance for cross-validating spaceborne biomass estimates using airborne lidar reference maps, and offers an expected accuracy reference for end-users to evaluate whether to adopt a multi-sensor mapping approach combining GEDI, NISAR, and ICESat-2.

Also using the RH-AGBD model recommended by GEDI, Qi et al. estimated footprint-level AGBD in temperate mixed deciduous forest, temperate montane coniferous forest, and tropical forest¹⁵⁸. Simultaneously, canopy height was retrieved from TanDEM-X data using the Random Volume over Ground model. By linking the derived TDX InSAR heights to the footprint-level GEDI biomass, calibration relationships were developed to calculate a 1-km resolution GEDI/TDX fused AGBD product. In the HBEF, TEF, and LSBS sites, the R^2 between GEDI AGBD and ALS AGBD were 0.90, 0.58, and 0.75, with RMSEs of 19.2 Mg/ha, 72.9 Mg/ha, and 32.0 Mg/ha, respectively. In comparison, the GEDI/TDX fused biomass showed R^2 of 0.90, 0.88, and 0.82 with the reference biomass, with RMSEs of 22.6 Mg/ha, 46.2 Mg/ha, and 25.1 Mg/ha. The study confirms that combining biomass-sensitive data such as TDX-derived heights (or other auxiliary datasets) with GEDI observations of ecosystem structure holds promise for simultaneously improving accuracy and/or delivering results at a higher spatial resolution than possible with a single data source.

Shendryk et al. employed a gradient boosting machine learning framework to generate 100 m and 200 m resolution AGB maps for Australia and the United States in 2020¹⁵⁹. The model integrated GEDI L4A AGBD data with Sentinel-1 SAR imagery, Sentinel-2 multispectral imagery, elevation data, and land cover data. The accuracy of the model showed a strong dependence on land cover types: across various scenarios, R^2 values ranged from 0.66 to 0.77 for trees, 0.34 to 0.52 for shrublands, 0.39 to 0.57 for grasslands, and 0.14 to 0.68 for croplands. Consequently, targeted calibration of GEDI data for different land cover types is needed to enhance the model's generality and accuracy.

To verify the role of GEDI data in mapping AGB and to overcome the significant uncertainties in biomass estimation caused by complex terrain in subalpine forest regions, Nian et al. integrated RIEGL miniVUX-1 and DJI M600 UAV data with GEDI metrics including RH100, canopy cover, FHD, PAI, and gap probability¹⁶⁰. Using a RF algorithm, they first generated discrete footprint-level AGB estimates for the Qilian Mountain National Park in Gansu Province. Subsequently, by incorporating Sentinel-1 and Landsat-8 data and continuing with the RF model, they extrapolated the discrete GEDI footprint biomass to produce a continuous, high-precision forest biomass distribution map covering the entire park, achieving an accuracy of $R^2 =$

0.66 and RMSE = 19.08 Mg/ha. To clarify the specific contribution of GEDI data, Nian et al. compared results obtained using only UAV, Sentinel-1, and Landsat-8 data (without GEDI). That approach yielded AGB estimates with $R^2 = 0.69$ and RMSE = 29.44 Mg/ha, which were less accurate than the scheme integrating GEDI data ($R^2 = 0.75$, RMSE = 28.75 Mg/ha). This comparison confirms that incorporating GEDI data as an intermediate bridge in the prediction process effectively enhanced the accuracy of forest biomass estimation. The study successfully extended biomass estimation from the plot scale to the regional scale, significantly improving the retrieval accuracy of mountain forest biomass through the synergy of multi-source data.

3.4.9 Mapping canopy height and AGB simultaneously from waveform metrics

To quantify the human-induced degradation of tropical moist forests, Bourgoïn et al. integrated long-term, spatially continuous spaceborne optical datasets with GEDI-derived estimates of canopy height and biomass¹⁶¹. They systematically examined the mean values, standard errors, and accuracy of RH98 and AGBD across 12 scenarios encompassing intact forests, degraded forests, forest edges, and forest regrowth in the Americas, Asia, and Africa. This analysis revealed the effects of forest edge influence, selective logging, fire, and forest regrowth on canopy height. The study quantified the extent and persistence of forest structural impacts from multiple degradation types, identifying RH50, RH98, and distance to the forest edge as strong predictors of deforestation probability. Selective logging and fire reduced forest height by approximately 15% and 50%, respectively, with very limited recovery even after 20 years. Forest edge effects driven by agricultural and road expansion led to reductions of 20–30% in canopy height and biomass, with measurable impacts extending up to 1.5 km into the forest interior. These edge effects affect an area equivalent to 18% of the remaining tropical moist forest, which is more than 200% higher than previous estimates. Furthermore, degraded forests with over 50% canopy loss were found to be more vulnerable to subsequent logging. This research clarifies, at a pantropical scale, the actual spatial extent and long-term ecological consequences of tropical forest degradation caused by selective logging, fire, and edge effects.

To estimate forest parameters at large scales and improve mapping performance, Zhang et al. extracted AGBD and RH95 metrics from GEDI data and integrated multi-source earth observation data from Sentinel-1, Sentinel-2, Landsat 8, ALOS-2, and GLO-30, constructing 432 feature variables across 14 categories¹⁶². Using machine learning methods, they established relationships between the GEDI metrics and these feature variables, subsequently generating high-spatial-resolution (10 m) and high-accuracy (500 m) maps of forest canopy height and AGB for the Hainan tropical rainforest. Six machine learning models were tested: deep neural

network, distributed random forest, gradient boosting machine, generalized linear model, stacked ensemble, and extremely randomized trees. The AutoML-based integration strategy ensured that the stacked ensemble model achieved the best performance among all models. At 10 m and 500 m resolutions, the predicted forest height showed R^2 values of 0.53 and 0.64 with GEDI RH50, with RMSEs of 11.4 m and 10.06 m, respectively. The predicted AGBD showed R^2 values of 0.53 and 0.66 with GEDI AGBD, with RMSEs of 133.56 Mg/ha and 114.24 Mg/ha, respectively. Validation in the Hainan Tropical Rainforest National Park showed that, using GEDI AGBD as reference, the R^2 values for this method and those of Santoro, Yang, and Zhang were 0.74, −0.01, −0.02, and 0.05, with RMSEs of 131.39 Mg/ha, 218.19 Mg/ha, 180.75 Mg/ha, and 181.74 Mg/ha, respectively. Using GEDI RH95 as reference, the R^2 values for this method and the maps of Lang, Liu, and Potapov were 0.73, 0.17, 0.01, and 0.06, with RMSEs of 11.40 m, 16.92 m, 21.63 m, and 19.62 m, respectively. Thus, compared to existing products, this method significantly improves parameter estimation accuracy and demonstrates consistent applicability in both biomass and canopy height estimation, providing a new and reliable solution for large-scale forest parameter retrieval.

3.4.10 Annual mapping of AGB from waveform metrics

Potapov noted that the Landsat-based forest canopy height model calibrated with GEDI data can be applied to historical imagery¹³. In fact, this approach can be extended to other ecological parameters as well. It enables the spatial extrapolation of GEDI-derived ecological parameters from discrete footprints to spatially continuous coverage at regional and global scales, while also facilitating a temporal extension from single time points to continuous time series.

Consequently, this framework allows for the retrieval and analysis of multiple ecological parameters across historical periods since the availability of optical imagery. It supports annual monitoring, reporting, and verification, enables accurate quantification of the climate benefits from ecological restoration practices, and helps systematically reveal the spatiotemporal evolution patterns of ecological parameters at regional to global scales. These capabilities make it valuable for applications in carbon market mechanisms, ecosystem management, vegetation dynamics research, and operational monitoring of forest degradation and recovery.

Liang et al. integrated 3D forest structure data from GEDI with Landsat optical time series¹⁶³. Using Landsat spectral indices from the limited hydric stress (LHS) period as predictor variables, they applied three RF models, LHSM_S97, LHSM_S98, and ANM_S98, to annual LHS Landsat composite data stabilized by the LandTrendr algorithm. This approach enabled the generation of annual forest AGBD maps for the Mabalane region in southern Mozambique from 2007 to 2019. The study systematically

reconstructed the interannual dynamics of forest AGBD at an annual scale, quantifying a total AGB loss of 2.12 ± 0.06 million metric tons over the decade-long period. These results provide important data support and scientific insights for understanding the shift in charcoal-driven degradation patterns and their ecological impacts.

Atzberger et al. proposed a regional AGB mapping approach achieving decameter-scale spatial resolution (10–30 m)¹⁶⁴. The method first converts the height difference between GEDI RH95 and RH10 using a scaling factor to generate a footprint-level AGB^*_{RH} with higher accuracy than the GEDI AGB data. This AGB^*_{RH} is then combined with 32 Barlow twin representations derived from Sentinel-2 and Landsat imagery to produce a 10 m resolution AGB map (based on Sentinel-2) for 2017–2024, as well as a 30-m resolution AGBD map for 2013–2019. Validation against 38 field biomass plots covering agroforestry systems (cocoa, oil palm) and forests of varying recovery ages showed that Atzberger's method outperformed other widely used products (GLAD CHM, ETH CHM, ESA CCI AGB, Meta CHM, GTDX AGB). The proposed method achieved an R^2 of ~ 0.66 , compared to ~ 0.41 , ~ 0.41 , ~ 0.18 , ~ 0.13 , and ~ 0.09 for the other products, respectively. Its RMSE remained below 25 Mg/ha, whereas the RMSE values of the other products were approximately 40, 32, 47, 50, and 57 Mg/ha. These results demonstrate that the proposed approach surpasses conventional AGB mapping methods, providing a more consistent and robust time series for biomass monitoring.

To generate annual maps of forest height, timber volume, and AGB at the stand scale, with a focus on forest growth dynamics rather than loss monitoring, Schwartz et al. employed a U-Net deep learning model that integrated multi-source data from Sentinel-1, Sentinel-2, and GEDI¹⁶⁵. They produced a 10 m resolution forest canopy height time series for France over a seven-year period (2018–2024). Based on leaf-types-specific allometric equations, these height maps were further converted into 30 m resolution maps of growing stock volume and AGB. Validation against French National Forest Inventory data showed that the height retrievals had a MAE of 3.07 m ($R^2 = 0.68$) across all years, with volume and biomass errors of 86 m³/ha ($R^2 = 0.41$) and 65.1 Mg/ha ($R^2 = 0.29$), respectively. The model demonstrated a strong ability to capture tree growth: when aggregated to a 5-km pixel scale, growth in the 10–15 m height class during 2020–2022 closely matched LiDAR-derived estimates ($R^2 = 0.60$, MAE = 0.27 m). At the regional scale, the model's annual wood production estimates also aligned well with inventory data (MAE = 1.45 m³ ha⁻¹ year⁻¹, $R^2 = 0.59$).

3.4.11 Mapping building height from waveform metrics

In recent decades, China has undergone an unprecedented rapid urbanization process. Existing research has predominantly focused on the horizontal expansion of cities, while

attention to vertical spatial growth patterns remains relatively limited. Acquiring nationwide 3D urban data not only offers new perspectives for understanding urban development and its impacts on population density, energy consumption, and carbon budgets, but also holds significant value for urban planning, disaster management, and sustainable development.

In the field of building height mapping, Ma et al. integrated GEDI metrics (RH95 and mean RH95) with multi-source remote sensing data including Landsat-8, Sentinel-1, and Sentinel-2 to develop a RF model¹²⁸. This model extrapolated discrete building height samples from GEDI (150 m grid) in the Yangtze River Delta region in 2019 into a continuous building height map, providing an effective technical pathway for mapping recent building height distributions at fine grid scales, both regionally and globally. The results showed that when GEDI data were combined individually with Landsat-8, Sentinel-2, and Sentinel-1, as well as jointly with all four datasets, the resulting building height maps for the Yangtze River Delta achieved R^2 of 0.47, 0.80, 0.25, and 0.85 with reference heights, respectively. The corresponding RMSE values were 8.03 m, 5.62 m, 9.27 m, and 5.03 m. This indicates that, among single-source datasets, Sentinel-2 had the strongest explanatory power for building height, followed by Landsat 8, while Sentinel-1 performed relatively weakest. Furthermore, the building height map produced in this study demonstrated high consistency with the average building heights from the mapping work of Li et al., with R^2 reaching 0.88 and 0.95 at the city and provincial levels, respectively, and RMSE values of 4.77 m and 4.41 m.

The study by Ma et al. focused solely on building heights in the Yangtze River Delta region for 2019. To reveal the long-term evolution patterns of high-resolution building heights across the entire country, Chen et al. integrated multi-source data including GEDI RH96, Landsat-5/6, and PALSAR, and proposed a Multi-Temporal Built-up Height estimation Network model¹⁶⁶. Utilizing sample migration techniques to generate reference building height labels and incorporating time-series disturbance features extracted by the Continuous Change Detection and Classification algorithm, they constructed a 30-meter resolution China Multi-Temporal Building Height dataset (CMTBH-30). Validation of this dataset against an independent GEDI test set yielded R^2 values of 0.60, 0.58, 0.52, and 0.57 for the years 2005, 2010, 2015, and 2020, with corresponding RMSEs of 5.10 m, 5.53 m, 6.16 m, and 6.21 m. Data analysis revealed that the national average building height increased from 10.28 m in 2005 to 10.92 m in 2020, indicating a consistent upward development trend. The standard deviation of building height expanded from 5.16 m to 7.71 m, suggesting increased spatial heterogeneity across the country. For 2020, the building height estimates for major Chinese cities achieved an R^2 of 0.67 and an RMSE of 4.54 m when compared to field-collected measurements. At the individ-

ual building scale, the RMSE values for CMTBH-30, CNBH-10, and the method by He et al. compared to floor-derived height data were 11.8 m, 12.22 m, and 14.70 m, respectively. At the pixel scale, the RMSE values for CMTBH-30, CHSL-2023, and GBH-2020 were 10.13 m, 11.34 m, and 10.67 m, respectively. As a result, compared to existing products, CMTBH-30 provides a more detailed and accurate characterization of spatiotemporal building height changes, effectively mitigating the underestimation of high-rise structures.

3.4.12 Mapping other ecological parameters from waveform metrics

In recent years, with in-depth exploration of the physical significance of echo waveforms, ecological parameter mapping based on full-waveform LiDAR has seen significant progress. Its applications have rapidly expanded from traditional parameters such as forest canopy height, AGB, and building height to a wider range of critical ecological variables, including forest age, composite burn index, clumping index, 3D canopy complexity, canopy structural complexity, fuel load, canopy bulk density, and commercial crop distribution. These advances not only help quantify the impact of various drivers on the ecological environment and deepen the understanding of ecosystem structure, function, and their response mechanisms but also provide an important data foundation and methodological support for modeling carbon-water cycles, assessing fire risk, and supporting ecosystem management and conservation at global scales.

Using GEDI data without integrating other technical datasets, it is possible to generate a 3D canopy complexity distribution map of forests. Based on the GEDI RH1–RH100 (with a step of RH1), de Conto et al. employed XGBoost trees to derive a waveform structure complexity index and produced a near-global 3D canopy complexity map at approximately 1 km resolution¹¹⁹. The map showed an R^2 of 0.68 and an RMSE of 8.31% against airborne laser scanning-observed complexity data. When stratified by vegetation type, R^2 values for Evergreen Broadleaf Trees, Deciduous Broadleaf Trees, Evergreen Needleleaf Trees, and Grasslands/Shrublands/Woodlands were 0.65, 0.61, 0.58, and 0.36, with corresponding RMSEs of 5.78%, 7.59%, 8.07%, and 13.31%, respectively. The distribution map revealed that, at a given tree height, tropical forests contain a significantly higher proportion of high-complexity observations than other biomes: tropical forests concentrate the vast majority of high-complexity values, while less than 20% of temperate forest samples reach the median complexity level of tropical forests. Using SHAP feature explainers to analyze the relative contributions of lower, middle, and upper waveform layers to structural complexity prediction across different models and biomes, the results indicated that tropical forest complexity is more strongly associated with mid-to-lower canopy attributes, whereas temperate forests are more influenced by mid-to-upper canopy characteristics. Further inte-

grated analysis demonstrated that complexity and RH98 follow a power-law scaling relationship in most forests, but the scaling exponent varies with plant functional type and biome: across more than ten vegetation types and biomes examined, scaling exponents ranged from 0.91 to 1.33. The work of de Conto et al. systematically elucidates the spatial distribution patterns of forest structural complexity and underscores the importance of considering biome-specific characteristics and fine-scale variability in ecological research and management practices.

To identify the key drivers, such as mean annual temperature, mean annual precipitation, potential evapotranspiration, solar radiation, species richness, forest age, soil nitrogen, and soil texture of forest canopy structural complexity (CSC), Liu et al. proposed a method that uses canopy entropy to characterize CSC⁴². They established a mapping relationship between canopy entropy and canopy cover as well as FHD derived from GEDI-simulated waveforms through a linear regression model, thereby developing a forest CSC estimation model. This model was subsequently applied to actual GEDI observation data to generate a global forest CSC map. Global-scale validation showed that the GEDI-based estimates achieved an R^2 of 0.58 and an nRMSE of 12.35% when compared to airborne LiDAR-measured forest CSC. When validated by major biome types, the R^2 values for tropical and subtropical moist broadleaf forests, temperate broadleaf and mixed forests, and tropical and subtropical savannas were 0.47, 0.32, and 0.47, with corresponding nRMSE values of 14.33%, 13.52%, and 19.72%, respectively. By analyzing the relationships between forest CSC and various environmental factors, the study quantified the relative importance of biotic, climatic, and edaphic factors in explaining the variation in forest CSC. At the global scale, CSC is generally significantly positively correlated with both forest ecosystem productivity and stability. Moreover, compared to primary forests, managed forests exhibit lower CSC but demonstrate a more pronounced enhancing effect on ecosystem productivity and stability.

Existing forest height maps can serve as a data foundation for generating maps of new ecological parameters. By integrating forest canopy height maps, topographic data, and climate data, Lin et al. developed a 30 m resolution forest age distribution map for Fujian Province¹⁶⁷. Their study quantified the influence of forest height data sources, stand growth equations, and machine learning methods on the accuracy of forest age mapping, and revealed the contributions of factors such as canopy height, slope, aspect, mean temperature of the warmest month, and precipitation of the driest month. For the six forest types studied and five commonly used stand growth equations, the R^2 values ranged from 0.66 to 0.91, and RMSE values changed from 2.19 to 8.56 years. Among the five categories of machine learning methods evaluated, Boosted RF performed best for most forest types, achieving an R^2 as high as 0.96 and an RMSE as low as 4.26 years. Importance analysis using

Boosted RF indicated that tree height, driest monthly precipitation, and highest monthly temperature were the three top-ranked parameters for predicting forest age, with weights ranging from 48% to 77%, from 8% to 17%, and from 3% to 15%, respectively, across forest types. The study further compared the effectiveness of different forest height map sources, finding that maps generated using the height data from Liu et al.¹²⁷ significantly outperformed those based on Potapov et al.¹³ in terms of accuracy. Moreover, models relying solely on integrated auxiliary data, without species classification or data on diameter at breast height and forest density, still achieved high mapping precision ($R^2 = 0.92$, and RMSE = 5.89 years for Boosted RF).

Zhao et al. established polynomial functions based on the simple ratio for different plant functional types (including evergreen broadleaf forest, deciduous broadleaf forest, evergreen needleleaf forest, deciduous needleleaf forest, mixed forest, shrubs, and wetlands) and varying vegetation foliage densities over flat terrain, and integrated GEDI canopy height data with MODIS CI to generate a new 30 m resolution global forest CI product¹⁶⁸. Validation across 39 sites on three continents yielded promising results: compared with field-measured CI, the model achieved an R^2 of 0.42 and RMSE of 0.0180 for broadleaf forests, an R^2 of 0.48 and RMSE of 0.0206 for needleleaf forests, and an overall R^2 of 0.84 with an RMSE of 0.0199. This study enhances the spatial resolution of the global forest CI map from 500 m to 30 m, substantially improving the representation of its spatial patterns.

Utilizing WorldView-2 and GEDI data, Lin et al. identified two assessment metrics, namely, Relative Spectral Change (RSC) and Relative Waveform Energy Change (RWC), to quantify burn severity, thereby supporting the understanding of post-fire vegetation recovery patterns and ecosystem resilience¹²⁰. RWC is defined as

$$RWC_i = \frac{CE_{pre} - CE_{post}}{CE_{pre}}. \quad (36)$$

Where subscript i represents the overstory or understory layer, and subscripts pre and post refer to pre-fire and post-fire, respectively, and CE refers to the cumulated waveform energy. By integrating RSC and RWC through weighted scoring, a novel remote sensing-based Hybrid Composite Burn Index (HCBI) was developed to assess burn severity at the overstory, understory, and total site levels. Results indicated that the index exhibited high sensitivity across varying burn severity and vegetation coverage levels. As the composite burn index (CBI) increased from 0 to 3.0, the RWC in the understory rose from being negligible to over 75%, while that in the overstory increased from less than 5% to 100%. At fractional vegetation cover values of 0.4, 0.6, and 0.8, the correlation between the HCBI and the CBI for the overstory, understory, and total site ranged from 0.969 to 0.995, with RMSE values between 0.07 and 0.15. Subsequently, spatially discrete HCBI maps for the overstory, understory, and total

site were generated for the Xiushan fire scar in the Greater Xing'an Mountains using WorldView-2 multispectral imagery and GEDI LiDAR data. A continuous, wall-to-wall HCBI map was then produced by applying a RF algorithm. Validation showed that the wall-to-wall HCBI map correlated with field-measured HCBI at 0.89, 0.84, and 0.87 for the overstory, understory, and total site levels, with RMSE values of 0.30, 0.34, and 0.26, respectively. Compared to field-based CBI measurements, the R^2 values were 0.93, 0.74, and 0.87, with RMSE values of 0.53, 0.31, and 0.28. Therefore, the wall-to-wall HCBI map can capture fine-scale horizontal and vertical variability even in cases where only the understory is burned. It serves as an advanced alternative to field-based CBI assessments, offering new opportunities for advancing overstory and understory burn severity evaluation in heterogeneous forests and supporting detailed forest management decision-making.

To accurately identify areas at high fire risk and to assess and mitigate potential fire impacts, Aragoneses et al. utilized metrics such as ground elevation and RH100 from GEDI data to achieve canopy energy separation, from which LAI and Leaf Area Density were derived¹²¹. Building upon a gap probability model and using these parameters as inputs, they developed estimation models for footprint-level canopy fuel load and canopy bulk density. Across three study areas in the United States, north of Sacramento, Idaho, and Yosemite National Park, as well as their combined dataset, the R^2 between GEDI footprint-level canopy bulk density and the LANDFIRE bulk density reference map were 0.86, 0.60, 0.77, and 0.79, with corresponding RMSE values of 0.02, 0.05, 0.02, and 0.03, respectively. Subsequently, by integrating the GEDI-derived footprint-level canopy fuel load and bulk density data with Landsat 8 imagery and L-band Phased Array SAR data, the study employed two machine learning models, RF and K-Nearest Neighbour, to generate wall-to-wall maps of canopy fuel load and canopy bulk density across Europe at a 1-km grid resolution. The results showed that both models achieved a correlation of 0.85 and an RMSE of 0.06 for canopy fuel load predictions against reference values. For canopy bulk density, the correlations were 0.75 and 0.74 for the two models, both with an RMSE of 0.03.

To map the distribution of tree crops, Adrah et al. employed Time-Weighted Dynamic Time Warping and a RF model to evaluate classification performance in distinguishing tree crops from other vegetation classes at the GEDI footprint level, testing different combinations of input features¹²². Under three study scenarios, Eastern Mediterranean, Southern France, and a combination of both, and across five zones graded from wettest to driest, the overall footprint-level classification accuracies changed between 0.85 and 0.98, with corresponding F1-scores between 0.80 and 0.92. Subsequently, using the GEDI footprint-level classification results as training samples, a RF classifier was trained on fused Sentinel-1 and Sentinel-2 imagery to gener-

ate wall-to-wall tree crop maps. Across the five zones from wettest to driest, the overall mapping accuracies achieved were between 0.73 and 0.92, with F1-scores between 0.70 and 0.90. This work reveals variations in accuracy across different agro-climatic zones and underscores the value of vegetation structural information for large-scale tree crop mapping.

To reveal how ecosystem structure and diversity influence the upscaling of eddy covariance carbon and water fluxes, Bu et al. integrated GEDI-derived metrics, including canopy height, FHD, total canopy cover, total PAI, and total gap probability, with data from over 90 flux towers in the AmeriFlux and NEON networks, MODIS-derived vegetation near-infrared reflectance, and Daymet meteorological data¹⁶⁹. They developed machine learning prediction models for gross primary productivity (GPP) and evapotranspiration (ET), and used XGBoost to generate daily-scale, 1-km resolution gridded carbon and water flux datasets over the contiguous United States from 2019 to 2023. The inclusion of GEDI canopy height and FHD features substantially improved model performance: for simulated GPP estimation, R^2 increased from 0.79 to 0.91, and RMSE decreased from $1.77 \text{ g C m}^{-2} \text{ d}^{-1}$ to $1.14 \text{ g C m}^{-2} \text{ d}^{-1}$; for simulated ET estimation, R^2 rose from 0.79 to 0.85, and RMSE dropped from 0.82 mm/d to 0.68 mm/d. Further analysis revealed that annual GPP and ET exhibit a positive logarithmic relationship with canopy structural complexity (represented by FHD), increasing with complexity but with a diminishing marginal effect. Higher canopy complexity was also found to reduce seasonal variability in both GPP and ET. During severe drought events, canopy structural complexity enhanced ecosystem drought resilience by mitigating losses in carbon and water fluxes.

3.4.13 Significance and applications of full-waveform LiDAR in ecological parameter retrieval

The aforementioned studies have established relational models between target ecological characteristics and echo waveform features through empirical formulas or machine learning methods, achieving high-precision retrieval of ecological parameters at the footprint level. By further integrating multi-source remote sensing data such as GEDI, Sentinel, and Landsat, regional- to global-scale ecological parameter maps have been generated. In recent years, GEDI data and its derived products have been widely applied across multiple fields, providing revolutionary tools for monitoring key climate variables, offering new perspectives for quantifying terrestrial carbon stocks and understanding forest ecosystem processes, and strongly supporting the formulation of ecological and environmental policies.

In the fields of topographic mapping and understanding of terrestrial ecosystems, Margruder et al. systematically reviewed the contributions of the ICESat, ICESat-2, and GEDI missions to earth and climate science¹⁷⁰. Over more than two decades of continuous observations, the data have

clearly revealed an ongoing contraction of the global cryosphere: between 2003 and 2021, the melting of the Greenland and Antarctic ice sheets led to a global land ice loss of approximately 320 billion tons per year, while Arctic winter sea ice volume declined by 30% over the same period. These observations have also served as crucial tools for understanding the dynamics of terrestrial ecosystems. At the hydrosphere level, studies showed that 57% of the global seasonal variation in terrestrial water storage can be attributed to the regulation of human-managed reservoirs. In the biosphere, despite carbon losses equivalent to approximately 8 billion tons of CO_2 due to land-use changes, global forest carbon stocks still achieved net growth through forest growth. Furthermore, this technology has played a significant role in atmospheric sciences, enabling the quantitative assessment of global vertical cloud distribution, aerosol composition, and the transport pathways of dust and smoke.

In the field of forest canopy height and AGB research, several studies have demonstrated the value of full-waveform LiDAR. Wang et al. employed a Gradient Boosted Machines model, combining GLAS AGB data with Landsat-8 imagery to generate an annual AGB dataset for North America¹⁷¹. This work quantified the combined impact of vegetation dynamics and disturbances on the regional carbon budget, suggesting that current earth system models may overestimate the accumulation of surface biomass by more than a factor of two.

Kalischek et al. used GEDI canopy height data as a foundation, integrated Sentinel-2 imagery, and applied a convolutional neural network to map the height distribution of cocoa trees in Côte d'Ivoire and Ghana¹⁷². This allowed them to quantify the expansion scale, yield, and economic benefits of cocoa cultivation within protected areas. The study provides key data supporting the understanding of the trade-offs between ecological conservation and economic development in cocoa-producing regions, aiding in the formulation of evidence-based sustainable development strategies.

With the support of canopy height data from GEDI, Sentinel-2, and Landsat, Liu et al. used 3 m resolution nanosatellite imagery to map canopy cover, height, and AGB across Europe¹⁷³. Their approach achieved cross-landscape spatial analytical capability, particularly enhancing the accurate quantification of trees outside forests, offering an innovative technical solution for improving national carbon accounting systems.

Xie et al. utilized GEDI RH95 data to construct the most comprehensive global alpine treeline species dataset to date, covering 38 mountain regions across 43 countries¹⁷⁴. They systematically analyzed the spatial patterns and environmental drivers shaping global treeline species distribution, revealing the synergistic role of heat and moisture in regulating community composition differentiation at the treeline. This provides crucial insights for studying treeline responses under climate change.

Xu et al. built a CNN classification model linking GEDI, Sentinel-1A, and Landsat data to produce a high-precision 30 m resolution forest extent map for 2020¹⁷⁵. Furthermore, they reconstructed the spatiotemporal upper range limit of trees at 100-year intervals since the Last Glacial Maximum. This work revealed the impact mechanism of historical fluctuations in the tree distribution limit on the spatial variation of alpine plant community composition specific to the Tibetan Plateau, indicating that understanding current patterns of alpine plant beta diversity requires simultaneous consideration of historical climate change and the dynamics of the upper tree line.

In the fields of carbon storage and global ecosystem research, Duncanson et al. estimated aboveground carbon density by multiplying GEDI-derived AGBD by a conversion factor of 0.49, producing an accurate global carbon stock map¹⁷⁶. Their work quantified the aboveground carbon stored within global protected areas and the additional carbon sequestration attributable to protection effects, highlighting the crucial role of conserving high-biomass forests for emissions reduction and maintaining future carbon sequestration capacity.

Reiner et al. integrated ultra-high-resolution imagery from the PlanetScope nanosatellite constellation with data such as GEDI RH95 to map forest and non-forest tree cover across the African continent in 2019¹⁷⁷. This achieved continent-scale, cross-nationally consistent mapping accurate to the individual tree level. The results revealed that 29% of tree cover is located in areas not classified as tree cover in the most advanced existing maps. This finding may redefine the understanding of land-use impacts in non-forest landscapes and provides a new foundation for nature-based climate solutions and related research.

Li et al. defined a Canopy Structural Index based on GEDI data and MODIS kNDVI, mapping forest structural density at a near-global scale, and clarified the relative impacts of climate versus human activity on forest structure: in 35.1% of global forests, human activity is the dominant driver of regional-scale differences in structural density¹⁷⁸.

Brodie et al. employed Kriging interpolation to align GEDI sample data across the pantropical region with the precise locations of biodiversity sampling points¹⁷⁹. Using five GEDI-derived 3D canopy structure metrics, the study assessed large-scale biodiversity conservation outcomes. It confirmed that protected areas not only enhance vertebrate diversity within their boundaries but also promote biodiversity conservation across broader regions, providing scientific support for the United Nations' target of protecting 30% of terrestrial areas by 2030.

Building upon the global forest height map produced by Lang et al.¹⁴⁸, Liu et al. applied a causality-oriented, explainable AI framework to parse the relative contributions of soil moisture supply and atmospheric moisture demand to gross primary productivity¹⁸⁰. This research deepened the under-

standing of how soil and atmospheric drought affect ecosystem photosynthesis, offering key insights into how terrestrial ecosystems respond to increasing aridity. It is particularly significant for understanding the potential shift of ecosystems from energy limitation to water limitation.

Forkel integrated multi-source earth observation data, including from TU Dresden and GEDI, with models such as RF, JULES, and OCN. The study revealed that fire emissions in the Amazon and Cerrado biomes primarily originate from the smoldering combustion of woody debris¹⁸¹. Compared to the Cerrado, higher CO emissions in the Amazon stem from a greater proportion and load of woody debris driving smoldering in understory vegetation and deforestation fires. Furthermore, landscape fragmentation, deforestation, and drought not only increase fire frequency but also intensify smoldering combustion, thereby amplifying fire activity in the Amazon. These compound effects underscore the necessity of mitigating the combustion of post-deforestation residual fuels and natural woody debris to reduce fire emissions and their impacts on health and the carbon cycle.

Brandt et al. systematically reviewed the latest advances and applications of artificial intelligence technologies, such as convolutional neural networks, vision transformers, and self-supervised learning models, in analyzing spaceborne LiDAR data and building forest monitoring systems¹⁸². Their work clarifies the contributions of these technologies to the accurate assessment of carbon stocks and the support of climate change mitigation efforts.

3.4.14 Characters of target characteristic inversion method based on waveform metrics

The procedure of retrieving the properties of targets in terms of waveform metrics is relatively straightforward: it does not require the complex processing steps inherent to waveform decomposition, nor does it involve the trade-off between accuracy and computational efficiency typical of waveform simulation methods. Consequently, it offers advantages of simplicity in processing and high computational efficiency. However, the method also has clear limitations: it typically ignores weak echo signals from small targets or those with low reflectivity, and it does not account for the subtle differences between "pseudo-Gaussian" echoes generated by adjacent targets and genuine single-target Gaussian echoes. Therefore, this method is less suitable for multi-target or structurally complex scenes and is primarily geared towards coarse parameter inversion where precision requirements are relatively lenient. Additionally, the method requires improvement in the following three key areas:

First, the selection of waveform metrics lacks universality. As shown in Table S6, the optimal RH metric for characterizing building or forest canopy height could be RH95, RH96, RH99, or RH100, with no single, universally optimal value applicable to all scenarios. Using RH5 to represent the forest

canopy bottom signal yields more accurate height inversion results compared to using RH0. Simultaneously, the performance of the optimal RH metrics varies significantly depending on the region, forest type, and terrain conditions, resulting in R^2 for inversion models that can be as high as > 0.9 or as low as < 0.4 .

Second, the construction of correlation models between ecological parameters and waveform features is complex. Such models are largely based on empirical relationships. As shown in Table S7, Kellner et al. constructed up to 14 complex AGB models based on combinations of different continents and forest types; these models primarily employed sqrt-sqrt transformations and largely took the form of RH98 combined with other RH parameters. Duncanson et al. proposed an original-log transformation form between RH98 and AGB. The accuracy range of these models fluctuates considerably: for instance, the R^2 for inversions using Duncanson's models ranged between 0.36 and 0.9. To correct for influences such as terrain and forest type, Nie, Ni, Li, and Ni-Meister et al. significantly improved AGB inversion accuracy by constructing linear models that incorporated AGB with a series of RH metrics, FHD, PAI, canopy cover, and their combinations.

Third, as exhibited in Tables S8 and S9, despite the various methods proposed for extrapolating discrete GEDI footprint ecological information to unobserved areas and subsequently constructing the global ecological mapping, such products show regional (e.g., topography and forest type) inconsistency. Taking forest canopy height as an example, researchers have utilized parameters such as RH100, RH98, and RH95, combined with multi-source data from Landsat, Sentinel, Tandem-X, and even Google Earth Engine, and applied various algorithms including RF, NNGL, probabilistic deep learning models, hybrid modeling frameworks, and XGBoost to generate regional and global canopy height maps with spatial resolutions ranging from 10 meters to kilometers. These studies have revealed the impact of data sources, modeling methods, and environmental variables on final map accuracy. The validation accuracy of the same canopy height map can vary dramatically across different regions. A notable example is the global 30 m resolution forest canopy height map for 2019 produced by Potapov et al. Its own validation reported an R^2 of 0.62, but independent validations by other researchers showed stark contrasts: Wang et al. reported R^2 values of 0.67 in SERC, 0.02 in Lantau Island, and 0.66 in CBHNP, Xiao et al. reported an R^2 of 0.32 for China (canopy height > 3 m), and Zhang et al. reported an R^2 as low as 0.06 in Hainan Tropical Rainforest National Park¹⁵⁴.

4 Application of full-waveform LiDAR for space surveillance

Currently, full-waveform LiDAR technology is primarily

focused on earth observation. Its application in the field of aerial and space target monitoring remains largely at the conceptual stage, and the payload systems comparable to those deployed on ICESat-1, GEDI, GaoFen-7, and TECIS have yet been developed for such purposes. Steinvall et al. have simulated and measured the echo waveforms modulated by various targets, ranging from fighter and attack aircraft (such as the JAS39, SK-60, Saab-37, F15, F16, and F117), and warships to unmanned aerial vehicles and small vessels of different classes and sizes^{183,184}. His work confirmed the feasibility of target classification based on echo waveform characteristics. However, Steinvall's studies are limited to targets within several kilometers but do not address the extraction of characteristic parameters from the waveforms.

In the domain of space target detection, the Shanghai Astronomical Observatory, the Purple Mountain Observatory, and the Yunnan Observatories have initiated laser ranging studies of high-velocity space objects such as space debris^{185,186}. Using a 1.2 m telescope laser ranging platform, the Yunnan Observatories have achieved ranging of space debris within a range of 400 to 900 km and for radar cross-sections between 9.0 and 20.0 m²¹⁸⁵. This space debris ranging is around-the-clock, despite daytime background noise being 10⁶ times stronger than at night, and thus demonstrated all-day laser ranging capability for space debris. The Shanghai Astronomical Observatory retrofitted a 1.2 m space optical telescope and achieved laser ranging of cooperative satellites in geosynchronous orbit with a 30-picosecond laser pulse width and an accuracy better than 2 cm¹⁸⁶. They performed laser ranging for a space debris target at a distance of 1620.5 km with a radar cross-section of 2.41 m², attaining an accuracy of 10.64 cm. Laser ranging of space targets is evolving from single-station to multi-station collaborative observation¹⁸⁶, but it has not been demonstrated to fine-scale map the characteristics such as material composition, 3D structure, attitude, and kinematic state of space debris using full-waveform LiDAR.

Reflective Tomography LiDAR (RTL) is a specialized operational mode of full-waveform LiDAR. It reconstructs 2D or 3D images of targets by collecting full-waveform echoes from different relative motion angles and applying tomographic principles¹⁸⁷. The theoretical foundation of this technology can be traced back to the pioneering experiment by Parker et al. in 1988, which first verified the feasibility of applying tomographic imaging principles to LiDAR¹⁸⁸. In 2001, Matson et al. applied RTL technology from a ground-based platform to detect an orbiting satellite, successfully reconstructing the relative positions of two corner reflectors spaced 6 m apart on a satellite at a distance of 514 kilometers, marking a breakthrough from principle validation to long-range application¹⁸⁹. Subsequent research focused on improving imaging resolution and practicality. In 2010, Murray et al. performed RTL imaging on a 1 m wide composite target at 22.4 km, achieving an imaging resolu-

tion of 15 cm, an order of magnitude improvement over conventional telescopic imaging¹⁹⁰. At the same year, the team of Henriksson pioneered the integration of single-photon time-correlated detection into RTL, realizing photon-level ultra-high-resolution tomographic imaging of complex targets in a laboratory setting, thereby opening a new avenue based on single-photon detection¹⁹¹.

Recent advancements have centered on algorithm optimization and system integration. Hu et al. optimized the data processing workflow and proposed a reconstruction algorithm more adaptable to real-world scenarios for incomplete angular data¹⁹². In 2023, they achieved super-resolution imaging of a complex target at 10.38 km with a state-of-the-art resolution of 1.7 cm. Furthermore, the team innovated a multi-angle, multi-field-of-view overlapping sampling and image fusion algorithm, and enhanced the experimental system, ultimately accomplishing 3D image reconstruction with resolutions of 2.0 cm in the orbital plane and 3.46 cm along the orbital axis. In 2025, Xu Feihu et al. demonstrated the feasibility of single-photon RTL, achieving an imaging resolution of 1.2 cm at 7.1 km, which fully underscores the significant application potential of this technology¹⁹³. Guo et al. proposed a vertical optical field framework to surpass the diffraction limit, achieving reflective Fourier ptychography imaging with a resolution of 11.2 mm at a distance of 3.2 km, which corresponds to 0.37 times the classical diffraction limit¹⁹⁴. Furthermore, by employing a laser with a pulse width of 7.2 ps and integrating time-correlated single-photon counting with superconducting nanowire single-photon detectors, Guo et al. realized RTL imaging with a resolution of 8 mm at a distance of 3.3 km. These advancements provide crucial technical support for space debris monitoring and satellite identification^{194,195}.

Both theoretical analysis and terrestrial observations have validated the space observation capabilities of full-waveform LiDAR. However, directly applying mature terrestrial observation methods to the space domain faces fundamental challenges, primarily due to inherent differences in the detection scenarios. First, the absence of a reference echo. In earth observation, the ground return serves as a reliable terminal signal, crucial for determining waveform metrics like length and RH, and acts as a reference baseline for inverting parameters such as forest canopy height in waveform decomposition or deconvolution. In space surveillance, no such stable, known reference echo exists, rendering the entire inversion methodology built around the ground return, especially for obtaining target depth information, inapplicable. Second, the complexity of target properties. Terrestrial targets are typically stationary, with fixed orientation and relatively uniform surface optical properties. In contrast, space objects are mostly artificial, with complex shapes, diverse surface optical characteristics, and they undergo high-speed motion with rapidly changing attitudes. Coupled with detection distances often ranging from hundreds to thousands of kilometers, these factors pose

immense challenges for target acquisition, echo collection, and data analysis.

5 Summary and outlook

This paper systematically introduces the processing methods for full-waveform LiDAR echoes, primarily including waveform simulation, analytical expression, and waveform decomposition and deconvolution. Building on this, the target feature inversion methods can be categorized into four classes, specifically, inversion based on waveform simulation, analytical echo waveform expressions, waveform decomposition and deconvolution, and waveform feature metrics, respectively. This paper reviews the latest research advances in each category and discusses the potential application in space object surveillance in detail.

5.1 Summary of the waveform simulation-based approach for target feature inversion

This method assigns the scene parameters used in the simulation to the actual target when the match between the simulated waveform and the actual exceeds a present threshold. Koetz, Hmida, Bye, and Yang et al. have confirmed that this method imposes no constraints on the target type or attributes. It can be widely applied to invert the properties ranging from building height to ecological parameters, including fractional vegetation cover, vertical crown extent, height to the first live branch, etc.

This method faces two major limitations: reliance on prior knowledge and the trade-off between the efficiency of constructing an echo waveform database and the accuracy of inverting target characteristics. This may explain the limited application of this method for target feature acquisition in recent years. Future breakthroughs are likely to hinge on: I) gaining a deeper understanding of the pulsed laser-target interaction mechanism and developing analytical models of the interaction process to replace the single-photon transport simulation under the radiative transfer model, thereby reducing the simulation time per waveform; II) employing multi-sensor fusion to decrease the number of unknown parameters; and III) incorporating approaches such as quantum computing, such as Jiuzhang and Zuchongzhi^{196,197}, to enhance processing capability. Progress in these three areas could lead to a significant revival of waveform simulation-based approaches.

5.2 Summary of the analytical formula-based approach for target feature inversion

This method mathematically characterizes target parameters and constructs echo waveform equations that incorporate these variables. Hu et al. developed analytical expressions for modulated echo waveforms from typical 2D and 3D targets, independent of factors such as attitude and posi-

tion, systematically revealing the modulation mechanisms by which target features affect echo waveforms. For vertically illuminated 2D targets, the method enables fine-grained retrieval of lateral size of non-extended target within the laser footprint, with an absolute error not exceeding 0.17 cm (a relative error below 3.4% relative to the footprint radius). For tilted 2D targets, a convolutional neural network shape discrimination algorithm integrating attention mechanisms and feature engineering was proposed, achieving 82.87% shape classification accuracy under high-noise conditions, along with a target size inversion error of 13.31% and a tilt angle inversion error of 6.42°.

This approach offers advantages over other full-waveform LiDAR methods in characterizing single-target properties. Its main challenges lie in the analytical description of target shapes, shape recognition, and matching the dimensionality of target attributes with waveform features. To apply this method to actual waveform analysis, the following breakthroughs are required: I) introducing descriptive functions for irregular shapes, such as superquadric surfaces and non-uniform rational B-splines^{198,199}; II) developing generalized and simplified analytical echo waveform equations for targets with similar shapes that cannot be distinguished via remote sensing; and III) proposing multi-pulse cooperative observation strategies to construct systems of echo waveform equations, advancing overdetermined optimization methods for target feature inversion, and quantifying inversion accuracy and influencing mechanisms.

5.3 Summary of the waveform decomposition-based approach for target feature inversion

Waveform decomposition and deconvolution methods employ analytical functions to process echo waveforms. The parameters of each component are used to invert various target features, ranging from forest canopy/building height to LAI and clumping index. Beyond feature retrieval, these methods help elucidate the impacts of factors like within- and between-crown aggregation effects, and geolocation bias on the inversion accuracy of ecological parameters. They are also widely applied in domains varying from tree species classification to forest pest/disease monitoring.

The mathematical procedure for waveform decomposition is relatively straightforward, whereas the key advantage of waveform deconvolution is its independence from prior knowledge or initial value estimation. Future efforts should focus on: I) establishing new waveform decomposition functions, such as the adaptive asymmetric Gaussian, based on an in-depth investigation of target-induced modulation mechanisms on echo waveforms; II) introducing the artificial intelligence techniques, III) exploring inter-channel signal correlations in hyperspectral data; and IV) quantifying the influencing factors and their magnitudes of waveform decomposition methods through simulations and measured results, clarifying the applicable boundaries of

different decomposition approaches. Ultimately, these will enable the decomposition of extremely weak and highly overlapped echo signals, as well as the correct interpretation of parameters for the decomposed echo components.

5.4 Summary of the waveform metrics-based approach for target feature inversion

This method revolves around extracting metrics from full-waveform LiDAR echoes, establishing correlation models between these metrics and target attributes. Commonly used waveform metrics include waveform length and RH. The application is extensive, including all continents, all terrain conditions, all vegetation and forest types. The invertible ecological parameters are highly diverse, ranging from forest growth to fuel load and canopy bulk density, ecosystem productivity and stability, burn severity assessment and recovery.

This method offers advantages of simplicity in processing and high computational efficiency, with the imperfections of ignoring weak echo signals and unidentifying pseudo-Gaussian echoes. It is thus less suitable for multi-target or structurally complex scenes. The relationship between waveform metric and target feature also lacks theoretical foundations. The future research should focus on the following three directions: I) Developing mechanism-driven, adaptive modeling to facilitate the extraction of more physically grounded correlation models adapted to specific scenarios; II) Advancing echo waveform processing methods adapted to regional and forest type conditions, such as by defining effective signal start/stop thresholds; and III) Optimizing multi-source fusion and mapping algorithms to achieve high-resolution, high-precision global mapping of ecological parameters.

5.5 Summary of application of full-waveform LiDAR in space surveillance

The application of full-waveform LiDAR in space surveillance is at a critical juncture, transitioning from methodological research to systematic development. Although various echo waveform analysis methods have been developed for earth observation, they are generally unsuitable for analyzing modulated echo waveforms from space targets due to factors such as the lack of reference echo signals and the more complex shapes and attitudes of space objects. Therefore, the successful application of full-waveform LiDAR to space surveillance necessitates a breakthrough beyond existing data processing frameworks and the establishment of the new target feature inversion system. Given that high-speed space targets are usually far apart, often resulting in only a single target within the laser footprint, and with single scattering being dominant, the physical scene is somewhat simplified. Under these conditions, a viable technical pathway involves: employing facet models or analytical formula

methods (combing with the formula to describe the irregular target) to systematically construct a comprehensive echo waveform database encompassing targets of various shapes, sizes, positions, attitudes, and surface optical properties; based on this, deeply analyzing the modulation mechanisms of target features on the echo waveform to excavate the mapping relationships between them; and finally, exploring and establishing a set of high-precision feature inversion methods specifically tailored for space objects, potentially leveraging array detectors or multi-station cooperative detection schemes used in laser ranging.

In summary, full-waveform LiDAR has demonstrated significant potential in earth observation, while its application in space surveillance represents a crucial frontier awaiting exploration. Currently, data analysis and target inversion for spaceborne full-waveform LiDAR rely primarily on GEDI data, and are increasingly integrated with optical, spectral, and radar technologies to achieve synchronous and holistic observation of surface structure, function, and composition. Satellites such as TECIS and those in the Gaofen series are equipped with a suite of sensors, granting them unique systematic advantages in coordinating and fusing multi-source remote sensing data. Looking ahead, we should leverage these advanced hardware platforms as an opportunity to deeply investigate the physical mechanisms of pulsed laser interaction with both terrestrial and space targets. The goal is to develop universal echo waveform analysis theories and target feature inversion methods, fully unlocking the potential of full-waveform data for fine-grained perception of multi-dimensional target information. This will greatly enrich the data processing workflows and product systems for domestic satellites like Gaofen-7 and TECIS. It also substantially enhances China's scientific influence and international discourse power in fields such as global topographic mapping, ecosystem characterization, carbon cycle research, and space situational awareness.

References

- Krieger G, Moreira A, Fiedler H et al. TanDEM-X: a satellite formation for high-resolution SAR interferometry. *IEEE Trans Geosci Remote Sens* **45**, 3317–3341 (2007).
- Wulder MA, Loveland TR, Roy DP et al. Current status of Landsat program, science, and applications. *Remote Sens Environ* **225**, 127–147 (2019).
- Liu YN, Sun DX, Hu XN et al. The advanced hyperspectral imager: aboard China's GaoFen-5 satellite. *IEEE Geosci Remote Sens Mag* **7**, 23–32 (2019).
- Aguilar MA, Nemmaoui A, Aguilar FJ et al. Quality assessment of digital surface models extracted from WorldView-2 and WorldView-3 stereo pairs over different land covers. *GISci Remote Sens* **56**, 109–129 (2019).
- Dubayah R, Blair JB, Goetz S et al. The Global Ecosystem Dynamics Investigation: high-resolution laser ranging of the earth's forests and topography. *Sci Remote Sens* **1**, 100002 (2020).
- Tang XM, Xie JF, Liu R et al. Overview of the GaoFen-7 laser altimeter system mission. *Earth Space Sci* **7**, e2019EA000777 (2020).
- Su HJ, Yao WJ, Wu ZY et al. Kernel low-rank representation with elastic net for China coastal wetland land cover classification using GF-5 hyperspectral imagery. *ISPRS J Photogram Remote Sens* **171**, 238–252 (2021).
- Chen LF, Letu, H, Fan M et al. An introduction to the Chinese high-resolution earth observation system: Gaofen-1~7 civilian satellites. *J Remote Sens* **2022**, 9769536 (2022).
- Wulder MA, Roy DP, Radeloff VC et al. Fifty years of Landsat science and impacts. *Remote Sens Environ* **280**, 113195 (2022).
- Yu X, Wang Z, Du XY et al. Multipass interferometric ISAR for three-dimensional space target reconstruction. *IEEE Trans Geosci Remote Sens* **60**, 5110317 (2022).
- Casalini E, Domínguez EM, Henke D. Multi-view 3-D radar imaging of moving targets with time-domain processing. *IEEE Trans Geosci Remote Sens* **60**, 5219212 (2022).
- Zwally HJ, Schutz B, Abdalati W et al. ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *J Geodyn* **34**, 405–445 (2002).
- Potapov P, Li XY, Hernandez-Serna A et al. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens Environ* **253**, 112165 (2021).
- Chen JY, Li GY, Peng J et al. Accuracy analysis of forest canopy height from waveform LiDAR of the Terrestrial Ecosystem Carbon Inventory Satellite (TECIS): a case study of a temperate coniferous-broadleaved mixed forest in Quebec, Canada. *Nat Remote Sens Bull* **29**, 2958–2975 (2025).
- Lefsky MA, Cohen WB, Parker GG et al. Lidar remote sensing for ecosystem studies. *Bioscience* **52**, 19–30 (2002).
- Steinvall O. Effects of target shape and reflection on laser radar cross sections. *Appl Opt* **39**, 4381–4391 (2000).
- Johnson SE. Effect of target surface orientation on the range precision of laser detection and ranging systems. *J Appl Remote Sens* **3**, 033564 (2009).
- Li YH, Wu ZS. Targets recognition using subnanosecond pulse laser range profiles. *Opt Express* **18**, 16788–16796 (2010).
- Hao Q, Cheng Y, Cao J et al. Analytical and numerical approaches to study echo laser pulse profile affected by target and atmospheric turbulence. *Opt Express* **24**, 25026–25042 (2016).
- Xu XB, Zhang H, Luo MZ et al. Research on target echo characteristics and ranging accuracy for laser radar. *Infrared Phys Technol* **96**, 330–339 (2019).
- Hu YH, Hou AH, Ma QL et al. Analytical formula to investigate the modulation of sloped targets using LiDAR waveform. *IEEE Trans Geosci Remote Sens* **60**, 5700512 (2022).
- Hu YH, Zhang XY, Hou AH et al. Laser echo waveform modulation modelling from lateral structure using a mathematical formula. *Int J Remote Sens* **44**, 2382–2399 (2023).
- Zhang ZJ, Xie H, Tong XH et al. Research progress of full waveform processing technology of satellite laser altimetry. *Sci Surv Mapp* **44**, 168–178 (2019).
- Huang JP, Li GY, Liu Z. Current status and development trend in estimating forest structural parameters with spaceborne LiDAR. *Trans Chin Soc Agric Mach* **55**, 18–33 (2024).
- Duncanson L, Neuenschwander A, Hancock S et al. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sens Environ* **242**, 111779 (2020).
- Xie JF, Huang GH, Liu R et al. Design and data processing of China's first spaceborne laser altimeter system for earth observation: GaoFen-7. *IEEE J Sel Top Appl Earth Obs Remote Sens* **13**, 1034–1044 (2020).
- Liu AB, Cheng X, Chen ZQ. Performance evaluation of GEDI and

- ICESat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sens Environ* **264**, 112571 (2021).
28. Duncanson L, Kellner JR, Armston J et al. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sens Environ* **270**, 112845 (2022).
 29. Zhu XX, Nie S, Wang C et al. Consistency analysis of forest height retrievals between GEDI and ICESat-2. *Remote Sens Environ* **281**, 113244 (2022).
 30. Li XX, Wessels K, Armston J et al. First validation of GEDI canopy heights in African savannas. *Remote Sens Environ* **285**, 113402 (2023).
 31. Li HT, Li XX, Kato T et al. Accuracy assessment of GEDI terrain elevation, canopy height, and aboveground biomass density estimates in Japanese artificial forests. *Sci Remote Sens* **10**, 100144 (2024).
 32. Li Y, Gao SJ, Fu HQ et al. Error analysis and accuracy improvement in forest canopy height estimation based on GEDI L2A product: a case study in the United States. *Forest* **15**, 1536 (2024).
 33. Du LM, Pang Y, Ni WJ et al. Forest terrain and canopy height estimation using stereo images and spaceborne LiDAR data from GaoFen-7 satellite. *Geo-spat Inf Sci* **27**, 811–821 (2024).
 34. Chen PM, Huang HB, Qin P et al. Unveiling the performance and influential factors of GEDI L2A for building height retrieval. *GISci Remote Sens* **62**, 2498785 (2025).
 35. Gastellu-Etchegorry JP, Yin TG, Lauret N et al. Discrete anisotropic radiative transfer (DART 5) for modeling airborne and satellite spectroradiometer and LIDAR acquisitions of natural and urban landscapes. *Remote Sens* **7**, 1667–1701 (2015).
 36. Guo K, Xu WX, Liu YX et al. Gaussian half-wavelength progressive decomposition method for waveform processing of airborne laser bathymetry. *Remote Sens* **10**, 35 (2018).
 37. Yang XB, Wang C, Xi XH et al. Extraction of multiple building heights using ICESat/GLAS full-waveform data assisted by optical imagery. *IEEE Geosci Remote Sens Lett* **16**, 1914–1918 (2019).
 38. Hu YH, Hou AH, Zhang XY et al. Assessment of lateral structural details of targets using principles of full-waveform light detection and ranging. *IEEE Trans Geosci Remote Sens* **61**, 5704116 (2023).
 39. Blair JB, Hofton MA. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophys Res Lett* **26**, 2509–2512 (1999).
 40. Hancock S, Armston J, Hofton M et al. The GEDI simulator: a large-footprint waveform lidar simulator for calibration and validation of spaceborne missions. *Earth Space Sci* **6**, 294–310 (2019).
 41. Tan S, Zhang Y, Qi JB et al. Exploring the potential of GEDI in characterizing tree height composition based on advanced radiative transfer model simulations. *J Remote Sens* **4**, 0132 (2024).
 42. Liu XQ, Feng YH, Hu TY et al. Enhancing ecosystem productivity and stability with increasing canopy structural complexity in global forests. *Sci Adv* **10**, ead1947 (2024).
 43. Wang ZY, Liu J, Sheng YH et al. Intercomparison of the DART model and GEDI simulator for simulating GEDI waveforms in forests. *Int J Appl Earth Obs Geoinf* **134**, 104148 (2024).
 44. Cardenas-Martinez A, Pascual A, Guisado-Pintado E et al. Using airborne LiDAR and enhanced-geolocated GEDI metrics to map structural traits over a Mediterranean forest. *Sci Remote Sens* **11**, 100195 (2025).
 45. Sun G, Ranson KJ. Modeling lidar returns from forest canopies. *IEEE Trans Geosci Remote Sens* **38**, 2617–2626 (2000).
 46. Li Y, Yan GJ, Li WH et al. Accurate spaceborne waveform simulation in heterogeneous forests using small-footprint airborne LiDAR point clouds. *ISPRS J Photogramm Remote Sens* **220**, 246–263 (2025).
 47. Ni-Meister W, Jupp DLB, Dubayah R. Modeling lidar waveforms in heterogeneous and discrete canopies. *IEEE Trans Geosci Remote Sens* **39**, 1943–1958 (2001).
 48. Yang XB, Wang YJ, Yin TG et al. Comprehensive LiDAR simulation with efficient physically-based DART-Lux model (I): theory, novelty, and consistency validation. *Remote Sens Environ* **272**, 112952 (2022).
 49. Yang XB, Wang C, Yin TG et al. Comprehensive LiDAR simulation with efficient physically-based DART-Lux model (II): validation with GEDI and ICESat-2 measurements at natural and urban landscapes. *Remote Sens Environ* **317**, 114519 (2025).
 50. Yang XB, Wang C, Xi XH et al. Assessment of multiple scattering in LiDAR canopy waveform. *IEEE J Sel Top Appl Earth Obs Remote Sens* **17**, 14394–14407 (2024).
 51. Bye IJ, North PRJ, Los SO et al. Estimating forest canopy parameters from satellite waveform LiDAR by inversion of the FLIGHT three-dimensional radiative transfer model. *Remote Sens Environ* **188**, 177–189 (2017).
 52. Winiwarter L, Pena AME, Weiser H et al. Virtual laser scanning with HELIOS++: a novel take on ray tracing-based simulation of topographic full-waveform 3D laser scanning. *Remote Sens Environ* **269**, 112772 (2022).
 53. Luo YT, Xie DH, Qi JB et al. LESS LiDAR: a full-waveform and discrete-return multispectral LiDAR simulator based on ray tracing algorithm. *Remote Sens* **15**, 4529 (2023).
 54. Li X, Li LY, Ni WJ, et al. Validating GEDI tree canopy cover product across forest types using co-registered aerial LiDAR data. *ISPRS J Photogramm Remote Sens* **207**, 326–337 (2024).
 55. Leite RV, Silva CA, Broadbent EN et al. Large scale multi-layer fuel load characterization in tropical savanna using GEDI spaceborne lidar data. *Remote Sens Environ* **268**, 112764 (2022).
 56. Zhao X, Qi JB, Xu HF et al. Evaluating the potential of airborne hyperspectral LiDAR for assessing forest insects and diseases with 3D Radiative Transfer Modeling. *Remote Sens Environ* **297**, 113759 (2023).
 57. Mkaouar A, Shean D, Yin TY et al. Refining satellite laser altimetry geolocation through full-waveform radiative transfer modeling and matching. *Sci Remote Sens* **12**, 100248 (2025).
 58. Wang YX, Zhang XY, Xu SL et al. Modulation of typical three-dimensional targets on the echo waveform using analytical formula. *Remote Sens* **17**, 3419 (2025).
 59. Hofton MA, Minster JB, Blair JB. Decomposition of laser altimeter waveforms. *IEEE Trans Geosci Remote Sens* **38**, 1989–1996 (2000).
 60. Wagner W, Ullrich A, Ducic V et al. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS J Photogramm Remote Sens* **60**, 100–112 (2006).
 61. Mallet C, Bretar F. Full-waveform topographic lidar: state-of-the-art. *ISPRS J Photogramm Remote Sens* **64**, 1–16 (2009).
 62. Duncanson LI, Niemann KO, Wulder MA. Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sens Environ* **114**, 138–154 (2010).
 63. Jutzi B, Stilla U. Range determination with waveform recording laser systems using a Wiener Filter. *ISPRS J Photogramm Remote Sens* **61**, 95–107 (2006).
 64. Qiu PR, Zhou H, Zhang QY et al. Modeling returned signal of a satellite laser altimeter over an urban area. *Appl Opt* **61**, 9888–9897 (2022).
 65. Li HP, Li GY, Cai ZJ et al. Full-waveform LiDAR echo decomposition method. *J Remote Sens* **23**, 89–98 (2019).
 66. Xie JF, Yang CC, Mei YK et al. Full waveform decomposition of

- spaceborne laser based on genetic algorithm. *Infrared Laser Eng* **49**, 20200245 (2020).
67. Lai XD, Yuan YF, Xu JZ et al. LiDAR waveform decomposition based on modified differential evolution algorithm. *J Infrared Millim Waves* **40**, 381–390 (2021).
 68. Zhou GQ, Deng RH, Zhou X et al. Gaussian inflection point selection for LiDAR hidden echo signal decomposition. *IEEE Geosci Remote Sens Lett* **19**, 6502705 (2022).
 69. Zhang FX, Wang X, Wang LG et al. A satellite full-waveform laser decomposition method for forested areas based on hidden peak detection and adaptive genetic optimization. *Remote Sens* **17**, 701 (2025).
 70. Zhang ZJ, Xie H, Tong XH et al. A combined deconvolution and Gaussian decomposition approach for overlapped peak position extraction from large-footprint satellite laser altimeter waveforms. *IEEE J Sel Top Appl Earth Obs Remote Sens* **13**, 2286–2303 (2020).
 71. Liu C, Li XL, Huang JB et al. B-spline based progressive decomposition of LiDAR waveform with low SNR. *IEEE Trans Instrum Meas* **71**, 8501812 (2022).
 72. Zhou GQ, Long SH, Xu JS et al. Comparison analysis of five waveform decomposition algorithms for the airborne LiDAR echo signal. *IEEE J Sel Top Appl Earth Obs Remote Sens* **14**, 7869–7880 (2021).
 73. Hakala T, Suomalainen J, Kaasalainen S et al. Full waveform hyperspectral LiDAR for terrestrial laser scanning. *Opt Express* **20**, 7119–7127 (2012).
 74. Song SL, Wang BH, Gong W et al. A new waveform decomposition method for multispectral LiDAR. *ISPRS J Photogramm Remote Sens* **149**, 40–49 (2019).
 75. Xia YH, Xu SL, Fang JJ et al. A novel waveform decomposition and spectral extraction method for 101-channel hyperspectral LiDAR. *Remote Sens* **14**, 5285 (2022).
 76. Wang BH, Song SL, Shi S et al. Multichannel interconnection decomposition for hyperspectral LiDAR waveforms detected from over 500 m. *IEEE Trans Geosci Remote Sens* **60**, 5515714 (2022).
 77. Xia YH, Xu SL, Shao H et al. Range resolution enhanced method with spectral properties for hyperspectral LiDAR. *IEEE Trans Geosci Remote Sens* **61**, 5703517 (2023).
 78. Bai J, Niu Z, Huang YR et al. Full-waveform hyperspectral LiDAR data decomposition via ranking central locations of natural target echoes (Rclonte) at different wavelengths. *Remote Sens Environ* **310**, 114227 (2024).
 79. Bai J, Niu Z, Bi KY et al. Toward an advanced method for full-waveform hyperspectral LiDAR data processing. *IEEE Trans Geosci Remote Sens* **62**, 5702516 (2024).
 80. Shi S, Gong CY, Xu Q et al. Waveform information accurate extraction for massive and complex waveform data of hyperspectral lidar. *IEEE J Sel Top Appl Earth Obs Remote Sens* **18**, 1020–1038 (2025).
 81. Liu XS, Guo YL, Xie H et al. Fast machine learning-based extraction of the peak number from ICESat full-waveform data. *Remote Sens Lett* **12**, 1004–1014 (2021).
 82. Liu GP, Ke J. Full-waveform LiDAR echo decomposition based on dense and residual neural networks. *Appl Opt* **61**, F15–F24 (2022).
 83. Liu GP, Ke J. End-to-end full-waveform echo decomposition based on self-attention classification and U-net decomposition. *IEEE J Sel Top Appl Earth Obs Remote Sens* **15**, 7978–7987 (2022).
 84. Xu XB, Wang JL, Wu JL et al. Full-waveform LiDAR echo decomposition method based on deep learning and sparrow search algorithm. *Infrared Phys Technol* **130**, 104613 (2023).
 85. Ji X, Yang BS, Wang Y et al. Full-waveform classification and segmentation-based signal detection of single-wavelength bathymetric LiDAR. *IEEE Trans Geosci Remote Sens* **60**, 4208714 (2022).
 86. Koetz B, Morsdorf F, Sun G et al. Inversion of a lidar waveform model for forest biophysical parameter estimation. *IEEE Geosci Remote Sens Lett* **3**, 49–53 (2006).
 87. Koetz B, Sun GQ, Morsdorf F et al. Fusion of imaging spectrometer and LIDAR data over combined radiative transfer models for forest canopy characterization. *Remote Sens Environ* **106**, 449–459 (2007).
 88. Hmida SB, Kallel A, Gastellu-Etchegorry JP et al. Crop biophysical properties estimation based on LiDAR full-waveform inversion using the DART RTM. *IEEE J Sel Top Appl Earth Obs Remote Sens* **10**, 4853–4868 (2017).
 89. Markus T, Neumann T, Martino A et al. The ice, cloud, and land elevation Satellite-2 (ICESat-2): science requirements, concept, and implementation. *Remote Sens Environ* **190**, 260–273 (2017).
 90. Wang YX, Fang JJ, Xu SL et al. Intelligent shape discrimination and multi-dimensional feature parameter inversion of target based on laser echo waveform. *Chin J Laser* **53**, 0210003 (2026).
 91. Yang XB, Wang C, Pan FF, et al. Retrieving leaf area index in discontinuous forest using ICESat/GLAS full-waveform data based on gap fraction model. *ISPRS J Photogramm Remote Sens* **148**, 54–62 (2019).
 92. Cui L, Jiao ZT, Zhao KG et al. Retrieving forest canopy elements clumping index using ICESat GLAS lidar data. *Remote Sens* **13**, 948 (2021).
 93. Chen S, Gong M, Sun H et al. Estimation of forest canopy height from spaceborne full-waveform LiDAR data using a bisection approximation decomposition method. *Forests* **16**, 145 (2025).
 94. Wang Y, Fang HL. Estimation of LAI with the LiDAR technology: a review. *Remote Sens* **12**, 3457 (2020).
 95. Jiang HL, Yan GJ, Tong YY et al. Correcting crown-level clumping effect for improving leaf area index retrieval from large-footprint LiDAR: a study based on the simulated waveform and GLAS data. *IEEE J Sel Top Appl Earth Obs Remote Sens* **14**, 12386–12402 (2021).
 96. Wang CJ, Jia D, Lei SG et al. Accuracy assessment and impact factor analysis of GEDI leaf area index product in temperate forest. *Remote Sens* **15**, 1535 (2023).
 97. Liang LJ, Shang R, Chen JM et al. Improved estimation of the underestimated GEDI footprint LAI in dense forests. *Geo-spat Inf Sci* **28**, 49–64 (2025).
 98. Tang H, Armston J, Hancock S et al. Characterizing global forest canopy cover distribution using spaceborne lidar. *Remote Sens Environ* **231**, 111262 (2019).
 99. Li Y, Fang HL, Wang Y et al. Validation of the vertical canopy cover profile products derived from GEDI over selected forest sites. *Sci Remote Sens* **10**, 100158 (2024).
 100. Zhuang W, Mountrakis G, Wiley Jr JJ et al. Estimation of above-ground forest biomass using metrics based on Gaussian decomposition of waveform lidar data. *Int J Remote Sens* **36**, 1871–1889 (2015).
 101. Luo SZ, Wang C, Xi XH et al. Estimating forest aboveground biomass using small-footprint full-waveform airborne LiDAR data. *Int J Appl Earth Obs Geoinf* **83**, 101922 (2019).
 102. Yang FL, Qi C, Su DP et al. An airborne LiDAR bathymetric waveform decomposition method in very shallow water: a case study around Yuanzhi Island in the South China Sea. *Int J Appl Earth Obs Geoinf* **109**, 102788 (2022).
 103. Qi C, Ma Y, Su DP et al. A method to decompose airborne LiDAR bathymetric waveform in very shallow waters combining deconvolution with curve fitting. *IEEE Geosci Remote Sens Lett* **19**, 7004905 (2022).
 104. Wu L, Chen YF, Le Y et al. A high-precision fusion bathymetry of

- multi-channel waveform curvature for bathymetric LiDAR systems. *Int J Appl Earth Obs Geoinf* **128**, 103770 (2024).
105. Chen YF, Wu L, Qian Y et al. A novel strategy of full-waveform light detection and ranging bathymetry based on spatial gram angle difference field conversion and deep-learning network architecture. *Remote Sens Environ* **318**, 114615 (2025).
 106. Guo YD, Xu WX, Liu YX et al. A pregeneration–recognition method of detecting weak seafloor echoes for full-waveform airborne LiDAR bathymetry. *Int J Appl Earth Obs Geoinf* **143**, 104798 (2025).
 107. Li N, Truong ML, Schwarz R et al. Deep learning assisted exponential waveform decomposition for bathymetric LiDAR. *Int Arch Photogramm Remote Sens Spatial Inf Sci XLVIII-2-2024*, 195–202 (2024).
 108. Bruggisser M, Roncat A, Schaepman ME et al. Retrieval of higher order statistical moments from full-waveform LiDAR data for tree species classification. *Remote Sens Environ* **196**, 28–41 (2017).
 109. Xu SL, Hu YH, Dong X et al. Solar panel scale inversion with higher-order statistical moments of echo laser pulse waveform. *Opt Eng* **61**, 026102 (2022).
 110. Liu XF, Liu XD, Wang ZH et al. Classification of laser footprint based on random forest in mountainous area using GLAS full-waveform features. *IEEE J Sel Top Appl Earth Obs Remote Sens* **15**, 2284–2297 (2022).
 111. Ran YH, Song SL, Hou XX et al. Multi-echo hyperspectral reflectance extraction method based on full waveform hyperspectral LiDAR. *ISPRS J Photogramm Remote Sens* **207**, 43–56 (2024).
 112. Coops NC, Tompalski P, Goodbody TRH et al. Modelling lidar-derived estimates of forest attributes over space and time: a review of approaches and future trends. *Remote Sens Environ* **260**, 112477 (2021).
 113. Lefsky MA, Harding DJ, Keller M et al. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophys Res Lett* **32**, L22S02 (2005).
 114. Lefsky MA, Keller M, Pang Y et al. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *J Appl Remote Sens* **1**, 013537 (2007).
 115. Lee S, Ni-Meister W, Yang WZ et al. Physically based vertical vegetation structure retrieval from ICESat data: validation using LVIS in white mountain national forest, New Hampshire, USA. *Remote Sens Environ* **115**, 2776–2785 (2011).
 116. Xing YQ, de Gier A, Zhang JJ et al. An improved method for estimating forest canopy height using ICESat-GLAS full waveform data over sloping terrain: a case study in Changbai mountains, China. *Int J Appl Earth Obs Geoinf* **12**, 385–392 (2010).
 117. Liang XJ, Pang Y, Wang CJ. Optimization of TECIS LiDAR waveform metrics for forest height estimation. *Int J Appl Earth Obs Geoinf* **142**, 104700 (2025).
 118. Wang Y, Ni WJ, Sun GQ et al. Slope-adaptive waveform metrics of large footprint lidar for estimation of forest aboveground biomass. *Remote Sens Environ* **224**, 386–400 (2019).
 119. de Conto T, Armston J, Dubayah R. Characterizing the structural complexity of the earth's forests with spaceborne lidar. *Nat Commun* **15**, 8116 (2024).
 120. Lin SM, Li LY, Liu SB et al. Stratified burn severity assessment by integrating spaceborne spectral and waveform attributes in Great Xing'an Mountain. *Remote Sens Environ* **307**, 114152 (2024).
 121. Aragonese E, García M, Tang H et al. A multi-sensor approach allows confident mapping of forest canopy fuel load and canopy bulk density to assess wildfire risk at the European scale. *Remote Sens Environ* **318**, 114578 (2025).
 122. Adrah E, Wong JP, Yin H. Integrating GEDI, Sentinel-2, and Sentinel-1 imagery for tree crops mapping. *Remote Sens Environ* **319**, 114644 (2025).
 123. Simard M, Pinto N, Fisher JB et al. Mapping forest canopy height globally with spaceborne lidar. *J Geophys Res Biogeosci* **116**, G04021 (2011).
 124. Liu LJ, Wang C, Nie S et al. Analysis of the influence of different algorithms of GEDI L2A on the accuracy of ground elevation and forest canopy height. *J Univ Chin Acad Sci* **39**, 502–511 (2022).
 125. Han MH, Xing YQ, Li GY et al. Comparison of the accuracy of the maximum canopy height and biomass inversion of the data of different GEDI algorithm groups. *J Central South Univ For Technol* **42**, 72–82 (2022).
 126. Min WK, Chen YM, Huang WL et al. Incorporating of spatial effects in forest canopy height mapping using airborne, spaceborne lidar and spatial continuous remote sensing data. *Int J Appl Earth Obs Geoinf* **133**, 104123 (2024).
 127. Liu XQ, Su YJ, Hu TY et al. Neural network guided interpolation for mapping canopy height of China's forests by integrating GEDI and ICESat-2 data. *Remote Sens Environ* **269**, 112844 (2022).
 128. Ma X, Zheng G, Chi X et al. Mapping fine-scale building heights in urban agglomeration with spaceborne lidar. *Remote Sens Environ* **285**, 113392 (2023).
 129. Tang HL, Huang HB, Zheng Y et al. Improved GEDI canopy height extraction based on a simulated ground echo in topographically undulating areas. *IEEE Trans Geosci Remote Sens* **61**, 5705915 (2023).
 130. Wang RQ, Lu YG, Lu DS et al. Improving extraction of forest canopy height through reprocessing ICESat-2 ATLAS and GEDI data in sparsely forested plain regions. *GISci Remote Sens* **61**, 2396807 (2024).
 131. Cho MS, Roy DP, Kashongwe HB et al. Assessment and improvement of GEDI canopy height estimation in tropical and temperate forests. *Sci Remote Sens* **11**, 100221 (2025).
 132. Zhou H, Wang H, Song H et al. Canopy height extraction over mountainous areas from GEDI Lidar deconvoluted waveforms. *IEEE Geosci Remote Sens Lett* **21**, 2502405 (2024).
 133. Kaya Y. Automated estimation of building heights with ICESat-2 and GEDI LiDAR altimeter and building footprints: the case of New York City and Los Angeles. *Buildings* **14**, 3571 (2024).
 134. Dubayah RO, Drake JB. Lidar remote sensing for forestry. *J For* **98**, 44–46 (2000).
 135. Drake JB, Dubayah RO, Clark DB et al. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sens Environ* **79**, 305–319 (2002).
 136. Kellner JR, Armston J, Duncanson L. Algorithm theoretical basis document for GEDI footprint aboveground biomass density. *Earth Space Sci* **10**, e2022EA002516 (2023).
 137. Nie S, Wang C, Zeng HC et al. Above-ground biomass estimation using airborne discrete-return and full-waveform LiDAR data in a coniferous forest. *Ecol Indic* **78**, 221–228 (2017).
 138. Ni WJ, Zhang ZY, Sun GQ. Assessment of slope-adaptive metrics of GEDI waveforms for estimations of forest aboveground biomass over mountainous areas. *J Remote Sens* **2021**, 9805364 (2021).
 139. Ni-Meister W, Rojas A, Lee S. Direct use of large-footprint lidar waveforms to estimate aboveground biomass. *Remote Sens Environ* **280**, 113147 (2022).
 140. Li XX, Wessels K, Armston J et al. Evaluation of GEDI footprint level biomass models in Southern African Savannas using airborne LiDAR and field measurements. *Sci Remote Sens* **10**, 100161 (2024).
 141. Zhang B, Wang ZC, Ma TT et al. Correcting forest aboveground biomass biases by incorporating independent canopy height retrieval with conventional machine learning models using GEDI

- and ICESat-2 data. *Ecol Inf* **86**, 103045 (2025).
142. Yang HF, Qin Z, Shu QT et al. Estimation of above-ground biomass for *Dendrocalamus Giganteus* utilizing spaceborne LiDAR GEDI data. *IEEE J Sel Top Appl Earth Obs Remote Sens* **18**, 5271–5286 (2025).
 143. Huang JP, Piao RX, Liang LH. Estimation of canopy coverage based on multisource remote sensing data. *IEEE J Sel Top Appl Earth Obs Remote Sens* **18**, 21342 (2025).
 144. Wu QH, Lee CKF, Wang JA et al. Improved assessment of post-fire recovery trajectory of forests in Amazon's protected areas. *Remote Sens Environ* **326**, 114802 (2025).
 145. Dubayah R, Armston J, Healey SP et al. GEDI launches a new era of biomass inference from space. *Environ Res Lett* **17**, 095001 (2022).
 146. Qi WL, Dubayah RO. Combining Tandem-X InSAR and simulated GEDI lidar observations for forest structure mapping. *Remote Sens Environ* **187**, 253–266 (2016).
 147. Lang N, Kalischek N, Armston J et al. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. *Remote Sens Environ* **268**, 112760 (2022).
 148. Lang N, Jetz W, Schindler K et al. A high-resolution canopy height model of the earth. *Nat Ecol Evol* **7**, 1778–1789 (2023).
 149. Wang SF, Liu C, Li WY et al. Hybrid model for estimating forest canopy heights using fused multimodal spaceborne LiDAR data and optical imagery. *Int J Appl Earth Obs Geoinf* **122**, 103431 (2023).
 150. Liu AB, Chen YT, Cheng X. Evaluating ICESat-2 and GEDI with integrated landsat-8 and PALSAR-2 for mapping tropical forest canopy height. *Remote Sens* **16**, 3798 (2024).
 151. Liu SQ, Csillik O, Ordway EM et al. Environmental drivers of spatial variation in tropical forest canopy height: insights from NASA's GEDI spaceborne LiDAR. *Proc Natl Acad Sci USA* **122**, e2401755122 (2025).
 152. Qi WL, Armston J, Choi C et al. Mapping large-scale pantropical forest canopy height by integrating GEDI lidar and TanDEM-X InSAR data. *Remote Sens Environ* **318**, 114534 (2025).
 153. Xiao YL, Wang QM, Xi HP et al. A 30 m canopy height map in China created by fusion of multiple relative height metrics. *IEEE Trans Geosci Remote Sens* **63**, 4411116 (2025).
 154. Li YS, Lu DS, Lu YG et al. Examining the impact of topography and vegetation on existing forest canopy height products from ICESat-2 ATLAS/GEDI data. *Remote Sens* **16**, 3650 (2024).
 155. Huang JP, Cao XY. Assessment and optimization of forest above-ground biomass in Liaoning Province. *Forest* **15**, 2095 (2024).
 156. May PB, Schlund M, Armston J et al. Mapping aboveground biomass in Indonesian lowland forests using GEDI and hierarchical models. *Remote Sens Environ* **313**, 114384 (2024).
 157. Ni-Meister W, Rojas A, Grant I. Enhancing aboveground biomass density estimates in tropical forests using GEDI waveform data. *Sci Remote Sens* **12**, 100286 (2025).
 158. Qi WL, Saarela S, Armston J et al. Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. *Remote Sens Environ* **232**, 111283 (2019).
 159. Shendryk Y. Fusing GEDI with earth observation data for large area aboveground biomass mapping. *Int J Appl Earth Obs Geoinf* **115**, 103108 (2022).
 160. Nian YY, Chen SW, Chen J et al. Mapping subalpine forest above-ground biomass in Qilian Mountain national park using UAV-LiDAR, GEDI, and multisource satellite data. *IEEE J Sel Top Appl Earth Obs Remote Sens* **18**, 12407–12420 (2025).
 161. Bourgoin C, Ceccherini G, Girardello M et al. Human degradation of tropical moist forests is greater than previously estimated. *Nature* **631**, 570–576 (2024).
 162. Zhang B, Zhang L, Yan M et al. High-resolution mapping of forest parameters in tropical rainforests through AutoML integration of GEDI with Sentinel-1/2, Landsat 8, and ALOS-2 data. *IEEE J Sel Top Appl Earth Obs Remote Sens* **18**, 9084–9118 (2025).
 163. Liang MY, Duncanson L, Silva JA et al. Quantifying aboveground biomass dynamics from charcoal degradation in Mozambique using GEDI Lidar and Landsat. *Remote Sens Environ* **284**, 113367 (2023).
 164. Atzberger C, Immitzer M, Hemes KS et al. A scalable, annual aboveground biomass product for monitoring carbon impacts of ecosystem restoration projects. *Remote Sens Environ* **327**, 114774 (2025).
 165. Schwartz M, Ciais P, Sean E et al. Retrieving yearly forest growth from satellite data: a deep learning based approach. *Remote Sens Environ* **330**, 114959 (2025).
 166. Chen PM, Huang HB, Qin P et al. Characterizing dynamics of built-up height in China from 2005 to 2020 based on GEDI, Landsat, and PALSAR data. *Remote Sens Environ* **325**, 114776 (2025).
 167. Lin XD, Shang R, Chen JM et al. High-resolution forest age mapping based on forest height maps derived from GEDI and ICESat-2 space-borne lidar data. *Agric For Meteorol* **339**, 109592 (2023).
 168. Zhao XM, Chen JM, Zhang YG et al. Global mapping of forest clumping index based on GEDI canopy height and complementary data. *ISPRS J Photogramm Remote Sens* **209**, 1–16 (2024).
 169. Bu JY, Xiao JF. Upscaling eddy covariance measurements of carbon and water fluxes to the continental scale by incorporating GEDI-derived canopy structural complexity metrics. *Remote Sens Environ* **329**, 114930 (2025).
 170. Magruder LA, Farrell SL, Neuenschwander A et al. Monitoring Earth's climate variables with satellite laser altimetry. *Nat Rev Earth Environ* **5**, 120–136 (2024).
 171. Wang JA, Baccini A, Farina M et al. Disturbance suppresses the aboveground carbon sink in North American boreal forests. *Nat Climate Change* **11**, 435–441 (2021).
 172. Kalischek N, Lang N, Renier C et al. Cocoa plantations are associated with deforestation in Côte d'Ivoire and Ghana. *Nat Food* **4**, 384–393 (2023).
 173. Liu SY, Brandt M, Nord-Larsen T et al. The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe. *Sci Adv* **9**, eadh4097 (2023).
 174. Xie YY, Shen ZH, Camarero JJ et al. Keys to the global treeline formation: thermal limit for its position and moisture for the taxon-specific variation. *Proc Natl Acad Sci USA* **122**, e2504685122 (2025).
 175. Xu JF, Wang T, Wang XY et al. Late Quaternary fluctuation in upper range limit of trees shapes endemic flora diversity on the Tibetan Plateau. *Nat Commun* **16**, 1819 (2025).
 176. Duncanson L, Liang M, Leitold V et al. The effectiveness of global protected areas for climate change mitigation. *Nat Commun* **14**, 2908 (2023).
 177. Reiner F, Brandt M, Tong XY et al. More than one quarter of Africa's tree cover is found outside areas previously classified as forest. *Nat Commun* **14**, 2258 (2023).
 178. Li W, Guo WY, Pasgaard M et al. Human fingerprint on structural density of forests globally. *Nat Sustain* **6**, 368–379 (2023).
 179. Brodie JF, Mohd-Azlan J, Chen C et al. Landscape-scale benefits of protected areas for tropical biodiversity. *Nature* **620**, 807–812 (2023).
 180. Liu JG, Wang QR, Zhan WW et al. When and where soil dryness matters to ecosystem photosynthesis. *Nat Plants* **11**, 1390–1400 (2025).
 181. Forkel M, Wessollek C, Huijnen V et al. Burning of woody debris dominates fire emissions in the Amazon and Cerrado. *Nat Geosci*

- 18, 140–147 (2025).
182. Brandt M, Chave J, Li SZ et al. High-resolution sensors and deep learning models for tree resource monitoring. *Nat Rev Electr Eng* 2, 13–26 (2025).
 183. Steinvall O, Chevalier T, Grönwall C. Simulation and modeling of laser range profiling and imaging of small surface vessels. *Opt Eng* 53, 013109 (2014).
 184. Steinvall O, Tuldahl M. Laser range profiling for small target recognition. *Opt Eng* 56, 031206 (2016).
 185. Li ZL, Zhai DS, Tang RF et al. Research and experiment on space debris daytime laser ranging based on 532 nm wavelength. *Laser Optoelectron Prog* 59, 1112003 (2022).
 186. Yang Y, Long ML, Zhang HF et al. A study of watt-scale picosecond laser space target ranging for a 1.2 m telescope at high altitude. *J Infrared Millim Waves* 43, 696–702 (2024).
 187. Guo YL, Hu YH, Fang JJ et al. Research status and development trend of LiDAR super-resolution imaging technology. *J Radars* 14, 501–527 (2025).
 188. Parker JK, Craig EB, Klick DI et al. Reflective tomography: images from range-resolved laser radar measurements. *Appl Opt* 27, 2642–2643 (1988).
 189. Matson CL, Mosley DE. Reflective tomography reconstruction of satellite features—field results. *Appl Opt* 40, 2290–2296 (2001).
 190. Murray J, Triscari J, Fetzer G et al. Tomographic Lidar. In *Lasers, Sources and Related Photonic Devices* (Optica, 2010). <https://doi.org/10.1364/LSC.2010.LSWA1>
 191. Henriksson M, Olofsson T, Grönwall C et al. Optical reflectance tomography using TCSPC laser radar. *Proc SPIE* 8542, 85420E (2012).
 192. Zhang XY, Shan LQ, Fang JJ et al. Super-resolution imaging using reflective tomography LiDAR over 10 km. *Adv Imaging* 2, 051002 (2025).
 193. Li ZP, Ye JT, Ye WL et al. Single-photon laser reflective tomography over 7 km. *Opt Lett* 50, 2970–2973 (2025).
 194. Guo YH, Pu MB, Li Y et al. Surpassing the diffraction limit in long-range laser engineering via cross-scale vectorial optical field manipulation: perspectives and outlooks. *Opto-Electron Adv* 8, 250256 (2025).
 195. Guo HS, Li XY, Wang ZW et al. Sub-diffraction limited imaging at 3.2km: SNSPD-based single-photon laser reflective tomography. *Proc SPIE* 13992, 139923F (2026).
 196. Deng YH, Gu YC, Liu HL et al. Gaussian boson sampling with pseudo-photon-number-resolving detectors and quantum computational advantage. *Phys Rev Lett* 131, 150601 (2023).
 197. Gao DX, Fan DJ, Zha C et al. Establishing a new benchmark in quantum computational advantage with 105-qubit Zuchongzhi 3.0 processor. *Phys Rev Lett* 134, 090601 (2025).
 198. Hughes TJR, Cottrell JA, Bazilevs Y. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Comput Methods Appl Mech Eng* 194, 4135–4195 (2005).
 199. Podlozhnyuk A, Pirkers S, Kloss C. Efficient implementation of superquadric particles in Discrete Element Method within an open-source framework. *Comput Part Mech* 4, 101–118 (2017).

Acknowledgements

The work was supported by the National University of Defense Technology (JS25-08 and 24-ZZCX-JDZ-43), and the National Natural Science Foundation of China (61871389). Research program for the college outstanding youth of Anhui Province (Grand No.2022AH020099).

Author contributions

Jiajie Fang, Xinyuan Zhang, and Yihua Hu conceptualized the original idea and supervised the project, Jiajie Fang and Xinyuan Zhang drafted the manuscript, Yongxiang Wang created the figures and participated in literature review, Liang Shi contributed to discussions on the paper's framework and wrote parts concerning target feature inversion, Xurong Duan conducted the literature review and compilation, Shilong Xu and Yuhao Xia contributed to writing the hyperspectral-related section, Nanxiang Zhao participated in the analysis of target waveform modulation mechanisms and inversion based on analytical formula, and Xiao Dong contributed to writing the space surveillance section.

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information Supplementary information for this paper is available at <https://doi.org/10.29026/oep.2026.260001>



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> ©The Author(s) 2026.