scientific data

Check for updates

OPEN

DATA DESCRIPTOR

# NEON-SD: A 30-m Structural Diversity Product Derived from the NEON Discrete-Return LiDAR Point Cloud

Jianmin Wang[1,9], Dennis H. Choi[1,9], Elizabeth LaRue[2], Jeff W. Atkins[3], Jane R. Foster[4,5], Jaclyn H. Matthes[6], Robert T. Fahey[7], Songlin Fei[1] & Brady S. Hardiman[1,8]✉

**Structural diversity (SD) characterizes the volume and physical arrangement of biotic components in an ecosystem which control critical ecosystem functions and processes. LiDAR data provides detailed 3-D spatial position information of components and has been widely used to calculate SD. However, the intensive computation of SD metrics from extensive LiDAR datasets is time-consuming and challenging for researchers who lack access to high-performance computing resources. Moreover, a lack of understanding of LiDAR data and algorithms could lead to inconsistent SD metrics. Here, we developed a SD product using the Discrete-Return LiDAR Point Cloud from the NEON Aerial Observation Platform. This product provides SD metrics detailing height, density, openness, and complexity at a spatial resolution of 30 m, aligned to the Landsat grids, for 211 site-years for 45 Terrestrial NEON sites from 2013 to 2022. To accommodate various ecosystems with different understory heights, it includes three different cut-off heights (0.5 m, 2 m, and 5 m). This structural diversity product can enable various applications such as ecosystem productivity estimation and disturbance monitoring.**

## Background & Summary

Three-dimensional (3-D) structural diversity (SD) characterizes the volumetric capacity and physical arrangement of biotic components in an ecosystem[1], and is associated with critical ecosystem functions such as forest productivity, species diversity, and forest health and disturbance[2–8]. Light detection and ranging (LiDAR) data provide detailed 3-D spatial position information on components of vegetation canopy and have been increasingly used to estimate SD. LiDAR-derived SD generally encompasses five categories of metrics that are canopy height, canopy density, openness, and interior and exterior complexity[5,9].

The LiDAR data from the National Ecological Observatory Network (NEON) Airborne Observation Platform (AOP)[10], publicly available for free, is widely used to estimate SD[6,8]. The NEON program monitors a variety of ecosystems across the United States and provides harmonized data from both ground surveys and remote sensing data since 2013 (https://www.neonscience.org/). Although individual studies have computed SD metrics at various NEON sites, variation in the workflows and parameter setup, such as cut-off height (a height threshold below which LiDAR returns are excluded from SD calculation), could limit their generalizability across ecosystems. Moreover, the intensive computation of SD metrics from extensive LiDAR datasets is time-consuming and challenging for researchers who lack access to high-performance computing resources. Given the rapid growth of NEON-based macrosystem research[1,6,8,11], there is a clear need for a comprehensive SD product derived from a consistent workflow and methodology across all the NEON sites.

[1]Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana, USA. [2]Department of Biological Sciences, The University of Texas at El Paso, El Paso, Texas, USA. [3]USDA Forest Service, Southern Research Station, New Ellenton, South Carolina, USA. [4]USDA Forest Service, Southern Research Station, Tennessee, Knoxville, USA. [5]Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, Vermont, USA. [6]Harvard Forest, Harvard University, Petersham, Massachusetts, USA. [7]Department of Natural Resources and the Environment and Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, Connecticut, USA. [8]Department of Environmental and Ecological Engineering, Purdue University, West Lafayette, Indiana, USA. [9]These authors contributed equally: Jianmin Wang, Dennis H. Choi. ✉e-mail: hardimanb@purdue.edu

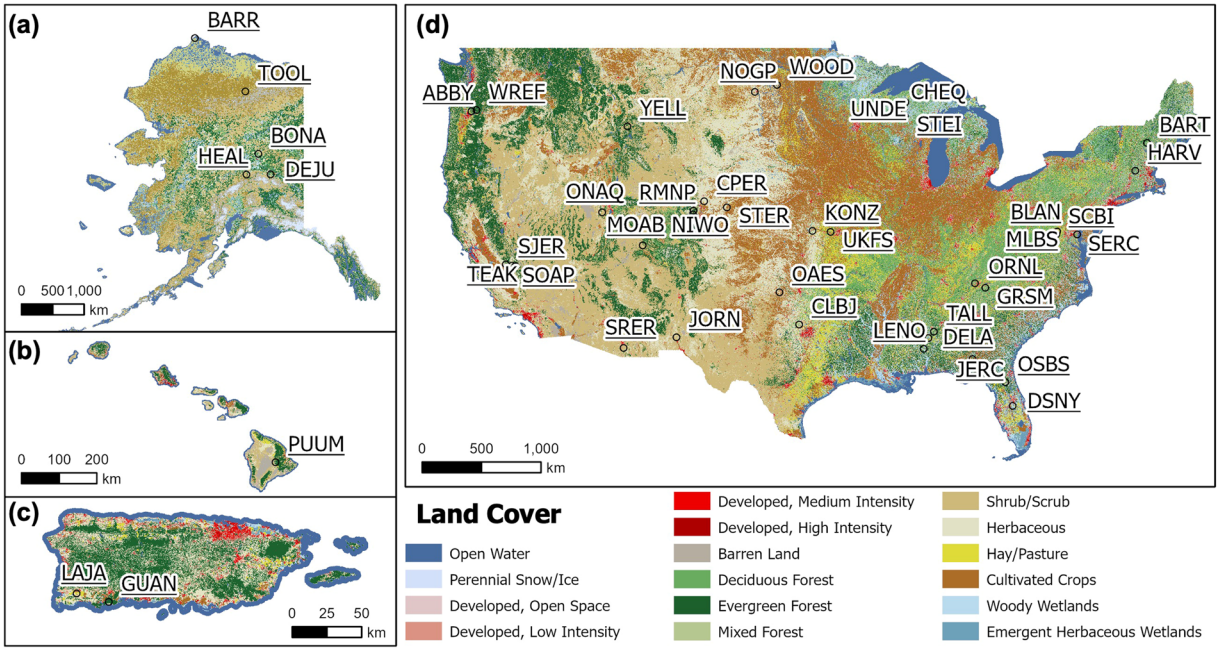


**Fig. 1** NEON field survey sites across the United States ((**a**): Alaska, (**b**): Hawaii, (**c**): Puerto Rico, (**d**): Contiguous US). Colors indicate land cover classes from National Land Cover Database.

To address this gap, we present a novel SD product, named NEON-SD, covering NEON's 45 terrestrial sites (Fig. 1) from 2013 to 2022, derived from the NEON AOP LiDAR data. NEON-SD delivers SD metrics of five categories (*i.e.*, height, density, openness, exterior complexity, and interior complexity) with three commonly used cut-off heights (0.5 m, 2 m, and 5 m) for 30-m pixels that are aligned with Landsat World Reference System-2 (WRS-2) grids. This product was developed to facilitate a broad spectrum of ecological research and applications, such as ecosystem productivity prediction and disturbance monitoring[5,12]. It provides the potential end users with access to SD metrics across NEON sites, eliminating the need for individual large data processing while providing flexibility to accommodate specific study requirements.

## Methods

**Datasets.** The NEON AOP Discrete return LiDAR point cloud product (DP1.30003.001; accessed in March 2023) was used to generate this product[10]. It provides 1 × 1 km pre-processed LiDAR data tiles, including coordinates (X, Y, Z), classification, and intensity, for each of the 45 NEON sites. AOP survey has been conducted for these sites since 2013, with the detailed site information and acquisition dates available in Tables S1 and S2. The AOP has employed three different commercial LiDAR systems: Optech Gemini (since 2013), Riegl LMS-Q780 (introduced in 2018 for some sites), and Optech Galaxy (introduced in 2021 for some sites) as detailed in Table S2. A key difference among the three sensors is the pulse repetition frequency (PRF), which is the rate at which the laser transmits and receives pulses. The Optech Gemini has a maximum PRF of 100 kHz, the Riegl Q780 can operate at up to 400 kHz, and the Optech Galaxy operates at up to 1000 kHz. The differences in PRF led to varying point densities among acquisitions with an average point density of ~5 points/m² for Gemini, 13 points/m² for Riegl, and 30 points/m² for Galaxy, respectively (Table S2). Typically, the LiDAR sensors are flown at an altitude of 1,000 m above ground level and the adjacent flight lines overlap by 30%[10], resulting in higher point density in overlapped areas.

**Generation of the product.** This product was derived using R programming, primarily the *lidR* (version: 4.0.3) and *leafR* (version: 0.3.5) packages[13,14]. The key processing steps include height normalization (Section 2.2.1) and metric calculation (Section 2.2.2). Parallel computing was applied for these steps to enhance efficiency.

**LiDAR data preprocessing.** For each tile of LiDAR data, we filtered the point cloud to remove noise and normalized heights to a flat surface. Specifically, we first filtered out noise points whose heights are greater than six standard deviations from the mean height and all points lower than ground points. Then, we set a 50 m buffer around each 1000 × 1000 m LiDAR tile, resulting 1100 × 1100 m range, to alleviate the edge effect when normalizing the heights. We normalized the heights using a digital terrain model (1 × 1 m) interpolated through the k-nearest neighbor approach with inverse-distance weighting using the *normalize_height* function in *lidR* package.

After normalization, we segmented the LiDAR points into individual 30 × 30 m pixels. We filtered out outlier points using the Noise Segmentation Algorithm (*ivf* filter) by setting a 3 × 3 × 3 m voxel resolution without other points in the surrounding 27 voxels and excluded the points that are not classified as unclassified (code: 1), ground (code: 2), low vegetation (code: 3), medium vegetation (code: 4), or high vegetation (code: 5). Then, we generated a canopy height model (CHM, 1 × 1 m) for the calculation of some SD metrics (Table 1).

| Category | Metrics | Description | Unit | Package Function | Minimum point density (pts/m$^2$) |
|---|---|---|---|---|---|
| Height | Quantile Heights (Q0, Q1, Q5, Q25, Q50, Q75, Q95, Q99, and Q100) | 0$^{th}$ to 100$^{th}$ quantiles of point height | m | *quantile(Z)* | 2.0–3.0 |
| | Mean Height (MH) | Mean of point height | m | *mean(Z)* | 2.4 |
| | Mean Canopy Height (MCH) | Mean of maximum height in 1m$^2$ grid | m | *mean(CHM_1m$^2$)* | 7.0 |
| Density | Vegetation Area Index (VAI)[abc] | Vegetation area per unit of ground area | m$^2$/m$^2$ | *leafR::lad.voxels(k = 1) leafR::lai* | 2.0 |
| Openness | Deep Gap Fraction (DGF)[a] | Fraction of 1 m$^2$ canopy gaps in the plot | 1 | Ngaps_1m$^2$/Ncells_1m$^2$ | 5.1 |
| | Gap Fraction Profile (GFP)[c] | Mean of gap fractions in 1-m horizontal slices | 1 | *lidR::gap_fraction_profile(dz = 1)* | 2.3 |
| Exterior Complexity | Top Rugosity (TopRugosity) | Standard deviation of outer canopy heights in 1 m$^2$ of plot | m | *sd(CHM_1m$^2$)* | 7.5 |
| | Rumple Index (RumpleIndex)[c] | Ratio between canopy surface area and its projected area on the ground | m$^2$/m$^2$ | *lidR::rumple_index(CHM_1m$^2$)* | 7.0 |
| Interior Complexity | Gini Coefficient Index (GiniCoeff)[c] | the Gini coefficient of point height | | *leafR::GC* | 2.2 |
| | Foliage Height Diversity (FHD)[abc] | Distribution of canopy cover among vertical canopy layers expressed as a diversity index. | | *leafR::lad.voxels(k = 1) leafR::FHD(evenness = F)* | 4.7 |
| | Standard Deviation of Height (SDH) | Standard deviation of point height | m | *sd(Z)* | 2.0 |
| | Coefficient Variation of Height (CVH) | Coefficient of variation of point height | | SDH/MH | 2.1 |
| | Mean of Standard Deviation of Height (MSDH) | Mean of standard deviation of point height in 3 × 3 m grid | m | *mean(sd(Z_9m$^2$))* | |
| | Standard Deviation of Standard Deviation of Height (SDSDH) | Standard deviation of standard deviation of height in 3 × 3 m grid | m | *sd(sd(Z_9m$^2$))* | 5.8 |
| Ancillary | Area[c] | Area of bounding box of points | m$^2$ | *lidR::area* | |
| | Point Density | Number of points per unit of ground area | pts/m$^2$ | N_points/Area | |

**Table 1.** Description of canopy structural diversity metrics in NEON-SD. The minimum point density needed for reliable SD estimation is extracted from LaRue *et al.*[9]. The vertical resolution for VAI and FHD is 1 m. Abbreviations and descriptions: Z – point height, CHM – canopy height model (of 1 × 1 m grids), Ngaps – number of gaps, Ncells – number of cells, Z_9m$^2$ – point heights in 3 × 3 m grids, N_points – number of points, dz – the thickness of the layers. [a]denotes the SD metrics were computed at three cut-off heights – 0.5, 2, and 5 m. [b]denotes the SD metrics were computed using three grain-sizes – 1, 10, and 30 m. [c]denotes the SD metrics were computed from existing packages: *lidR* or *leafR*.

**Metric calculation.** Based on the normalized LiDAR point cloud with a ground at 0 m, we calculated various SD metrics (Table 1) in 30 × 30 m pixels aligned with Landsat WRS-2 grids. Mean Canopy Height (MCH), Deep Gap Fraction (DGF), and Top Rugosity (Table 1) were calculated from the CHM and the rest of SD metrics in Table 1 were directly calculated from LiDAR points. To expedite the computation and avoid memory outage, we divided the LiDAR point cloud into 480 × 480 m blocks for each site-year and applied parallel computing for individual pixels' noise filtering and SD computation for each block. The computed metrics from these blocks were then merged and saved in two formats, CSV table and raster TIF files, for each site-year.

We applied three grain sizes (the horizontal resolution of voxels) to calculate NEON-SD metrics (1 m, 10 m, and 30 m) for Vegetation Area Index (VAI) and Foliage Height Diversity (FHD) calculations which were based on vegetation area density (VAD; vegetation area per unit of volume). Grain size has a complex effect on VAD estimation[15,16]. On the one hand, a larger grain size mitigates the effect of low pulse densities in lower portion of canopy caused by the occlusion of leafy crowns and produces a more stable VAD. On the other hand, a grain size as large as a tree or branches could cause strong leaf clumping, violating the MacArthur-Horn assumption of random leaf distribution[16,17]. Comparing the VAD profile in a tropical forest derived from airborne LiDAR and Portable Canopy profiling Lidar, de Almeida *et al.*[16] recommended a small grain size (<10 m) when point density is >15 points/m$^2$ and a 10-m grain size for lower density point cloud[16]. Therefore, we applied three grain sizes – 1 m, 10 m, and 30 m – for the VAD estimation in this product. Practically, VAD was computed with the function *leafR::lad.voxels*, which has a fixed vertical resolution of 1 m, with the extinction coefficient k set to 1. VAI and FHD were then calculated from VAD using *leafR::lai* and *leafR::FHD* functions, respectively.

In addition to the SD metrics, NEON-SD also provides ancillary information including the area of bounding box of LiDAR points (Area) and point density in individual pixels, serving as quality assessment of SD metrics. Particularly, point density significantly influences SD characterization and its impact depends on SD metric and forest type[9]. By artificially reducing point density, LaRue *et al.*[9] found that SD metric values stabilize when point density reaches or exceeds a certain threshold. The minimum point density required for robust estimation ranging from 2.0 to 7.5 points/m$^2$ across different SD metrics[9], which is documented in Table 1.

Note that all SD metrics were computed with a cut-off height of 0.5 m, effectively excluding the ground and very low vegetation points to consider only those ≥0.5 m above the ground[9]. However, to accommodate the complex and stratified nature of the vegetation systems, metrics including VAI, DGF, and Foliage Height Diversity (FHD) (Table 1) were also computed at two additional cut-off heights (2 m and 5 m) (Fig. 2). These thresholds were used to exclude small shrubs (<2 m)[18] and shrubs (<5 m), thus adapting NEON-SD for varied research needs across ecosystem types. For example, studies in regions with taller shrubs might opt for a higher

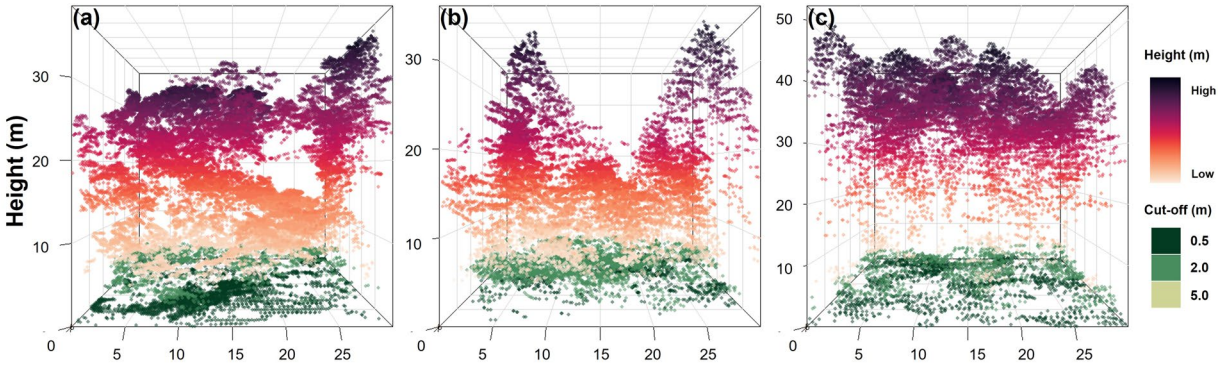


**Fig. 2** Examples of data excluded using three different cut-off heights in sample pixels of deciduous (**a**, SERC), mixed (**b**, ABBY) and evergreen forests (**c**, ABBY) for individual 30 x 30 m pixels.

cut-off to reduce the impact of the understory species. Additionally, comparing SD between cut-off heights helps characterize canopy structure within specific strata.

## Data Records

The 30 m NEON-SD dataset is publicly available through the Environmental Data Initiative (EDI, https://doi.org/10.6073/pasta/e02f855d69193a46571168575b35291d)[19]. It provides SD across 45 Terrestrial NEON sites spanning the years 2013–2022, resulting in 211 site-years. The data are presented in two file formats: CSV and TIF. These two formats contain the same information and are offered for user convenience. SD metrics are represented as columns in CSV files and band files in TIF folders. The naming conventions for CSV and TIF files are as follows:

$$SD\_YYYY\_SITE\_N\_EPSGDDDDD.csv$$

$$SD\_YYYY\_SITE\_N\_METRIC.tif$$

where *YYYY*, *SITE*, and *N* represents the acquisition year, site ID, and visit number, collectively forming the site-year ID. *EPSGDDDDD* is the EPSG code for the UTM projection of the X and Y coordinates, and METRIC specifies the metric name. The descriptions of individual metrics can be found in the metadata https://portal.edirepository.org/nis/metadataviewer?packageid=edi.1519.1.

An example from the site-year *2021_TALL_6* illustrates SD metrics (MCH, VAI, DGF, TopRugosity, and FHD, each representing one of the five SD categories) of various vegetation types including deciduous, evergreen, mixed forests and shrubland/grasslands (Fig. 3). These metrics generally correspond with the national land cover database (NLCD) map[20], with forested areas displaying higher values of MCH, VAI, TopRugosity, and FHD, and lower DGF than shrubland/grassland. An increase in the cut-off height led to a reduction in VAI and FHD, and a rise in DGF, particularly more evident in shrub/grasslands compared to trees (Fig. S1). Point density varied spatially along flight lines, with increased densities in areas where flight lines overlap. The average point density is $18 \pm 8$ points/m$^2$ (Table S1), meeting the requirement for reliable SD estimation, and no clear association of spatial pattern between SD metrics and point density was observed because of the relatively high point density.

## Technical Validation

We compared our NEON_SD metrics with a variety of independent products including other NEON products and spaceborne LiDAR data that provide the same or similar metrics. The goal of these comparisons is to place this data product in the context of other commonly used sources of data rather than attempting to establish a single "best" source of data. Our comparison was performed on height metrics, DGF, FHD, and VAI based on data availability. The cut-off height of 0.5 m for NEON-SD metrics was used in all evaluations because this scenario most closely aligns with reference products that typically measure the entire canopy. Linear regression modeling was conducted to compare NEON-SD with reference datasets, using $R^2$ and Root Mean Square Deviation (RMSD) for quantitative analysis. RMSD is calculated from the estimated metric ($\hat{y}$) and the reference ($y$) as in Eq. 1:

$$RMSD = \sqrt{\frac{\sum (y - \hat{y})^2}{n}} \quad (1)$$

**Height metrics.** We compared NEON-SD's height metrics with both ground surveys and spaceborne LiDAR data. For ground survey reference, we used the NEON's Vegetation Structure Data Product (DP1.10098.001)[21], which provides in-situ structure measurements using rangefinders, measuring sticks, or equivalent tools and mapped positions of individual plants for NEON Distributed and Tower plots. Meanwhile, we compared

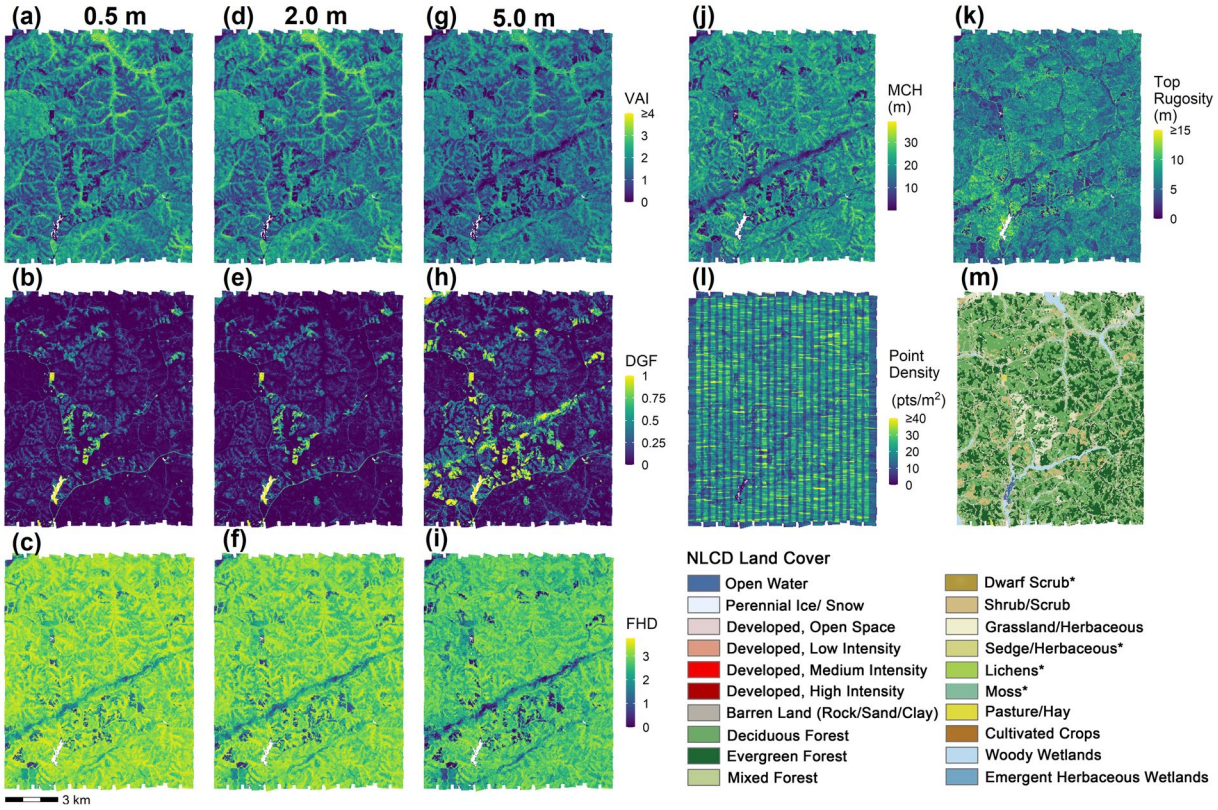


**Fig. 3** Maps of structural diversity in 2021_TALL_6. Vegetation Area Index (VAI), Deep Gap Fraction (DGF), and Foliage Height Diversity (FHD) with a cut-off of 0.5 m (**a**–**c**), 2 m (**d**–**f**), and 5 m (**g**–**i**); Mean Canopy Height (MCH) (**j**); Top Rugosity (**k**), Point Density (**l**), and NLCD land cover (**m**).

NEON-SD's height metrics with GEDI L2A product from the GEDI mission that was launched in December 2018 aboard the International Space Station[22]. While airborne LiDAR measurements are generally more accurate than those of GEDI, we compared our product with GEDI to better understand the consistency of SD metrics between airborne and spaceborne LiDAR products. The information can be useful for scaling up from ALS-derived structural diversity at high resolution to GEDI-derived SD at moderate resolutions (e.g., Hakkenberg *et al.*[4]).

For MH and MCH, we selected 34 terrestrial NEON sites where the NEON's Vegetation Structure data are available. Ground-based tree height surveys by NEON have been conducted using rangefinders, measuring sticks, or equivalent tools. We calculated the yearly plot-level height by averaging the individual tree heights in each plot. Then, we compared them with NEON-SD's MH and MCH by the year. For Q50 and Q99 evaluations, we selected 23 sites covered by the GEDI L2A data during leaf-on seasons (between June to September in 2020 to 2022). Following GEDI's product guidelines[22], we filtered out daytime acquisitions with sensitivity less than 0.95 and those marked with a quality flag of 0. To compare NEON-SD with GEDI products, we calculated the weighted mean values by computing zonal statistics with 12.5 m buffers surrounding each GEDI beam's center location. We matched the years in comparisons between our height metrics with GEDI L2A data.

The NEON-SD heights were consistent ($R^2 > 0.6$) with ground surveys. The comparison between MH and ground survey mean height demonstrated an almost 1:1 correspondence (slope = 1.1), with the $R^2$, RMSD, and RMSD% being 0.71, 3.94 m, and 28.69%, respectively (Fig. 4a). The comparison of MCH indicated that the $R^2$, RMSD, and RMSD% were 0.61, 6.11 m, and 44.46%, respectively (Fig. 4b). Ground survey mean height was 0.84 m and 1.26 m lower than NEON-SD MH and MCH, respectively (Fig. 4a,b). We speculated that the difficulties in measuring the height of tall trees and the complexity of the topography could influence these underestimations in the ground survey data[23].

When compared to GEDI L2A's RH99 and RH50, NEON-SD's Q99 and Q50 were higher by 0.2 and 1.4 m, respectively (Fig. 5). The Q99 comparison yielded a $R^2$ of 0.64, RMSD of 6.9 m, and RMSD% of 37.5%, respectively, and the values were 0.59, RMSD of 5.72 m, and RMSD% of 67.7% for the Q50 comparison. Moreover, the correlation between the GEDI and NEON-SD heights varied across the forest types (deciduous, evergreen, mixed forests, and shrub) with $p$-values $< 0.001$. Specifically, evergreen forests showed the highest $R^2$ (0.59 and 0.59 for Q99 and Q50, respectively), followed by deciduous forests (0.32 and 0.35), mixed forests (0.26 and 0.28) and shrub/scrub (0.35 and 0.16). The height comparisons appeared to be unaffected by the LiDAR point density, despite a significant association between point density and both Q99 and Q50. The scaled coefficients for density in each model are 0.9, respectively, while those for scaled RH99 and RH50 are 8.9 and 6.2, respectively, with all p-values $< 0.001$ (Fig. 5c,d). The disagreement between these two datasets might result from uncertainties in GEDI beam locations, slopes, aspects, and canopy heterogeneity[24,25], misalignment between the GEDI

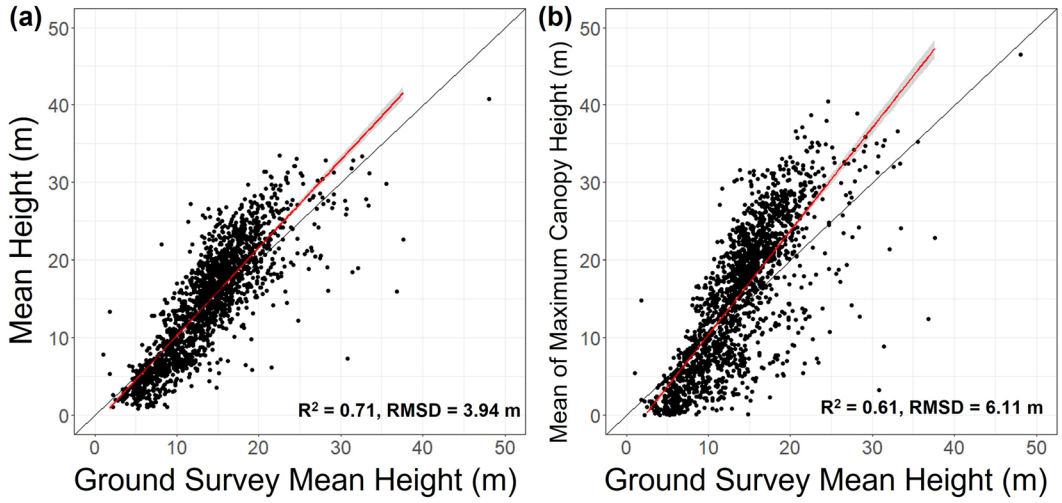


**Fig. 4** Height comparisons between the ground surveys and our products at NEON plots. The solid black line represents the 1:1 line. The red line and surrounding grey shadow represent the regression line and 95% confidence intervals, respectively.

footprint (25 m-radius) and our gridded product (30 m-grid), and land cover changes between AOP and GEDI acquisitions. Ground observations of tree height and LiDAR estimates of canopy height both have their own well-known sources of measurement error. Thus, while our data product broadly agrees with other sources, all values should be considered estimates and interpreted with caution.

We also compared FHD and DGF between NEON-SD and GEDI L2B product (Figs. S3, S4). They showed the highest $R^2$ in evergreen forests (FHD: 0.46 and DGF: 0.26), followed by shrub/scrub (0.26, 0.20), deciduous (0.25, 0.10), and mixed forests (0.16, 0.09), with point density not appearing to affect the comparison results (Fig. S2). NEON-SD's FHD aligned well with GEDI when the value exceeded 1.5. In the comparison of DDF. In the case of the DGF comparison, NEON-SD generally showed lower values compared to GEDI L2B across the forest types, except for the shrub/scrub type (Fig. S3).

**VAI evaluation.** We evaluated NEON-SD's VAI using NEON's 1-m Leaf Area Index (LAI) product (DP3.30012.001) as an independent reference[21]. The LAI product (hereafter called spectra-LAI) is derived using vegetation index from NEON Imaging Spectrometer, whereas the LiDAR-VAI in NEON-SD is based on the Beer-Lambert Law[15]. Although spectra-LAI represents the photosynthetic components (primarily leaves) and LiDAR-VAI reflects all components that interact with LiDAR pulses, it is generally assumed that a linear relationship exists between them[3,26,27]. We compared LiDAR-VAI with Spectra-LAI to assess their correlation and to investigate how point density and grain size—factors known to influence VAD estimation[15,16]—affect this relationship. This comparison helps clarify the relationship between LiDAR-VAI and Spectra-LAI across different conditions and evaluates the influence of point density and grain size on VAI estimation.

The comparison between NEON-SD VAI and spectra-LAI was conducted on a site-year basis to account for the variations in canopy properties and point-density across site-years which influence the extinction coefficient and VAI calculation[28]. Thus, we randomly selected three NEON sites representing different forest types and canopy properties: MLBS (deciduous forest), SOAP (evergreen forest), and TALL (mixed forest). For each site, we included two acquisitions to capture variations in point density. The site-years 2021_MLBS_4, 2021_SOAP_5, and 2021_TALL_6 (captured with either Galaxy or Riegl) displayed higher average point densities, and site-years 2017_MLBS_2, 2019_SOAP_4, and 2019_TALL_5 (captured with Gemini) displayed relatively lower point densities (Table S2). Spectra-LAI was averaged into 30-m pixels to match with NEON-SD. To examine the impact of grain size on VAI, we evaluated six grain sizes (1 m, 3 m, 5 m, 10 m, 15 m, and 30 m), though NEON-SD ultimately included only three of them (1 m, 10 m, and 30 m). Specifically, $R^2$ was calculated from linear modeling between spectra-LAI and LiDAR-VAI, and RMSD was calculated between spectra-LAI and the LAI predicted from NEON-SD VAI ($\hat{LAI}$) (Eq. 2):

$$\hat{LAI} = b + a \times VAI \tag{2}$$

Results showed relatively low $R^2$ (0.08–0.44) between LiDAR-VAI and spectra-LAI, which could result from the difference in the definitions and retrieving methods between LiDAR-VAI and spectra-LAI, sensor artifacts, and sensor configuration (e.g., point density). Acquisitions with higher point densities (24, 20, and 18 points/$m^2$ for 2021_MLBS_4, 2O21_SOAP_5 and 2021_TALL_6, respectively) retrieved higher $R^2$ and comparable *RMSD* values than those with lower point densities (5, 5, and 6 points/$m^2$ for 2017_MLBS_2, 2O19_SOAP_4, and 2019_TALL_5, respectively). This indicates more effective VAI characterization with denser points.

For both acquisitions with higher and lower point densities, smaller grain sizes (1 m or 3 m) resulted in larger $R^2$ and smaller RMSD. Particularly, acquisitions with higher point densities saw a closer VAI-LAI relationship

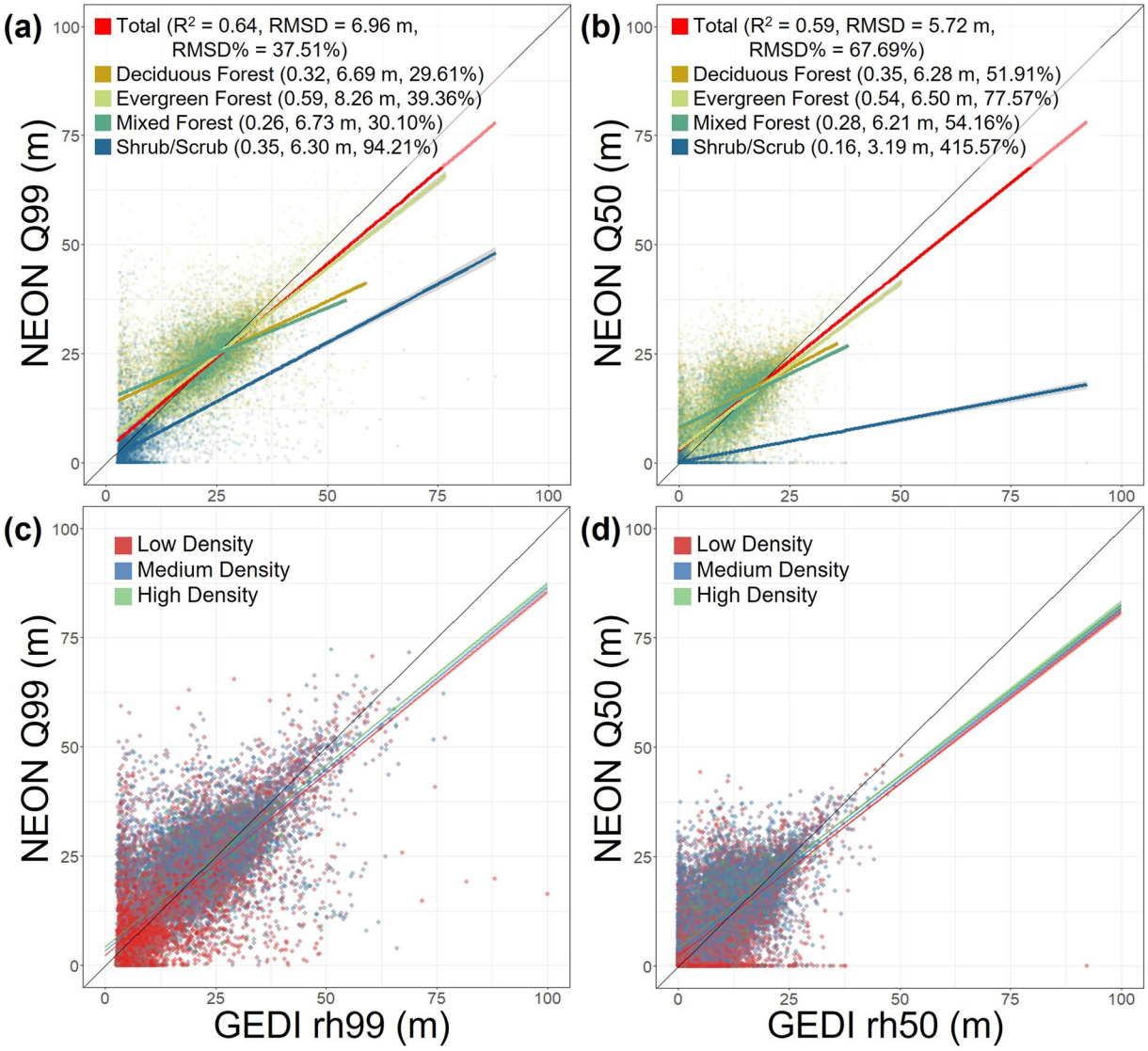


**Fig. 5** Comparisons of quantile 99$^{th}$ and 50$^{th}$ heights between NEON-SD and GEDI by land cover (**a,b**) and point density (**c,d**). The colors of points and lines represent forest type in (**a,b**), and surrounding grey shadow represent the regression line and 95% confidence intervals. Low to high categories of point density (**c,d**) were classified by the mean ± 1 standard deviation of the point density. The solid black line represents the 1:1 line.

at the 1 m grain size (Fig. 6a,b); whereas those with lower densities found the closest match at a 3 m grain size (Fig. 6c,d). This pattern aligns with the finding that a finer grain size results in a more accurate estimation of LAI for denser point clouds[16].

Interestingly, a grain size of 10 m for lower-density acquisitions obtained relatively lower $R^2$ and higher *RMSD* values than other grain sizes (Fig. 6c,d), deviating from the recommendation in de Almeida[16] of a 10-m grain size for a stable VAI estimation when point density is $<$10 points/m$^2$. The discrepancy could stem from two main differences. First, the forest type between the studies may affect the results – this study selected temperate forests, while de Almeida *et al.*[16] focused on a tropical rainforest[16]. Second, the evaluation methods differ in that this study compares VAI with Spectra-LAI, whereas de Almeida[16] assessed the stability of VAI relative to grain size.

The spatial pattern of NEON-SD's VAI and Spectral LAI generally matches land cover for site-years with higher point densities (2021_MLBS_4, 2O21_SOAP_5 and 2021_TALL_6) (Fig. 7). Forested areas display higher VAI or LAI than other land cover types. We presented only VAI maps with a grain size of 1 m as the spatial pattern of VAI is consistent across different grain sizes. We note that Spectra-LAI exhibits striping in 2021_MLBS_4 and 2021_TALL_5 (Fig. 7b,e), as well as in 2017_MLBS_2 (Fig. S4d), potentially due to differences in atmosphere conditions, sensor artifacts, and/or solar angle between flight lines. In contrast, acquisitions with lower point densities (2017_MLBS_2, 2O19_SOAP_4, and 2019_TALL_5) show striping associated with point density variation in VAI at grain size of 1 m (Figs. S4a, S4m), but not in VAI at grain sizes of 10 m or 30 m. It suggests that a grain size of 10 m or 30 m should be used for NEON-SD VAI of acquisitions with lower point density (captured with Gemini).

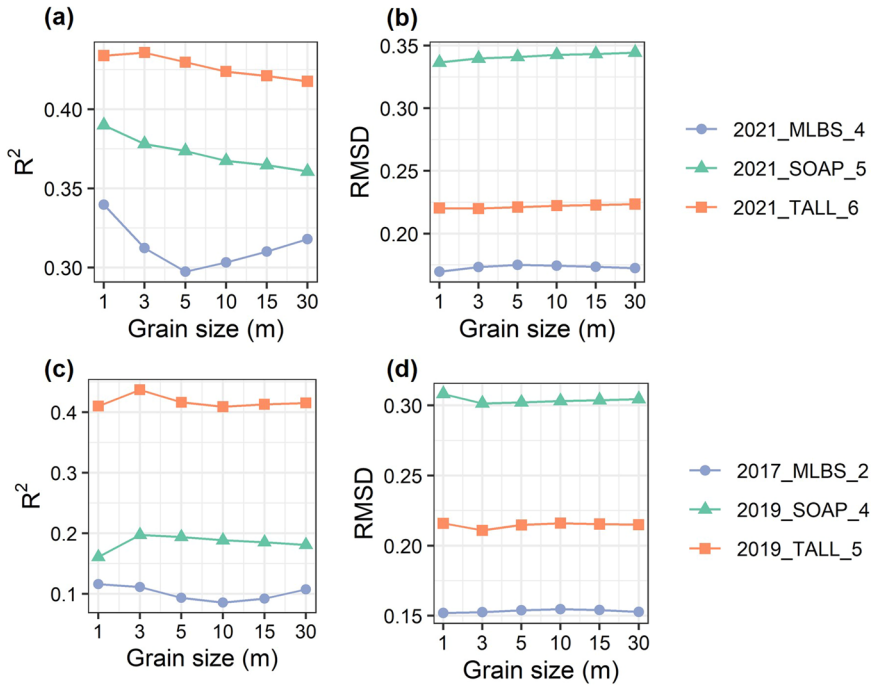


**Fig. 6** Comparisons between NEON-SD's VAI and Spectra-LAI from NEON. $R^2$ (**a**) and *RMSD* (**b**) with grain size for site-years 2021_MLBS_4 (24 points/m$^2$), 2O21_SOAP_5 (20 points/m$^2$), and 2021_TALL_6 (18 points/m$^2$). $R^2$ (**c**) and *RMSD* (**d**) with grain size for site-years 2017_MLBS_2 (5 points/m$^2$), 2O19_SOAP_4 (5 points/m$^2$), and 2019_TALL_5 (6 points/m$^2$).

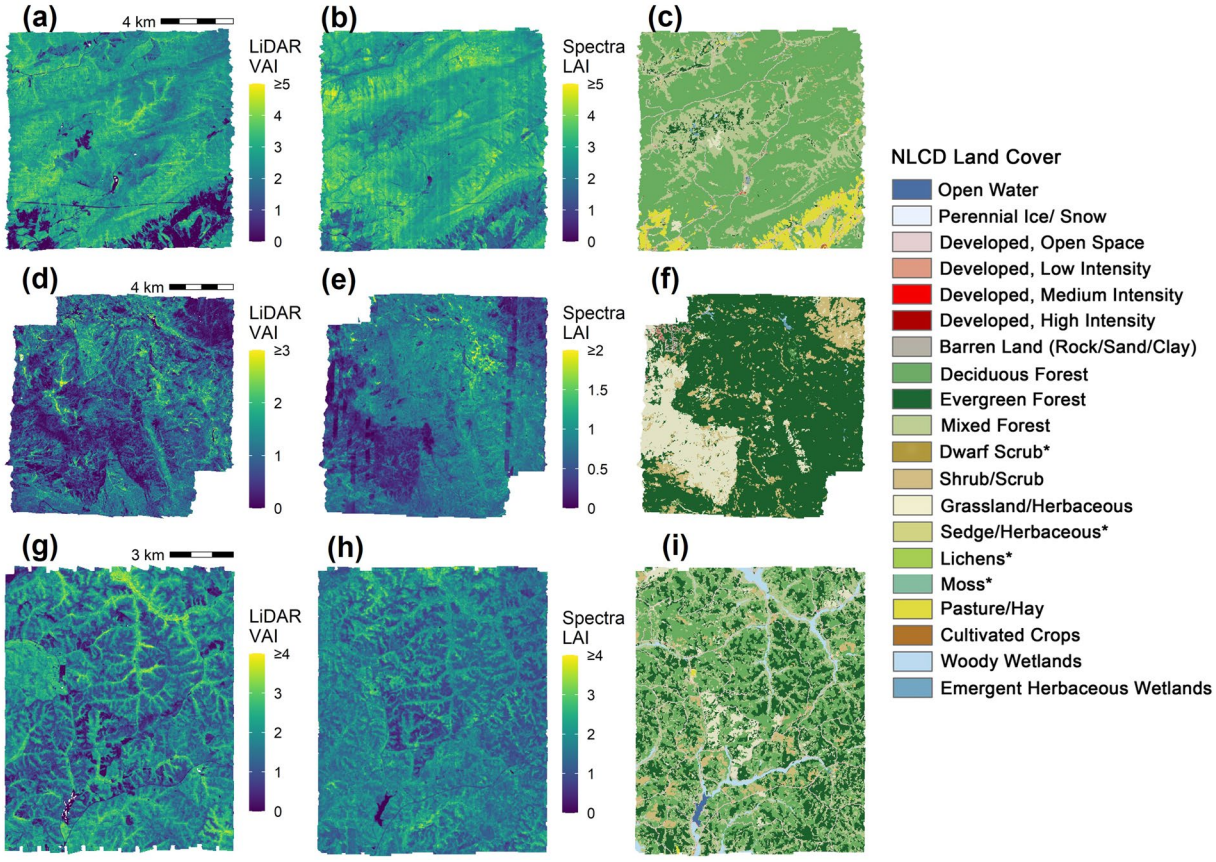


**Fig. 7** Maps of NEON-SD's VAI (cut-off height = 0.5 m, grain size = 1 m), Spectra-LAI from NEON and NLCD land cover. VAI (**a**), LAI (**b**), and land cover (**c**) for site-year 2021_MLBS_4; VAI (**d**), LAI (**e**), and land cover (**f**) for site-year 2O21_SOAP_5; and VAI (**g**), LAI (**h**), and land cover (**i**) for site-year 2021_TALL_6.

## Usage Notes

Our NEON-SD product provides structural diversity metrics encompassing height, density, openness, and exterior and interior complexity at three cut-off heights (0.5 m, 2 m, and 5 m). Users can select an appropriate cut-off height based on their specific needs. For studies focusing on all vegetation, a cut-off height of 0.5 meters is advisable; for studies focusing on canopy vegetation, a cut-off height of 2 m or 5 m (depending on height of understory vegetation) is recommended. Additionally, three grain sizes (1 m, 10 m, and 30 m) were used for Vegetation Area Index (VAI) and Foliage Height Diversity. Our findings suggest a 1 m grain size is optimal for VAI in site-years captured by higher density LiDAR systems like Optech Galaxy or Riegl LMS-Q780, while a 10 m or 30 m grain size is suitable for those captured by Optech Gemini (lower point density). Further, point density for each pixel is provided as an important factor influencing the estimation of SD metrics. Users are advised to use metrics that meet the minimum point density requirements (Table 1) for reliable SD metric estimation, particularly when studying the spatial or temporal dynamics in SD. In addition, users can compare SD metrics across sites and years, calculated within the same grains and cut-off heights, with application to a wide range of ecological research topics (e.g., Choi *et al.*[8]). Last, steep slopes and complex local terrains may influence SD values, particularly FHD[29]. We recommend that users apply our SD product in conjunction with NEON's topography product (DP3.30025.001[30]) with caution in areas of complex topography, especially in applications that rely on the actual vertical vegetation structure.

## Code availability

The product NEON-SD and the computer code with example input data are openly available on the Environmental Data Initiative: https://doi.org/10.6073/pasta/e02f855d69193a46571168575b35291d[19].

Received: 18 July 2024; Accepted: 18 October 2024;
Published online: 29 October 2024

## References

1. LaRue, E. A. *et al*. A theoretical framework for the ecological role of three-dimensional structural diversity. *Front. Ecol. Environ.* **21**, 4–13 (2023).
2. LaRue, E. A. *et al*. Structural diversity as a reliable and novel predictor for ecosystem productivity. *Front. Ecol. Environ.* **21**, 33–39 (2023).
3. Hardiman, B. S., Bohrer, G., Gough, C. M., Vogel, C. S. & Curtisi, P. S. The role of canopy structural complexity in wood net primary production of a maturing northern deciduous forest. *Ecology* **92**, 1818–1827 (2011).
4. Hakkenberg, C. R. *et al*. Inferring alpha, beta, and gamma plant diversity across biomes with GEDI spaceborne lidar. *Environ. Res.: Ecology* **2**, 035005 (2023).
5. Atkins, J. W. *et al*. Application of multidimensional structural characterization to detect and describe moderate forest disturbance. *Ecosphere* **11** (2020).
6. Atkins, J. W., Shiklomanov, A., Mathes, K. C., Bond-Lamberty, B. & Gough, C. M. Effects of forest structural and compositional change on forest microclimates across a gradient of disturbance severity. *Agric. For. Meteorol.* **339**, 109566 (2023).
7. Gough, C. M. *et al*. Disturbance has variable effects on the structural complexity of a temperate forest landscape. *Ecol. Indic.* **140**, 109004 (2022).
8. Choi, D. H. *et al*. Short-term effects of moderate severity disturbances on forest canopy structure. *J. Ecol.* **111**, 1866–1881 (2023).
9. LaRue, E. A. *et al*. Evaluating the sensitivity of forest structural diversity characterization to LiDAR point density. *Ecosphere* **13** (2022).
10. NEON. Discrete return LiDAR point cloud (DP1.30003.001). https://data.neonscience.org/data-products/DP1.30003.001 (2023).
11. Silveira, E. M. O. *et al*. Multi-grain habitat models that combine satellite sensors with different resolutions explain bird species richness patterns best. *Remote Sens. Environ.* **295**, 113661 (2023).
12. Gough, C. M., Atkins, J. W., Fahey, R. T. & Hardiman, B. S. High rates of primary production in structurally complex forests. *Ecology* **100**, e02864 (2019).
13. Roussel, J.-R. *et al*. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sens. Environ.* **251**, 112061 (2020).
14. de Almeida, D. R. A., Stark, S. C., Silva, C. A., Hamamura, C. & Valbuena, R. leafR: Calculates the Leaf Area Index (LAD) and Other Related Functions. https://CRAN.R-project.org/package=leafR (2021).
15. Kamoske, A. G., Dahlin, K. M., Stark, S. C. & Serbin, S. P. Leaf area density from airborne LiDAR: Comparing sensors and resolutions in a temperate broadleaf forest ecosystem. *For. Ecol. Manage.* **433**, 364–375 (2019).
16. de Almeida, D. R. A. *et al*. Optimizing the Remote Detection of Tropical Rainforest Structure with Airborne Lidar: Leaf Area Profile Sensitivity to Pulse Density and Spatial Sampling. *Remote Sensing* **11**, 92 (2019).
17. Atkins, J. W. *et al*. Scale dependency of lidar-derived forest structural diversity. *Methods Ecol. Evol.* **14**, 708–723 (2023).
18. Filippelli, S. K., Lefsky, M. A. & Rocca, M. E. Comparison and integration of lidar and photogrammetric point clouds for mapping pre-fire forest structure. *Remote Sens. Environ.* **224**, 154–166 (2019).
19. Wang, J. *et al*. Structural Diversity from the NEON Discrete-Return LiDAR Point Cloud in 2013-2022. Environmental Data Initiative https://doi.org/10.6073/pasta/e02f855d69193a46571168575b35291d (2023).
20. Dewitz, J. National Land Cover Database (NLCD) 2021 products: U.S. Geological Survey data release. https://doi.org/10.5066/P9JZ7AO3 (2023).
21. NEON. LAI - spectrometer - mosaic (DP3.30012.001). https://doi.org/10.48443/Q59T-3788 (2023).
22. Dubayah, R. *et al*. GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002. https://doi.org/10.5067/GEDI/GEDI02_A.002 (2021).
23. Wang, Y. *et al*. Is field-measured tree height as reliable as believed – A comparison study of tree height estimates from field measurement, airborne laser scanning and terrestrial laser scanning in a boreal forest. *ISPRS J. Photogramm. Remote Sens.* **147**, 132–145 (2019).
24. Oliveira, P. V. C., Zhang, X., Peterson, B. & Ometto, J. P. Using simulated GEDI waveforms to evaluate the effects of beam sensitivity and terrain slope on GEDI L2A relative height metrics over the Brazilian Amazon Forest. *Egypt. J. Remote Sens. Space Sci.* **7**, 100083 (2023).
25. Roy, D. P., Kashongwe, H. B. & Armston, J. The impact of geolocation uncertainty on GEDI tropical forest canopy height estimation and change monitoring. *Egypt. J. Remote Sens. Space Sci.* **4**, 100024 (2021).
26. Fang, H., Baret, F., Plummer, S. & Schaepman-Strub, G. An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Rev. Geophys.* **57**, 739–799 (2019).

27. Gough, C. M., Vogel, C. S., Schmid, H. P., Su, H.-B. & Curtis, P. S. Multi-year convergence of biometric and meteorological estimates of forest carbon storage. *Agric. For. Meteorol.* **148**, 158–170 (2008).
28. Tian, L. & Qu, Y. Assessing Factors That Affect the Estimation of a Canopy's Gap Fraction and Extinction Coefficient Using Discrete Airborne LiDAR Data. *IEEE Trans. Geosci. Remote Sens.* **61**, 1–14 (2023).
29. Liu, J., Skidmore, A. K., Heurich, M. & Wang, T. Significant effect of topographic normalization of airborne LiDAR data on the retrieval of plant area index profile in mountainous forests. *ISPRS J. Photogramm. Remote Sens.* **132**, 77–87 (2017).
30. NEON. Slope and Aspect - LiDAR (DP3.30025.001). National Ecological Observatory Network (NEON) (2024).
31. Silva, C. A. *et al*. rGEDI: NASA's Global Ecosystem Dynamics Investigation (GEDI) Data Visualization and Processing. https://github.com/carlos-alberto-silva/rGEDI (2024).

## Acknowledgements

The National Ecological Observatory Network (NEON) is a program sponsored by the National Science Foundation and operated under cooperative agreement by Battelle. DP1.30003.001 and DP3.30012.001 used in this research were obtained through NEON Research Support Services. We thank Bridget Hass and Tristan Goulden from NEON for coordinating delivering the AOP LiDAR data. The GEDI L2A and L2B data were downloaded by using rGEDI package (https://github.com/carlos-alberto-silva/rGEDI)[31]. This work was supported by NSF awards 1926538 to Hardiman and Foster, 1926454 to Matthes, and 1926442 to Fahey and by USDA NIFA grant 2023-68012-38992 to Fei and Hardiman. Additional funding and computational capacity was provided by the Institute for Digital Forestry at Purdue University. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

## Author contributions

Jianmin Wang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing. Dennis H. Choi: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing. Elizabeth La Rue: Methodology, Writing – Review & Editing. Jeff W. Atkins: Methodology, Writing – Review & Editing. Jane R. Foster: Writing – Review & Editing. Jaclyn H. Matthes: Writing – Review & Editing, Robert T. Fahey: Writing – Review & Editing. Songlin Fei: Writing – Review & Editing. Brady Hardiman: Conceptualization, Writing – Review & Editing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41597-024-04018-0.

**Correspondence** and requests for materials should be addressed to B.S.H.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024