https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

DigiSylva: a plot-level dataset of forest attributes integrating field and airborne laser scanning measurements across the boreal-temperate ecotone of the northeastern United States

Xinyuan Wei [1, 2, 3], Daniel J. Hayes [3, 4], Lei Ma [5], Jianheng Zhao [3, 4], Grant Daniels [3], Joanne C. White [6], Archibald Owen [3], David Sandilands [3], Shambhu Katel [3], Chenye Zhang [3], George Hurtt [5], and Aaron Weiskittel [3, 4]

[1] Blackland Research & Extension Center, Texas A&M AgriLife Research, Temple, TX, 76502, USA
[2] Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, 77843, USA
[3] School of Forest Resources, University of Maine, Orono, ME, 04473, USA
[4] Center for Research on Sustainable Forests, University of Maine, Orono, ME, 04473, USA
[5] Department of Geographical Sciences, University of Maryland, College Park, MD, 20742, USA
[6] Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, Victoria, British Columbia V8Z 1M5, Canada

Correspondence to: Xinyuan Wei (xinyuan.wei@ag.tamu.edu)

Abstract Enhancing estimates of forest attributes such as structure, composition, volume, and biomass across large spatial extents relies on field inventory measurements to calibrate and validate relationships with metrics derived from airborne laser scanning (ALS) point clouds. However, mixed temperate and boreal forests in the northeastern United States remain underrepresented in publicly available datasets, especially those that co-locate field inventory and ALS observations. In this study, we developed DigiSylva, a spatially explicit database integrating field inventory and various remote sensing products across five representative forest ecosystems within the boreal-temperate ecotone of the northeastern United States. This database includes 308 fixed-area plots spanning inland, managed, experimental, coastal, and northern hardwood forest systems. For each plot, individual tree measurements, including species, diameter at breast height, and spatial position, were recorded. Plot centers were surveyed with sub-meter GNSS accuracy to enable precise spatial alignment with airborne and spaceborne observations. Tree height, aboveground biomass, and stem volume were estimated using multiple allometric methods. Multi-temporal ALS point clouds acquired between 2010 and 2026 were clipped to plot boundaries and processed to derive standardized forest structural metrics. In addition, co-registered NAIP imagery and Harmonized Landsat and Sentinel-2 (HLS) data were provided for each plot. In total, the dataset contains 6,137 trees, 308 plot-level forest attribute records, and 1,408 plot-level ALS point clouds. This integrated dataset can support forest structural analysis, development of ALS-based forest attribute predictive models (e.g., volume), and multi-temporal biomass monitoring in structurally complex, mixed-species forest ecosystems.

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

1 Introduction

Forests play a key role in biodiversity, ecosystem services, and global carbon cycles (Mori et al., 2017; Pan et al., 2024; Poulter et al., 2025); however, mapping of their structure, biomass, and volume across large areas remains challenging (Kangas and Maltamo, 2006). Traditional, ground-based forest inventories provide detailed measurements (e.g., size, species, and spatial location), but such surveys are labor-intensive and spatially limited (McRoberts et al., 2010). In addition, interpolating these local measurements to regional or global extents is difficult without the support of remote sensing products (McRoberts and Tomppo, 2007). Remote sensing, especially airborne laser scanning (ALS), has been demonstrated as a powerful complement to field-based forest surveys, enabling the upscaling of plot-level forest inventory measurements to large-area forest maps (Huang et al., 2019; Tang et al., 2021; White et al., 2016).

Because ALS does not directly measure forest attributes such as basal area, biomass, timber volume, or species diversity, field measurements are generally required to calibrate and validate area-based models that estimate these attributes from ALS-derived structural metrics (White et al., 2016). ALS-derived metrics, such as height percentiles, canopy density, and vertical distribution indices, have been shown to be strongly related to basal area, volume, and biomass across a range of forest types (Knapp et al., 2020). By developing and calibrating models that link field-based measurements with ALS-derived structural metrics, these relationships can be used to predict forest attributes across large spatial extents (Coops et al., 2025; Guo et al., 2021; Pascual et al., 2013). Therefore, integrating ALS point clouds with co-located inventory plots is critical for enhanced forest inventory and supporting large-area forest attribute mapping.

In the United States, several forest inventory datasets support the linkage between field measurements and ALS-derived structural metrics. The Forest Inventory and Analysis (FIA) program provides a nationwide database of field-measured forest attributes (Tinkham et al., 2018). The FIA database, available at both plot and tree levels, includes records such as tree species, tree size, location, volume, and biomass for a large number of plots across all states. However, the publicly available FIA data are spatially fuzzed to protect plot confidentiality and therefore do not provide the precise plot location information required for direct alignment with ALS point clouds (Wei et al., 2026a). Another important resource is the National Ecological Observatory Network (NEON), which collects standardized multi-sensor observations at more than 40 field sites spanning diverse ecosystems (Atkins et al., 2023; Atkins et al., 2025). NEON openly provides co-registered in situ tree measurements, including stem locations, species, and crown attributes, together with high-resolution ALS point clouds and hyperspectral imagery (Atkins et al., 2025; Scholl et al., 2020). Globally, national forest inventory programs, such as Canada's National Forest Inventory (NFI) (Gillis et al., 2005) and Europe's LUCAS and national inventory networks (Vidal et al., 2016), could be integrated with ALS surveys to improve large-area estimation of forest structure, volume, and biomass. These efforts highlight the growing importance of integrated forest inventory and ALS datasets for advancing model calibration, large-scale forest attribute mapping, and forest monitoring.

Numerous forest types across the United States remain underrepresented in publicly available forest inventory and ALS datasets. Mixed temperate and boreal forests in the northeastern United States are a representative region. These forests

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



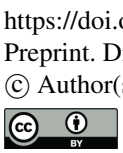
occur within a well-defined ecotone between boreal spruce-fir forests and northern temperate hardwood forests and therefore include species associated with both climatic regimes (Hollinger et al., 2021), as well as a range of management histories and practices (Zhao et al., 2025). The high biodiversity, structural complexity, economic importance, and carbon storage potential of this region requires accurate, large-area estimates of forest attributes; however, few publicly available datasets provide co-located plot-level inventory data and ALS point clouds for these mixed forest systems. To provide co-located forest inventory data and corresponding plot-level ALS point clouds for this region, we developed DigiSylva, a dataset spanning five representative forest ecosystems. The objectives of this database are to (1) provide the scientific community with a freely accessible dataset integrating field-measured forest inventory data, co-located ALS point clouds, and optical imagery data at the plot level; (2) enable the development, calibration, and evaluation of area-based models linking inventory-derived forest attributes with ALS-derived metrics; and (3) create opportunities to examine spatial and temporal variation in forest structure, improve large-area forest attribute estimation, and facilitate method development for structurally complex, mixed-species transitional forest ecosystems.

2 Data and Methods

2.1 Study sites

The field inventory data were collected in five representative forest ecosystems of the northeastern United States, including Howland Research Forest, Dwight B. Demeritt Forest, Penobscot Experimental Forest, Schoodic Peninsula Forest, and Western Maine Forest (Fig. 1). These sites collectively span the boreal-temperate transition zone and encompass a gradient of climatic conditions, management histories, and structural characteristics. Howland Research Forest, located in central Maine, USA, covers more than 2 km^2 and represents an ecotonal spruce-hemlock system transitioning between boreal spruce-fir and northern temperate hardwood forests (Savage et al., 2009). The region experiences a cool, humid continental climate with a mean annual temperature of approximately 6 °C and annual precipitation of 1000-1150 mm (Richardson et al., 2019). The forest is dominated by red spruce (*Picea rubens*) and eastern hemlock (*Tsuga canadensis*), forming a mature, multi-aged canopy with complex vertical structure (Wei et al., 2025; Wei et al., 2026c). Dwight B. Demeritt Forest is a long-term research and teaching forest situated within the northern temperate forest zone and is representative of managed mixedwood systems in the region (University of Maine, 2026). Forest composition includes eastern white pine (*Pinus strobus*), red maple (*Acer rubrum*), eastern hemlock (*Tsuga canadensis*), and northern red oak (*Quercus rubra*). Active forest management has produced a heterogeneous mosaic of age classes, canopy densities, and vertical structural conditions. Penobscot Experimental Forest, located within the Acadian Forest Region, is one of the longest-running experimental forests in the United States and has been extensively managed under different silvicultural treatments (Kenefic and Brissette, 2014; Rogers and Kenefic, 2017). Its composition is dominated by conifers such as red spruce, balsam fir (*Abies balsamea*), eastern hemlock, and eastern white pine, with minor hardwood components, resulting in substantial variability in stand structure and canopy height. Schoodic Peninsula Forest, situated along the eastern coast of Maine within Acadia National

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

Park, represents a maritime-influenced coastal forest ecosystem (Workman, 2014). Climatic conditions are characterized by high humidity, coastal fog, and moderate precipitation. Forest stands are shaped by shallow soils, wind exposure, and disturbance history, leading to comparatively lower canopy heights and heterogeneous structural patterns. Western Maine Forest, located in western Maine, extends the dataset into a northern hardwood-dominated system at the western edge of the study region. Its stands are dominated by American beech (*Fagus grandifolia*), northern red oak (*Quercus rubra*), red maple (*Acer rubrum*), and eastern hemlock (*Tsuga canadensis*). These five regions provide a robust representation of inland, managed, experimental, coastal, and northern hardwood forest ecosystems within the broader boreal-temperate transitional landscape of the northeastern United States.

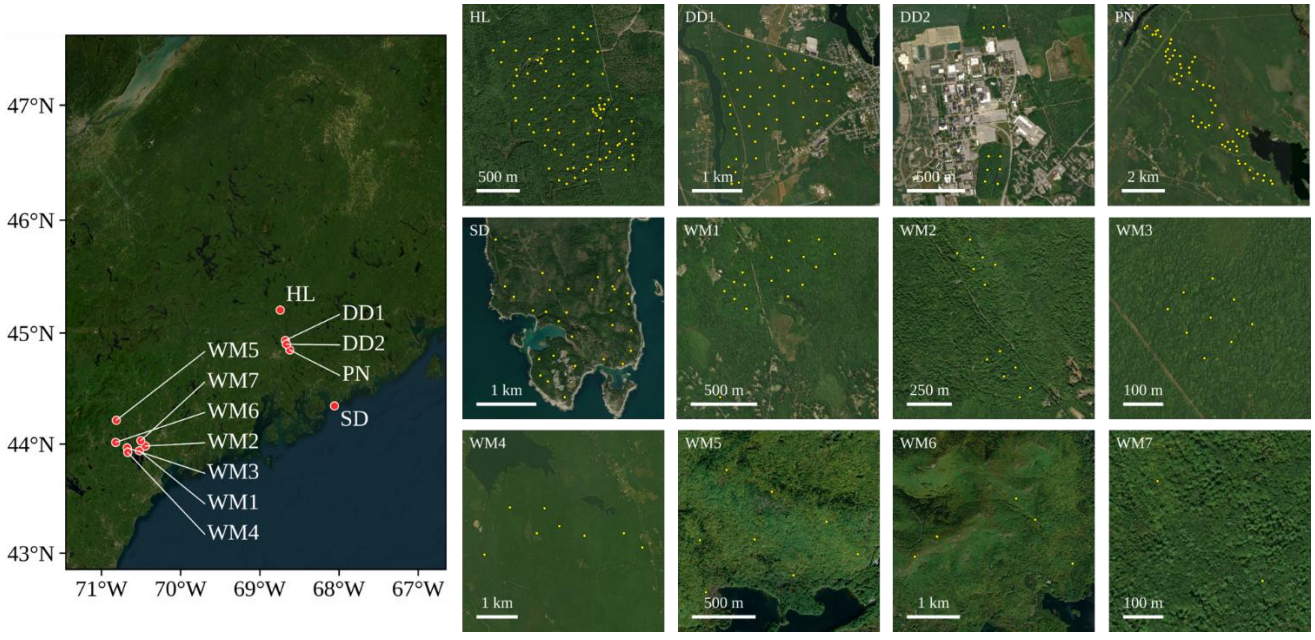


Figure 1. Locations of the five study forests and 308 inventory plots. The left panel shows the locations of the study forests in Maine, USA (red circles); the remaining panels show the locations of individual inventory plots within each forest (yellow circles). HL, Howland Research Forest; DD, Dwight B. Demeritt Forest; PN, Penobscot Experimental Forest; SD, Schoodic Peninsula Forest; WM, Western Maine Forest. DD and WM plots are located in separate areas (DD1 and DD2; WM1 to WM7). The map was created using the Esri World Imagery basemap layer. Sources: Esri, Vantor, Earthstar Geographics, and the GIS User Community | Powered by Esri.

2.2 Data collection

Forest inventories were conducted during the summers of 2024, 2025, and 2026 using a fixed-area sampling design across five forest ecosystems. A total of 308 circular plots were established, including 95 in Howland Research Forest, 57 in Dwight B. Demeritt Forest, 65 in Penobscot Experimental Forest, 30 in Schoodic Peninsula Forest, and 61 in Western Maine Forest (Fig. 1). Plots were systematically and uniformly distributed to capture spatial variability in stand structure and

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

species composition within each study site. Each plot was defined as a circular fixed-area plot with a radius of 8.92 m (area = 250 m^2). Within each plot, all live and standing dead trees (snags) with a diameter at breast height (DBH) ≥ 10 cm were measured and recorded (Fig. 2a). DBH was measured using a calibrated diameter tape. Tree species were identified in the field based on observer expertise, with assistance from experienced field technicians and relevant identification references as needed. The azimuth from the plot center to each tree was measured using a compass corrected for a magnetic declination of 15° and recorded clockwise from true north, ranging from 0° to 359°. The horizontal distance was obtained using a Haglöf Vertex IV hypsometer, which has a reported distance error rate of ≤ 1% (Haglöf, 2007). The hypsometer was calibrated daily using a distance tape prior to field measurements to minimize potential errors associated with environmental conditions such as temperature and humidity (Haglöf, 2007).

The geolocation of each plot center was measured using a Trimble Geo 7X handheld GNSS receiver connected to a Trimble Zephyr Model 2 antenna mounted on a survey tripod (Fig. 2b). The Geo 7X supports simultaneous tracking of GPS, GLONASS, and Galileo constellations and can achieve post-processed horizontal accuracy of 0.30-0.50 m under open-sky conditions and less than 1 m under forest canopy (Trimble, 2013). The receiver tracks up to 216 channels, while the Zephyr Model 2 antenna uses a choke-ring design to reduce multipath effects (Trimble, 2013). At the beginning of each inventory day, the tripod was placed in an open area to establish connections with more than 25 satellites, which typically required about 20 minutes. The system was then moved to each plot center without shutting down the receiver to maintain satellite lock. At each plot center, at least 175 positional readings were collected over approximately 5 minutes. This occupation time represented a practical compromise for operational forest inventory. Although longer occupation times may improve absolute GNSS accuracy (Andersen et al., 2022), under dense canopy they do not necessarily lead to further improvements in positioning accuracy. GNSS data were post-processed in Trimble Pathfinder Office (version 6.0) using differential correction from the nearest continuously operating reference station in the National Geodetic Survey network. This procedure reduced common-mode errors, including satellite clock drift, atmospheric delays, and orbital errors (Trimble, 2013). The corrected positions were averaged for each plot center, and the associated horizontal uncertainty was estimated from satellite geometry, number of satellites, receiver type, and related factors. All post-processed plot centers had estimated horizontal uncertainties below 1 m, confirming sub-meter horizontal accuracy.



https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Figure 2. Forest conditions and field equipment at Howland Research Forest. Forest near plot HL_001 in 2025 (a). Trimble Geo 7X handheld GNSS receiver connected to a Trimble Zephyr Model 2 antenna mounted on a survey tripod and used to determine the plot-center coordinates (b).

2.3 Forest attributes estimation

Based on the individual tree measurements, tree-level attributes were derived using the observed DBH and species information together with species-specific allometric equations. Tree height, aboveground biomass, and stem volume were estimated for each tree. Tree height was estimated using DBH-to-height allometric relationships developed by MacPhee et al. (2018) (Table S1), which are parameterized for northeastern forest species and suitable for mixed conifer and hardwood systems. These equations provide biologically consistent height estimates that account for interspecific differences in growth form and size structure. Aboveground biomass (AGB) at the individual tree level was calculated using multiple allometric models to ensure robustness and facilitate methodological comparison. We applied three DBH-based biomass equations developed by Young et al. (1980); Jenkins et al. (2003); Chojnacky et al. (2014) (Tables S2 to S4). These models estimate biomass as a function of DBH and species or species group and have been widely validated in temperate and boreal forest ecosystems. In addition, the allometric approach proposed by Westfall et al. (2024), which incorporates both DBH and tree height, was employed to account for vertical structural variation and to improve biomass estimation accuracy, particularly in structurally heterogeneous stands. Individual tree stem volume was estimated using species-specific volume equations based on DBH and the estimated tree height developed by Westfall et al. (2024). All biomass estimates were converted to carbon mass using the species-specific carbon fractions reported by Westfall et al. (2024) and the AGB values in the database are expressed in units of carbon (kgC/m^2 at the plot level). Tree density was calculated and expressed as trees per hectare (TPH) by expanding the number of measured trees within each fixed area plot to a per hectare basis. Plot-level attributes were calculated from individual tree estimates by averaging tree heights to obtain mean tree height and summing biomass and volume estimates to obtain total biomass and total volume, respectively.

To reconstruct tree- and plot-level AGB and total volume for each ALS acquisition year, we applied a backward DBH growth approach anchored to the field inventory measurements. For each measured tree, historical DBH was reconstructed by assigning a literature-informed annual diameter increment based on species group (Kenefic and Seymour, 2000; Seymour and Kenefic, 2002; Weiskittel et al., 2016; Kershaw et al., 2017) (Table S5). This approach provided a consistent basis for estimating the corresponding biomass and total volume values for the 2010, 2012, 2013, 2015, 2016, 2017, 2021, 2022, 2023, 2024, 2025, and 2026 ALS acquisitions (Eq. 1). For ALS acquisitions that were later than a plot's inventory year, the same species-group increment was applied forward in time.

$$DBH_{i,y} = \max\left[DBH_{i,inv} - g_{s(i)} \cdot (inv - y), 0\right] \qquad \text{Equation 1}$$

where $DBH_{i,inv}$ is the measured DBH in the inventory year (inv = 2024, 2025, or 2026) for tree i, $g_{s(i)}$ is the annual DBH increment (cm per year) assigned to the species group $s(i)$ (Table S5), and y represents the target ALS acquisition year. The

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



maximum function constrains reconstructed DBH to non-negative values. Trees with reconstructed DBH below the 10 cm inventory threshold were assigned zero biomass and zero total volume for that year, consistent with the field inventory design. Historical tree-level biomass was then estimated by scaling the inventory-year biomass estimate according to the ratio between reconstructed historical DBH and observed inventory-year DBH (Eq. 2).

$$AGB_{i,m,y} = AGB_{i,m,inv}\left(\frac{DBH_{i,y}}{DBH_{i,inv}}\right)^{\beta_{s(i)}} \quad \text{Equation 2}$$

where $AGB_{i,m,inv}$ is the field measurement-derived AGB in the inventory year for tree i from AGB estimator m (i.e., Young et al. (1980), Jenkins et al. (2003), Chojnacky et al. (2014), and Westfall et al. (2024)), and $\beta_{s(i)}$ is the species-group-specific allometric exponent (Table S5). This scaling was applied to each of the four AGB estimates included in the dataset to maintain consistency among historical AGB estimates. Historical total volume was reconstructed using the same DBH-ratio scaling framework (Eq. 3).

$$Volume_{i,y} = Volume_{i,inv}\left(\frac{DBH_{i,y}}{DBH_{i,inv}}\right)^{\beta_{s(i)}} \quad \text{Equation 3}$$

where $Volume_{i,inv}$ is the field measurement-derived total volume in the inventory year for tree i. Tree-level biomass and total volume for each target year were then aggregated to the plot level by summing over all trees i within each plot p, using the plot area A_p (250 m^2) (Eq. 4 and Eq. 5). Plot-level AGB was expressed in kgC/m^2 (Eq. 4), and the factor of 10,000 (Eq. 5) was used to convert plot-level total volume to m^3/ha.

$$AGB_{p,m,y} = \left(\frac{1}{A_p}\right)\sum_i AGB_{i,m,y} \quad \text{Equation 4}$$

$$Volume_{p,y} = \left(\frac{10000}{A_p}\right)\sum_i Volume_{i,y} \quad \text{Equation 5}$$

This approach provides a transparent and reproducible method for generating temporally aligned historical biomass and total volume estimates for ALS acquisition years while maintaining direct linkage to the measured tree inventory. The resulting estimates should be interpreted as approximate backcasts or forecasts rather than direct observations of past conditions or comprehensive predictions of future conditions. The method assumes a constant diameter increment within each species group and monotonic tree growth and does not explicitly account for recruitment, harvesting, storm disturbance, insect outbreaks, or mortality.

2.4 Airborne laser scanning data

ALS data acquired between 2010 and 2026 were used to characterize forest structural attributes across the five study forests. Multi-temporal datasets were obtained from the United States Geological Survey (USGS) lidar acquisitions funded by the American Recovery and Reinvestment Act of 2009 (USGS ARRA), the USGS 3D Elevation Program (USGS 3DEP) (Sugarbaker et al., 2014), and the NASA G-LiHT (Goddard's LiDAR, Hyperspectral, and Thermal) (Cook et al., 2013) airborne system (Table 1). The USGS ARRA data were collected in October 2010, before the 3DEP program began, under the federal lidar funding that led to the first USGS lidar base specification (Sugarbaker et al., 2014). These acquisitions have

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



a lower point density of approximately 1.5 points/m^2. The USGS 3DEP datasets acquired from 2013 to 2016 have lower point densities, ranging from approximately 5 to 6 points/m^2. In contrast, acquisitions from 2021 onward were collected at higher densities, ranging from approximately 8 to 27 points/m^2. NASA G-LiHT data are collected using a low-altitude airborne platform and typically provide high point densities, ranging from approximately 17 to 50 points/m^2. The USGS ARRA and 3DEP acquisitions were primarily conducted during hardwood leaf-off phenology conditions, which facilitate ground detection and more accurate terrain representation, whereas the NASA G-LiHT data were collected during leaf-on conditions, providing more detailed information on canopy structure and vertical complexity. In addition, we collected high-density ALS data in the summer of 2025 over Howland Research Forest, and in the spring of 2026 over Howland Research Forest and Dwight B. Demeritt Forest, using a Phoenix RANGER-U580 LiDAR system mounted on a crewed Cessna 172 aircraft, producing substantially denser point clouds (mean plot-level density up to 86.6 points/m^2) suitable for detailed canopy structural analysis (Fig. 3).

Table 1. Characteristics of the ALS acquisitions. Pulse density is the nominal pulse density (pulses/m^2), footprint is the laser footprint diameter (m) at the flying height, scan angle (°), and datum is the horizontal and vertical datum with the geoid model. A blank cell indicates that the information is unavailable.

Year	Source	Sensor	Pulse Density	Footprint	Scan Angle	Datum
2010	ARRA		0.25			NAD83; NAVD88 (GEOID09)
2012	G-LiHT	Riegl VQ-480	7.4	0.10	60	WGS 84; EGM96
2013	3DEP	Leica ALS70	2.0	0.44	40	NAD83(2011); NAVD88 (GEOID12A)
2015	3DEP	Leica ALS70	2.4	0.44	40	NAD83(2011); NAVD88 (GEOID12A)
2016	3DEP	Leica ALS70	2.2	0.46	40	NAD83(2011); NAVD88 (GEOID12B)
2017	G-LiHT	Riegl VQ-480	7.4	0.10	60	WGS 84; EGM96
2021	3DEP	Riegl VQ-1560i, VQ-780ii	3.0	0.34	58	NAD83(2011); NAVD88 (GEOID18)
2021	G-LiHT	Riegl VQ-480	7.4	0.10	60	WGS 84; EGM96
2022	3DEP	Riegl VQ-780i	4.7	0.24	60	NAD83(2011); NAVD88 (GEOID18)
2023	3DEP	Riegl VQ-1560ii-S	8.2	0.35	54	NAD83(2011); NAVD88 (GEOID18)
2024	3DEP	Riegl VQ-1460	2.0	0.50	58	NAD83(2011); NAVD88 (GEOID18)
2025	DigiSylva	Riegl VQ-580II-S	80.0	0.16	35	NAD83(2011); NAVD88 (GEOID18)
2026	DigiSylva	Riegl VQ-580II-S	20.0	0.21	35	NAD83(2011); ellipsoidal

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

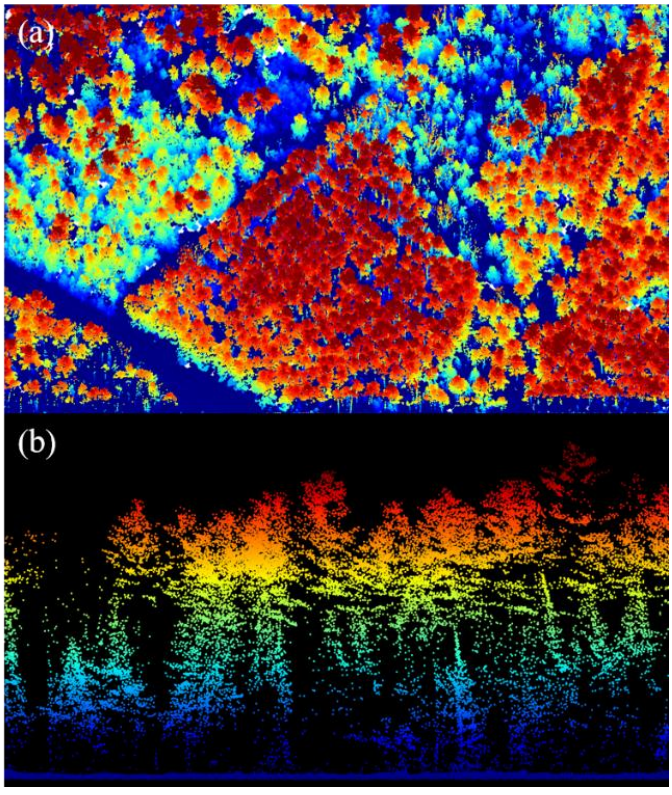


Figure 3. Airborne laser scanning point cloud of the Dwight B. Demeritt Forest, acquired on 26 April 2026 using a Phoenix RANGER-U580 LiDAR system. Oblique view of a 290 × 240 m area that includes a forest road (a). Vertical cross-sectional profile along a 56 m transect, extracted from a 34-meter wide swath within the same stand and displayed without vertical exaggeration (b). Returns are colored by height above ground, ranging from 0 m (blue) to ≥26 m (red). Point heights were normalized relative to a ground surface derived from ground-classified returns.

Based on the high positional accuracy plot center locations, plot-level point clouds were extracted to match the dimensions of the plots. All ALS returns within each circular plot were clipped and height-normalized by subtracting a plot-level ground elevation estimate, computed as the median elevation of the returns classified as ground within the plot, thereby generating height-above-ground (HAG) values for each return. When a plot-level point cloud contained fewer than three ground-classified returns, the elevation of the lowest return was used instead. Negative HAG values were set to zero. Returns more than 40 m above the ground were removed as noise, because no tree in these forests reaches that height. As all five study sites are forested areas without permanent built structures within the plot boundaries, no building masking was applied prior to HAG calculation. Plot-level structural metrics were then computed from the normalized point clouds, including ground elevation, total number of returns, plot-specific point density (points/m^2, calculated using the plot area), and the maximum and mean canopy heights (maximum and mean HAG). To characterize the vertical distribution of canopy structure, a set of

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.


height percentiles (p10, p20, p25, p30, p40, p50, p60, p70, p75, p80, and p90) was calculated from the HAG distribution, where each percentile represents the height below which a given proportion of ALS returns occurs (e.g., p90 corresponds to the height exceeded by the tallest 10% of returns). These percentile metrics provide a quantitative description of the vertical profile of forests, with lower percentiles reflecting near-ground and understory structure and higher percentiles representing upper canopy height and vertical complexity. In addition, the cumulative fraction of returns below each percentile threshold was computed to further describe the relative concentration of ALS returns along the vertical axis.

The same procedure was applied to each ALS acquisition, using the ground classification (class 2 in all data) provided by the source data, and no returns were reclassified. The USGS data were classified by the data providers according to the USGS project specifications, the G-LiHT data were classified by the G-LiHT processing chain, and the ground returns of the DigiSylva data were classified in SpatialExplorer (software). The acquisitions differ in horizontal datum, vertical datum, and geoid model (Table 1), and the data were used as delivered without changing these references. All acquisitions were projected in UTM zone 19N with meter units, so the plot circles were clipped directly from each acquisition without reprojection. The plot centers, the USGS data, and the DigiSylva data are referenced to NAD83 or NAD83(2011), whereas the G-LiHT data are referenced to WGS 84. The USGS data and the 2025 DigiSylva data report NAVD88 orthometric heights based on GEOID09, GEOID12A, GEOID12B, or GEOID18, the G-LiHT data report orthometric heights based on the EGM96 geoid, and the 2026 DigiSylva data report NAD83(2011) ellipsoidal heights. These differences do not affect the calculation of structural metrics, because the ground elevation and the HAG values were computed within each acquisition. For the same plot, the ground elevations of the NAVD88 acquisitions agree within 0.2 m (95^{th} percentile of the absolute differences), the G-LiHT ground elevations are on average 0.3 m lower, and the 2026 DigiSylva ground elevations are about 24.4 m lower because of the ellipsoidal reference. The ground elevation in the plot-level file is therefore reported in the vertical datum of each acquisition.

2.5 Optical observations

We compiled aerial imagery from the National Agriculture Imagery Program (NAIP) and multispectral satellite data from the NASA Harmonized Landsat and Sentinel-2 (HLS) product for each inventory plot. NAIP is acquired during the agricultural growing season and provides high-resolution orthorectified imagery across the conterminous United States (Maxwell et al., 2017). NAIP imagery typically has a spatial resolution of 0.6 to 1 m and commonly includes four spectral bands (red, green, blue, and near-infrared) (Allred et al., 2025). For each plot, the corresponding NAIP image was identified and clipped to a square window corresponding to the plot extent, with a 3 m buffer to ensure spatial consistency with the ALS point cloud extraction. NAIP imagery complements the structural information derived from ALS by providing high-resolution spectral and visual information that can support canopy interpretation, disturbance assessment, and land-cover characterization. While ALS captures three-dimensional canopy structure, NAIP imagery provides fine-resolution image context for each plot and its surrounding area. All NAIP imagery was retained in its original projection and spatial resolution to remain consistent with the publicly available NAIP products. For each plot, all available NAIP images acquired during the

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

leaf-on growing season in 2011, 2013, 2015, 2018, 2021, and 2023 were retained (six images per plot, 1,848 in total), and all pixels within the plot window were extracted (Fig. 4).

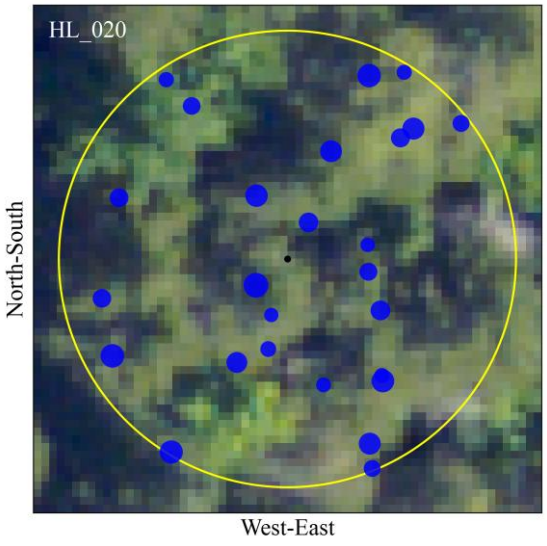
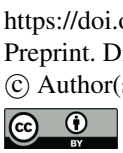


Figure 4. Example of NAIP imagery for Howland Research Forest plot 020 (HL_020).

In addition, we provided time series of multispectral satellite observations from the NASA HLS product. HLS provides harmonized surface reflectance data from Landsat (L30) and Sentinel-2 (S30) sensors on a common 30 m grid, enabling temporally consistent analysis across platforms (Claverie et al., 2018). For each plot, all growing-season (April to October) HLS scenes with reported cloud cover below 30% were collected from 2015 (L30) and 2016 (S30) through 2026, and each scene was clipped to a 7 × 7 pixel (210 m × 210 m) window centered on the plot center. Each clipped scene was stored as one stacked GeoTIFF containing all spectral bands, the Fmask quality layer used for per-pixel cloud and cloud-shadow screening (Zhu et al., 2015), and the solar and view geometry layers. The clipped scenes were organized by product (L30 and S30) and plot ID, with one folder per plot holding its full scene time series; in total, the database includes 66,412 plot-level HLS scenes (22,727 L30 and 43,685 S30). All HLS data were retained in the native 30 m grid and spatial reference framework. Because plot centers were surveyed with sub-meter GNSS accuracy, users can directly identify the HLS pixel containing each plot center and evaluate the spatial correspondence between field plots and 30 m satellite observations (Fig. 5). These co-located multispectral remote sensing data complement the ALS observations and support a wide range of applications linking ground-based measurements with airborne and satellite remote sensing (e.g., area-based AGB model development).

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

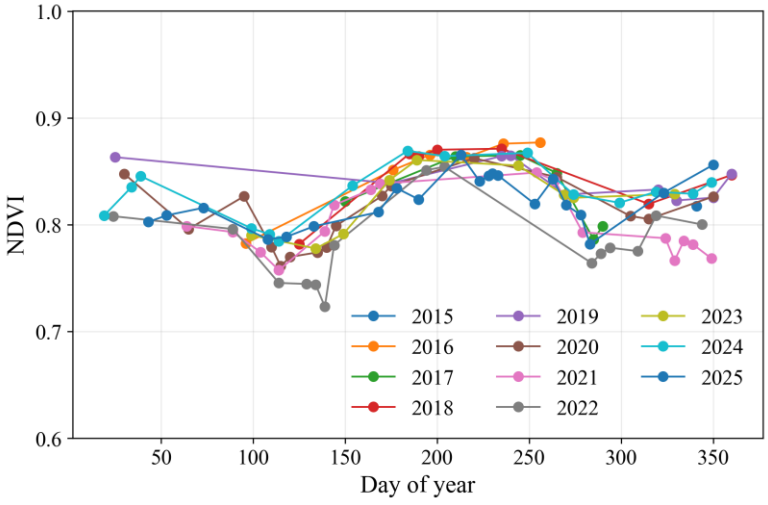


Figure 5. Example of seasonal dynamics in the normalized difference vegetation index (NDVI) derived from NASA Harmonized Landsat and Sentinel-2 (HLS) for Howland Research Forest plot 020 (HL_020).

3 Results

3.1 Database structure

The DigiSylva database is organized into three hierarchical components. The first level of the database consists of primary forest inventory data at the individual-tree level (Fig. 6). These data include directly recorded species, measured DBH (cm), and tree spatial position relative to the plot center, determined from azimuth (measured in degrees clockwise from north) and horizontal distance (m) from the center. The tree-level data are stored in a comma-separated values file *(DigiSylva_Tree_Data.csv)*. In addition to the measured attributes, estimated tree height, AGB, and volume are provided for each tree. A snag flag identifies standing dead trees. The second level of the database contains plot-level remote sensing data, including ALS point clouds in LAS format (.las), NAIP aerial imagery in GeoTIFF format (.tif), and HLS imagery in GeoTIFF format (.tif) (Fig. 6). These remote sensing data provide spatially aligned forest structural and spectral information for each plot. The third level of the database consists of plot-level summary metrics stored in a comma-separated values file *(DigiSylva_Plot_Data.csv)*. These include inventory-derived attributes such as TPH (trees per hectare), AGB (kgC/m^2), and stem volume (m^3/ha) for each plot. In addition, ALS-derived structural metrics are provided for each acquisition, including ground elevation (m), total number of returns, plot-level point density (points/m^2), maximum height (m), mean height (m), and a set of height percentiles (p10, p20, p25, p30, p40, p50, p60, p70, p75, p80, and p90). The variables in both data files are explained in Table S6.

Plot identifiers are standardized across all data components to ensure relational consistency. Each plot ID consists of a two-letter site code followed by a three-digit numeric identifier. The site codes are HL for Howland Research Forest, DD for

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Dwight B. Demeritt Forest, PN for Penobscot Experimental Forest, SD for Schoodic Peninsula Forest, and WM for Western Maine Forest. Plot numbers within each forest begin at 001. This unique plot ID serves as the primary key linking the tree-level inventory data, plot-level summaries, ALS point clouds, and associated remote sensing imagery.

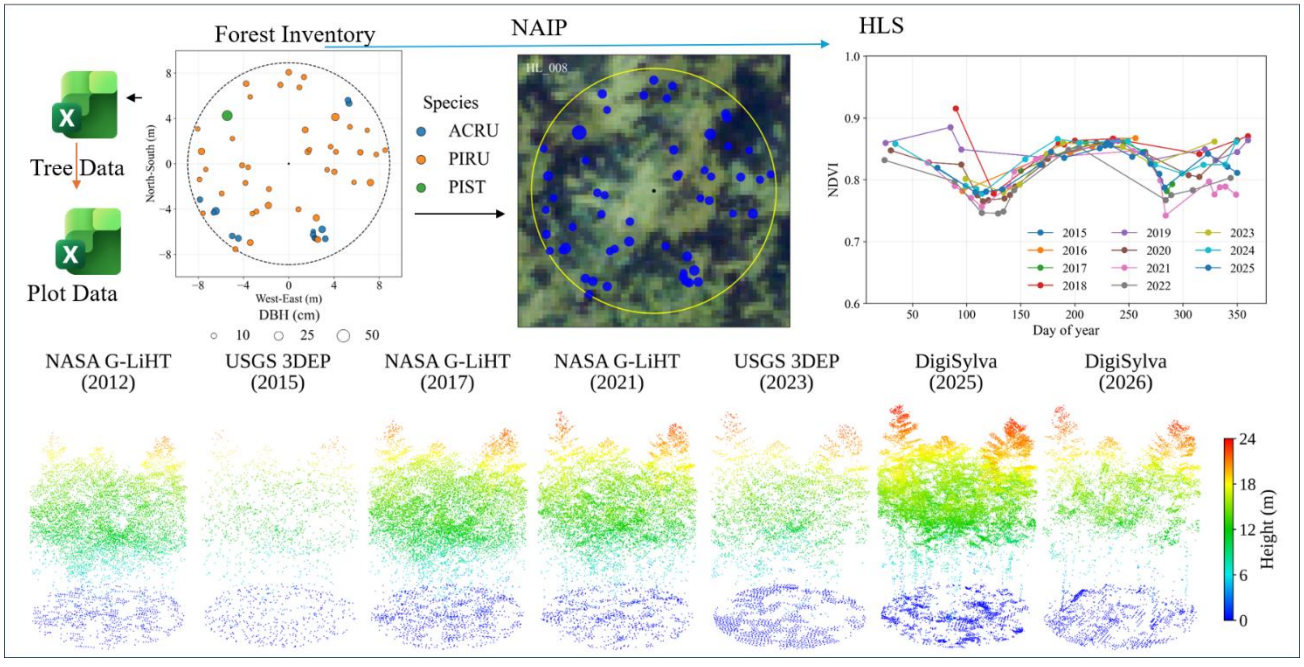


Figure 6. Structure of the DigiSylva database, including forest inventory data archived in the tree-level data file *(DigiSylva_Tree_Data.csv)*, plot-level summary metrics stored in the plot-level data file *(DigiSylva_Plot_Data.csv)*, NAIP imagery, Harmonized Landsat and Sentinel-2 (HLS) data, and airborne laser scanning point clouds. The example shown is for plot HL_008 in Howland Research Forest. ACRU: red maple (*Acer rubrum*); PIRU: red spruce (*Picea rubens*); PIST: eastern white pine (*Pinus strobus*).

3.2 Database summary

The database contains a total of 308 inventory plots distributed across five forest sites. Howland Research Forest (HL) includes 95 plots, Dwight B. Demeritt Forest (DD) includes 57 plots, Penobscot Experimental Forest (PN) includes 65 plots, Schoodic Peninsula Forest (SD) includes 30 plots, and Western Maine Forest (WM) includes 61 plots (Table 2). Across all sites, 6,137 individual trees were measured. HL includes 2,223 trees, DD includes 1,022 trees, PN includes 1,168 trees, SD includes 796 trees, and WM includes 928 trees (Table 2). Among all measured trees, 222 are snags. They are broadly distributed across the 308 plots rather than concentrated in a particular plot or region. The plot SD_022 contains the largest number of recorded trees, with 61 individuals, followed by SD_026 with 56 individuals. Four plots, HL_022, HL_023, PN_037, and PN_062, contained no trees with DBH ≥ 10 cm; therefore, no trees were recorded in these plots under the

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



inventory protocol. The maximum plot-level AGB was observed in DD_055, with a mean value of 21.0 kgC/m^2 estimated as the average across four allometric approaches. The maximum volume was observed in DD_055, with a value of 801.6 m^3/ha.

Table 2. Summary of forest inventory data included in this dataset for the five forest ecosystems. HL, Howland Research Forest; DD, Dwight B. Demeritt Forest; PN, Penobscot Experimental Forest; SD, Schoodic Peninsula Forest; and WM, Western Maine Forest. Trees/Plot$_{max}$ is the maximum number of trees recorded in a single plot at each site. DBH$_{max}$ (cm), and Height$_{max}$ (m) represent the maximum individual-tree values, whereas AGB$_{max}$ (kgC/m^2, mean of the four allometric methods), and Volume$_{max}$ (m^3/ha) represent the maximum plot-level values at each site.

	HL	DD	PN	SD	WM	Total
Plots	95	57	65	30	61	308
Trees	2,223	1,022	1,168	796	928	6,137
Trees/Plot$_{max}$	55	43	51	61	31	61
DBH$_{max}$	82.0	81.9	106.6	62.5	68.5	106.6
Height$_{max}$	28.2	28.2	30.8	23.8	26.2	30.8
AGB$_{max}$	15.1	21.0	18.8	16.5	13.1	21.0
Volume$_{max}$	552.4	801.6	704.4	597.2	402.0	801.6

A total of 1,408 plot-level ALS point cloud files were compiled across all sites and acquisition years (Table 3). Every plot is covered by at least two acquisitions, with 4.6 acquisitions on average. At Howland Research Forest, 80 plots have seven repeated ALS acquisitions in 2012, 2015, 2017, 2021, 2023, 2025, and 2026, and the remaining 15 plots have five acquisitions (2012, 2015, 2023, 2025, and 2026). At Dwight B. Demeritt Forest, 44 plots have four acquisitions (2010, 2021 3DEP, 2021 G-LiHT, and 2026), 6 plots have three (2010, 2021 3DEP, and 2026), 4 plots have two (2021 G-LiHT and 2026), and 3 plots have five (2010, 2017, 2021 3DEP, 2021 G-LiHT, and 2026). At Penobscot Experimental Forest, plots have between four and seven acquisitions drawn from 2010, 2012, 2015, 2017, 2021 3DEP, 2021 G-LiHT, and 2023, with 43 plots covered by 2010, 2012, 2017, 2021 3DEP, and 2021 G-LiHT. At Schoodic Peninsula Forest, all 30 plots have three ALS acquisitions in 2010, 2017, and 2021. At Western Maine Forest, 52 plots have two acquisitions in 2013 and 2022, 8 plots have two acquisitions in 2016 and 2024, and 1 plot has three acquisitions in 2013, 2016, and 2022. These multi-temporal ALS acquisitions provide repeated structural observations that enable analysis of canopy development, disturbance, and structural change over time.

The earliest datasets were acquired in 2010 through the USGS ARRA program, covering Dwight B. Demeritt Forest, Penobscot Experimental Forest, and Schoodic Peninsula Forest, with point densities ranging from 1.3 to 1.5 points/m^2. The USGS 3DEP acquisitions provided intermediate densities, from 5.4 points/m^2 in 2015 to 27.2 points/m^2 in 2023. NASA G-LiHT acquisitions between 2012 and 2021 provided the highest airborne survey densities among the public sources, ranging from 17.4 to 49.5 points/m^2 depending on site and year. The highest point density was obtained in 2025 at Howland Research Forest from the acquisition with the Phoenix RANGER-U580 airborne system, with a mean density of 86.6 points/m^2 and a mean canopy height of 14.3 m. These datasets span a range of acquisition conditions and sensor

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



configurations, providing variability in point density and canopy representation that supports structural comparison across sites and years.

Table 3. Summary of multi-temporal airborne laser scanning (ALS) point clouds included in this dataset, including acquisition year, data source, forest site coverage, number of plots, mean point density (points/m^2), and mean canopy height (m).

Year	Data Source	Coverage	Number of Plots	Point Density (points/m^2)	Mean Height (m)
2010	ARRA	Dwight B. Demeritt Forest	53	1.5	9.53
		Penobscot Experimental Forest	58	1.4	8.06
		Schoodic Peninsula Forest	30	1.3	5.43
2012	G-LiHT	Howland Research Forest	95	28.9	12.42
		Penobscot Experimental Forest	65	27.7	11.40
2013	3DEP	Western Maine Forest	53	5.6	5.41
2015	3DEP	Howland Research Forest	95	5.9	10.51
		Penobscot Experimental Forest	11	5.4	6.75
2016	3DEP	Western Maine Forest	9	6.3	3.02
2017	G-LiHT	Howland Research Forest	80	40.3	12.50
		Dwight B. Demeritt Forest	3	17.4	17.87
		Penobscot Experimental Forest	65	49.2	11.40
		Schoodic Peninsula Forest	30	46.0	6.28
2021	3DEP	Dwight B. Demeritt Forest	53	9.2	10.20
		Penobscot Experimental Forest	65	7.9	8.07
		Schoodic Peninsula Forest	30	9.5	5.95
2021	G-LiHT	Howland Research Forest	80	40.7	12.67
		Dwight B. Demeritt Forest	51	48.7	12.93
		Penobscot Experimental Forest	65	49.5	10.94
2022	3DEP	Western Maine Forest	53	15.1	5.58
2023	3DEP	Howland Research Forest	95	27.2	11.31
		Penobscot Experimental Forest	14	25.9	7.72
2024	3DEP	Western Maine Forest	8	9.9	5.70
2025	DigiSylva	Howland Research Forest	95	86.6	14.34
2026	DigiSylva	Howland Research Forest	95	21.9	13.84
		Dwight B. Demeritt Forest	57	17.9	11.06

To complement the repeated ALS observations, we also provided reconstructed tree- and plot-level aboveground biomass (AGB) and volume estimates for each ALS acquisition year (2010, 2012, 2013, 2015, 2016, 2017, 2021, 2022, 2023, 2024, 2025, and 2026). For each tree, DBH at a prior year was estimated using a backward growth approach in which the observed DBH was reduced by the cumulative species-group-specific annual diameter increment between the inventory year and the target year (for acquisition years later than the inventory year, the increment was applied forward). Trees with reconstructed DBH below the 10 cm inventory threshold were assigned zero biomass and zero volume for that year. Historical biomass and volume were then calculated by scaling the corresponding inventory-year tree-level estimates according to the ratio between

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



reconstructed and observed DBH using species-group allometric exponents. Plot-level AGB and volume for each year were subsequently obtained by summing the reconstructed tree-level values within each plot and expressing them as kgC/m^2 and m^3/ha, respectively. These retrospective estimates provide temporally matched forest attribute layers for direct comparison with the multi-temporal ALS observations included in DigiSylva.

3.3 Database application

To illustrate the scientific applications of the DigiSylva database, we present an example analysis using the released plot-level forest attributes, ALS structural metrics, and tree-level inventory data. The example demonstrates the use of DigiSylva for area-based modeling of AGB (White et al., 2017). We calibrated an area-based AGB prediction model using Penobscot Experimental Forest (PN) plots with complete 2021 ALS summary metrics (NASA G-LiHT data) and released plot-level AGB estimates in 2021. The predictor variables included the 50th, 70th, and 90th height percentiles (p50, p70, and p90), mean canopy height (H_{mean}), maximum canopy height (H_{max}), and plot-level ALS point density (D). A random forest regression model with 500 trees was fitted using the PN plots and then transferred to Howland Research Forest using the same set of 2021 ALS metrics (NASA G-LiHT data). The calibration results for PN plots (n = 65) showed no statistically significant difference between observed and predicted AGB based on a two-tailed t-test (p-value = 0.752), with an RMSE of 0.97 kgC/m^2 and a relative bias of -0.6%. When the calibrated model was transferred to HL plots (n = 80), it retained reasonable predictive performance, with an RMSE of 2.64 kgC/m^2, a relative bias of 0.1%, and no statistically significant difference between observed and predicted AGB based on a two-tailed t-test (p-value = 0.982; Fig. 7). This example demonstrates that DigiSylva can support calibration of area-based ALS models for estimating plot-level AGB using the released structural metrics.

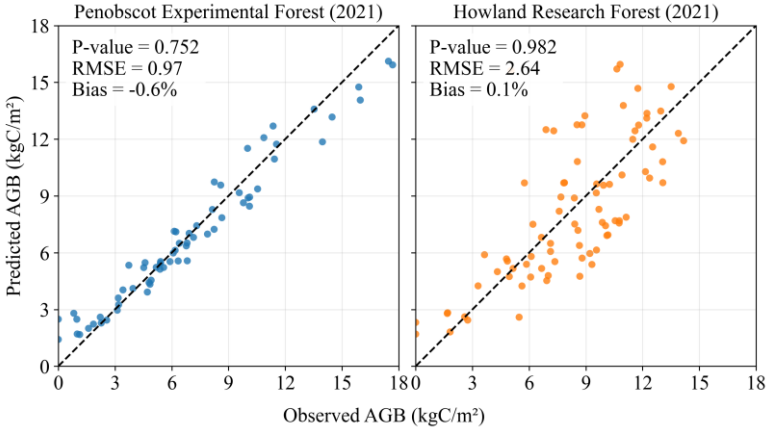


Figure 7. Application of area-based AGB modeling using 2021 ALS summary metrics and plot-level AGB estimates. A random forest model was calibrated using Penobscot Experimental Forest plots (n = 65) and then transferred to Howland Research Forest plots (n = 80) for cross-site evaluation. The dashed line represents the 1:1 relationship between observed and predicted AGB. The reported p-value indicates the result of a two-tailed t-test comparing observed and predicted AGB.

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

4 Discussion

This study presents a spatially explicit, multi-source forest inventory with co-located ALS and multispectral airborne and satellite observations spanning five representative forest ecosystems within the boreal-temperate transition zone of the northeastern United States. The significance of this dataset lies in its precise spatial alignment between detailed field measurements and multi-temporal ALS point clouds, supported by sub-meter plot geolocation accuracy. Accurate co-registration between forest inventory plots and remotely sensed data is a critical requirement for reliable ALS-based model calibration, structural analysis, and carbon estimation. Many publicly available forest inventory datasets either lack precise plot coordinates due to confidentiality constraints or do not provide direct linkage to corresponding ALS point clouds. By integrating individual tree measurements, plot-level summaries, normalized ALS structural metrics, and co-registered aerial imagery within a unified relational framework, this database provides a comprehensive reference dataset for mixed temperate and boreal forest ecosystems. The inclusion of repeated ALS acquisitions from 2010 to 2026 further enhances the scientific value of the dataset by enabling structural comparison across years and sensor configurations. The five study forests collectively represent managed, experimental, inland, coastal, and northern hardwood ecosystems, capturing structural heterogeneity and compositional variability characteristic of the northeastern United States. As mixed spruce-hardwood forests in the boreal-temperate transition zone remain underrepresented in open-access inventory and ALS datasets, this database fills an important regional data gap and provides a high-quality benchmark for forest-related studies.

The potential applications of this dataset extend across multiple research and management domains. First, the co-located tree-level measurements, plot-level forest attributes, and ALS-derived structural metrics provide a robust foundation for calibrating and validating biomass and volume prediction models using statistical regression, machine learning, or other area-based modeling approaches. Because tree spatial positions relative to plot centers are included, the dataset can also support individual-tree detection, crown segmentation, and analyses of spatial forest structure, particularly for plots with high-density ALS data. The availability of multi-temporal ALS acquisitions enables investigations of canopy structural dynamics, disturbance effects, and forest development trajectories over time. Differences among acquisitions in point density, seasonality, and sensor platform further provide opportunities to evaluate how ALS configuration influences derived structural metrics and forest attribute estimation. In addition, the inclusion of NAIP imagery and HLS multispectral time series provides complementary passive optical information that may support land-cover interpretation, canopy condition assessment, spectral-index analysis, and integration of structural and spectral predictors. Beyond biomass and volume modeling, DigiSylva can contribute to studies of vertical complexity, forest stratification, habitat structure, and stand structural heterogeneity. The plot-level structural metrics and tree-level attributes may provide useful calibration and evaluation data for ecosystem models and data assimilation frameworks. Because structural metrics were derived independently for each ALS acquisition year while preserving consistent plot identifiers, the dataset supports reproducible cross-year comparisons and methodological testing. Together, these characteristics make DigiSylva suitable for applied forest management analysis, remote-sensing model development, and fundamental ecological research.

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.


In the future, DigiSylva is intended to serve as a long-term, evolving resource rather than a static data product. Continued field inventory and ALS campaigns are planned to expand spatial coverage within existing study sites and to incorporate additional forest types across broader climatic, ecological, and management gradients. Repeated plot measurements will enable direct evaluation of tree growth, mortality, harvesting, disturbance impacts, and biomass change, thereby strengthening temporal validation of ALS-based models and improving estimates of forest carbon dynamics. Additional high-density ALS data will be collected across expanded forest ecosystems using the Phoenix RANGER-U580 airborne LiDAR system. Database maintenance will include version control, metadata documentation, standardized processing pipelines, and quality-control procedures for both field and remote sensing data to ensure consistency, reproducibility, and transparency across future releases. Code used for dataset processing and example applications will be updated as methods evolve, enabling users to replicate, evaluate, or extend the workflow. By combining continued field measurements, multi-temporal ALS acquisitions, optical imagery, methodological transparency, and open data maintenance, DigiSylva provides a scalable framework for sustained monitoring of forest structure, total volume, and biomass dynamics in transitional forest ecosystems and can support ongoing and future NASA and ESA missions, including GEDI, ICESat-2, NISAR, BIOMASS, and EDGE.

5 Conclusion

DigiSylva provides co-located forest inventory measurements and multi-temporal ALS point clouds for 308 plots in the boreal-temperate ecotone of the northeastern United States. The database includes 6,137 measured trees with estimated height, biomass, and volume, 1,408 plot-level ALS point clouds acquired from 2010 to 2026 with standardized structural metrics, and co-registered NAIP and HLS imagery. Sub-meter plot positioning allows direct alignment between the field and remote sensing data. Biomass and volume reconstructed for each ALS acquisition year support model calibration across sensors and years. The database will be expanded with new inventories and ALS campaigns.

Code and data availability

The DigiSylva database is available through the Harvard Dataverse at https://doi.org/10.7910/DVN/A3EUOQ (Wei et al., 2026b).

Author contributions

XW and LM conceptualized and designed this study. GD led the forest inventory. JZ calculated the volume and biomass. AO contributed to the aboveground biomass calculation. SK downloaded and processed the airborne laser scanning (ALS) data.

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



LM processed the NAIP and HLS data. DS collected and processed the ALS data. XW summarized all the data and developed the database. XW wrote the original manuscript. XW, DH, LM, and AW supervised the project. XW, DH, LM, JZ, GD, JW, AO, DS, SK, CZ, GH, and AW contributed to revising and editing the manuscript.

Competing interests

The authors declare no conflict of interest.

Acknowledgements

We appreciate the contributions of Gregory McHale, Peter Howe, Christian Hettwer, Zachary Whitehead, Isabella Pontecorvo, and Christian West-McCauley to the field inventory. During the preparation of this manuscript, the authors used Anthropic Claude and OpenAI ChatGPT to refine the language and improve readability.

Financial support

This work was supported by NASA Carbon Monitoring System grants (80NSSC25K7219, 80NSSC25K7221), USDA Forest Service Northeastern States Research Cooperative grant (24-DG-11242311-081), Maine Space Grant Consortium grant (SG-27-03), USDA NIFA Sustainable Agricultural Systems grant (2023-68012-38992), GEDI Competed Science Team (80NSSC24K0599), and NASA Early Career Investigator Program in Earth Science (80NSSC24K1632).

References

Allred, B. W., McCord, S. E., and Morford, S. L.: Canopy height model and NAIP imagery pairs across CONUS, Scientific Data, 12, 322, https://doi.org/10.1038/s41597-025-04655-z, 2025.

Andersen, H.-E., Strunk, J., and McGaughey, R. J.: Using high-performance global navigation satellite system technology to improve Forest Inventory and Analysis plot coordinates in the Pacific region, Gen. Tech. Rep. PNW-GTR-1000, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR, USA, https://doi.org/10.2737/PNW-GTR-1000, 2022.

Atkins, J. W., Aho, K. S., Chen, X., Elmore, A. J., Fiorella, R., Luo, W., Lombardozzi, D., Lunch, C., Manak, L., de Pablo, L. X., Myers-Pigg, A. N., Record, S., Qiu, T., Reed, S., Ruddell, B., Strange, B., Torrens, C. L., Yule, K., and Richardson, A. D.: Recommendations for developing, documenting, and distributing data products derived from NEON data, Ecosphere, 16, e70159, https://doi.org/10.1002/ecs2.70159, 2025.

Atkins, J. W., Costanza, J., Dahlin, K. M., Dannenberg, M. P., Elmore, A. J., Fitzpatrick, M. C., Hakkenberg, C. R., Hardiman, B. S., Kamoske, A., LaRue, E. A., Silva, C. A., Stovall, A. E. L., and Tielens, E. K.: Scale dependency of lidar-derived forest structural diversity, Methods in Ecology and Evolution, 14, 708–723, https://doi.org/10.1111/2041-210X.14040, 2023.

Chojnacky, D. C., Heath, L. S., and Jenkins, J. C.: Updated generalized biomass equations for North American tree species, Forestry, 87, 129–151, https://doi.org/10.1093/forestry/cpt053, 2014.

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., and Justice, C.: The Harmonized Landsat and Sentinel-2 surface reflectance data set, Remote Sensing of Environment, 219, 145–161, https://doi.org/10.1016/j.rse.2018.09.002, 2018.

Cook, B. D., Corp, L. A., Nelson, R. F., Middleton, E. M., Morton, D. C., McCorkel, J. T., Masek, J. G., Ranson, K. J., Ly, V., and Montesano, P. M.: NASA Goddard's LiDAR, Hyperspectral and Thermal (G-LiHT) Airborne Imager, Remote Sensing, 5, 4045–4066, https://doi.org/10.3390/rs5084045, 2013.

Coops, N. C., Irwin, L. A., Seely, H. S., and Hardy, S. J.: Advances in laser scanning to assess carbon in forests: From ground-based to space-based sensors, Current Forestry Reports, 11, 11, https://doi.org/10.1007/s40725-024-00242-4, 2025.

Gillis, M. D., Omule, A. Y., and Brierley, T.: Monitoring Canada's forests: The National Forest Inventory, The Forestry Chronicle, 81, 214–221, https://doi.org/10.5558/tfc81214-2, 2005.

Guo, Q., Su, Y., Hu, T., Guan, H., Jin, S., Zhang, J., Zhao, X., Xu, K., Wei, D., Kelly, M., and Coops, N. C.: Lidar boosts 3D ecological observations and modelings: A review and perspective, IEEE Geoscience and Remote Sensing Magazine, 9, 232–257, https://doi.org/10.1109/MGRS.2020.3032713, 2021.

Haglöf: Users Guide Vertex IV and Transponder T3, Haglöf Sweden AB, Långsele, Sweden, 2007.

Hollinger, D. Y., Davidson, E. A., Fraver, S., Hughes, H., Lee, J., Richardson, A. D., Savage, K., Sihi, D., and Teets, A.: Multi-decadal carbon cycle measurements indicate resistance to external drivers of change at the Howland Forest AmeriFlux site, Journal of Geophysical Research: Biogeosciences, 126, e2021JG006276, https://doi.org/10.1029/2021JG006276, 2021.

Huang, W., Dolan, K., Swatantran, A., Johnson, K., Tang, H., O'Neil-Dunne, J., Dubayah, R., and Hurtt, G.: High-resolution mapping of aboveground biomass for forest carbon monitoring system in the Tri-State region of Maryland, Pennsylvania and Delaware, USA, Environmental Research Letters, 14, 095002, https://doi.org/10.1088/1748-9326/ab2917, 2019.

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., and Birdsey, R. A.: National-scale biomass estimators for United States tree species, Forest Science, 49, 12–35, https://doi.org/10.1093/forestscience/49.1.12, 2003.

Kangas, A. and Maltamo, M. (Eds.): Forest Inventory: Methodology and Applications, Managing Forest Ecosystems, vol. 10, Springer, Dordrecht, the Netherlands, https://doi.org/10.1007/1-4020-4381-3, 2006.

Kenefic, L. S. and Brissette, J. C.: History of the Penobscot Experimental Forest, 1950–2010, in: Penobscot Experimental Forest: 60 years of research and demonstration in Maine, 1950–2010, edited by: Kenefic, L. S. and Brissette, J. C., Gen. Tech. Rep. NRS-P-123, U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA, USA, 1–20, 2014.

Kenefic, L. S. and Seymour, R. S.: Growth patterns of Tsuga canadensis in managed uneven-aged northern conifer stands, in: Proceedings: Symposium on Sustainable Management of Hemlock Ecosystems in Eastern North America, edited by: McManus, K. A., Shields, K. S., and Souto, D. R., Gen. Tech. Rep. NE-267, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA, USA, 29–33, 2000.

Kershaw, J. A., Weiskittel, A. R., Lavigne, M. B., and McGarrigle, E.: An imputation/copula-based stochastic individual tree growth model for mixed species Acadian forests: a case study using the Nova Scotia permanent sample plot network, Forest Ecosystems, 4, 15, https://doi.org/10.1186/s40663-017-0102-2, 2017.

Knapp, N., Fischer, R., Cazcarra-Bes, V., and Huth, A.: Structure metrics to generalize biomass estimation from lidar across forest types from different continents, Remote Sensing of Environment, 237, 111597, https://doi.org/10.1016/j.rse.2019.111597, 2020.

MacPhee, C., Kershaw, J. A., Weiskittel, A. R., Golding, J., and Lavigne, M. B.: Comparison of approaches for estimating individual tree height–diameter relationships in the Acadian forest region, Forestry: An International Journal of Forest Research, 91, 132–146, https://doi.org/10.1093/forestry/cpx039, 2018.

Maxwell, A. E., Warner, T. A., Vanderbilt, B. C., and Ramezan, C. A.: Land cover classification and feature extraction from National Agriculture Imagery Program (NAIP) Orthoimagery: A review, Photogrammetric Engineering & Remote Sensing, 83, 737–747, https://doi.org/10.14358/PERS.83.10.737, 2017.

McRoberts, R. E. and Tomppo, E. O.: Remote sensing support for national forest inventories, Remote Sensing of Environment, 110, 412–419, https://doi.org/10.1016/j.rse.2006.09.034, 2007.

McRoberts, R. E., Tomppo, E. O., and Næsset, E.: Advances and emerging issues in national forest inventories, Scandinavian Journal of Forest Research, 25, 368–381, https://doi.org/10.1080/02827581.2010.496739, 2010.

Mori, A. S., Lertzman, K. P., and Gustafsson, L.: Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology, Journal of Applied Ecology, 54, 12–27, https://doi.org/10.1111/1365-2664.12669, 2017.

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Pan, Y., Birdsey, R. A., Phillips, O. L., Houghton, R. A., Fang, J., Kauppi, P. E., Keith, H., Kurz, W. A., Ito, A., Lewis, S. L., Nabuurs, G.-J., Shvidenko, A., Hashimoto, S., Lerink, B., Schepaschenko, D., Castanho, A., and Murdiyarso, D.: The enduring world forest carbon sink, Nature, 631, 563–569, https://doi.org/10.1038/s41586-024-07602-x, 2024.

Pascual, C., García-Montero, L. G., Arroyo, L. A., and García-Abril, A.: Increasing the use of expert opinion in forest characterisation approaches based on LiDAR data, Annals of Forest Science, 70, 87–99, https://doi.org/10.1007/s13595-012-0232-1, 2013.

Poulter, B., Murray-Tortarolo, G., Hayes, D. J., Ciais, P., Andrew, R. M., Bastos, A., Byrne, B., Butman, D., Canadell, J. G., Chatterjee, A., Domke, G., Feldman, A., Foster, K., Hunka, N., Jackson, R. B., Kurz, W. A., Lindquist, A., Liu, M., Luijkx, I., Mialon, A., Michalak, A. M., Miller, J., Obermeier, W. A., Pan, N., Randerson, J. T., Raymond, P. A., Regnier, P., Resplandy, L., Rocher-Ros, G., Rodriguez-Fernandez, N., Rosentreter, J., Salazar-Neira, J. C., Tank, S. E., Tian, H., Vargas, R., Villalobos, Y., Wang, J. A., Wei, X., Wickland, K. P., Williams, C., Windham-Myers, L., Woodall, C., Ying, Q., and Zhang, Z.: The North American Greenhouse Gas Budget: Emissions, Removals, and Integration for CO_2, CH_4, and N_2O (2010–2019): Results From the Second REgional Carbon Cycle Assessment and Processes Study (RECCAP2), Global Biogeochemical Cycles, 39, e2024GB008310, https://doi.org/10.1029/2024GB008310, 2025.

Richardson, A. D., Hollinger, D. Y., Shoemaker, J. K., Hughes, H., Savage, K., and Davidson, E. A.: Six years of ecosystem-atmosphere greenhouse gas fluxes measured in a sub-boreal forest, Scientific Data, 6, 117, https://doi.org/10.1038/s41597-019-0119-1, 2019.

Rogers, N. S. and Kenefic, L. S.: The cutting practice level study at the Penobscot Experimental Forest in Maine, Journal of Forestry, 115, 58–61, https://doi.org/10.5849/jof.16-006, 2017.

Savage, K., Davidson, E. A., Richardson, A. D., and Hollinger, D. Y.: Three scales of temporal resolution from automated soil respiration measurements, Agricultural and Forest Meteorology, 149, 2012–2021, https://doi.org/10.1016/j.agrformet.2009.07.008, 2009.

Scholl, V. M., Cattau, M. E., Joseph, M. B., and Balch, J. K.: Integrating National Ecological Observatory Network (NEON) airborne remote sensing and in-situ data for optimal tree species classification, Remote Sensing, 12, 1414, https://doi.org/10.3390/rs12091414, 2020.

Seymour, R. S. and Kenefic, L. S.: Influence of age on growth efficiency of Tsuga canadensis and Picea rubens trees in mixed-species, multiaged northern conifer stands, Canadian Journal of Forest Research, 32, 2032–2042, https://doi.org/10.1139/x02-120, 2002.

Sugarbaker, L. J., Constance, E. W., Heidemann, H. K., Jason, A. L., Lukas, V., Saghy, D. L., and Stoker, J. M.: The 3D Elevation Program initiative: a call for action, U.S. Geological Survey Circular 1399, U.S. Geological Survey, Reston, VA, USA, https://doi.org/10.3133/cir1399, 2014.

Tang, H., Ma, L., Lister, A., O'Neil-Dunne, J., Lu, J., Lamb, R. L., Dubayah, R., and Hurtt, G.: High-resolution forest carbon mapping for climate mitigation baselines over the RGGI region, USA, Environmental Research Letters, 16, 035011, https://doi.org/10.1088/1748-9326/abd2ef, 2021.

Tinkham, W. T., Mahoney, P. R., Hudak, A. T., Domke, G. M., Falkowski, M. J., Woodall, C. W., and Smith, A. M. S.: Applications of the United States Forest Inventory and Analysis dataset: a review and future directions, Canadian Journal of Forest Research, 48, 1251–1268, https://doi.org/10.1139/cjfr-2018-0196, 2018.

Trimble: GPS Pathfinder Office Software, Trimble Navigation Limited, Sunnyvale, CA, USA, 2013.

University of Maine: UMaine's Forest Land Resources, School of Forest Resources, University of Maine, Orono, ME, USA, https://forest.umaine.edu/forest-land-resources/ (last access: 18 November 2026), 2026.

Vidal, C., Alberdi, I. A., Hernández Mateo, L., and Redmond, J. J. (Eds.): National Forest Inventories: Assessment of Wood Availability and Use, Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-319-44015-6, 2016.

Wei, X., Hayes, D. J., Weiskittel, A., and Zhao, J.: Warming-driven shifts in dominant tree species potentially reduce aboveground biomass in northeastern United States forests, Forest Ecology and Management, 580, 122536, https://doi.org/10.1016/j.foreco.2025.122536, 2025.

Wei, X., Hayes, D., McHale, G., Zhao, J., Weiskittel, A., and Daigneault, A.: PyFIA: analyzing and visualizing forest attributes using the United States Forest Inventory and Analysis database, Carbon Balance and Management, 21, 18, https://doi.org/10.1186/s13021-025-00364-7, 2026a.

Wei, X., Hayes, D., Ma, L., Daniels, G., Owen, A., Zhao, J., Sandilands, D., Katel, S., Hurtt, G., and Weiskittel, A.: DigiSylva Dataset, Harvard Dataverse [data set], https://doi.org/10.7910/DVN/A3EUOQ, 2026b.

https://doi.org/10.5194/essd-2026-773
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Wei, X., Hayes, D., McHale, G., and Weiskittel, A.: Influence of geolocation misalignment on airborne laser scanning-based forest biomass model calibration in structurally complex mixed forests, Forestry: An International Journal of Forest Research, 99, cpag089, https://doi.org/10.1093/forestry/cpag089, 2026c.

Weiskittel, A., Kuehne, C., McTague, J. P., and Oppenheimer, M.: Development and evaluation of an individual tree growth and yield model for the mixed species forest of the Adirondacks Region of New York, USA, Forest Ecosystems, 3, 26, https://doi.org/10.1186/s40663-016-0086-3, 2016.

Westfall, J. A., Coulston, J. W., Gray, A. N., Shaw, J. D., Radtke, P. J., Walker, D. M., Weiskittel, A. R., MacFarlane, D. W., Affleck, D. L. R., Zhao, D., Temesgen, H., Poudel, K. P., Frank, J. M., Prisley, S. P., Wang, Y., Sánchez Meador, A. J., Auty, D., and Domke, G. M.: A national-scale tree volume, biomass, and carbon modeling system for the United States, Gen. Tech. Rep. WO-104, U.S. Department of Agriculture, Forest Service, Washington, DC, USA, https://doi.org/10.2737/WO-GTR-104, 2024.

White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., and Tompalski, P.: Remote sensing technologies for enhancing forest inventories: A review, Canadian Journal of Remote Sensing, 42, 619–641, https://doi.org/10.1080/07038992.2016.1207484, 2016.

White, J. C., Tompalski, P., Vastaranta, M., Wulder, M. A., Saarinen, N., Stepper, C., and Coops, N. C.: A model development and application guide for generating an enhanced forest inventory using airborne laser scanning data and an area-based approach, Information Report FI-X-018, Natural Resources Canada, Canadian Forest Service, Canadian Wood Fibre Centre, Victoria, BC, Canada, 2017.

Workman, A. K.: Schoodic Point: History on the Edge of Acadia National Park, The History Press, Charleston, SC, USA, 2014.

Young, H. E., Ribe, J. H., and Wainwright, K.: Weight tables for tree and shrub species in Maine, Miscellaneous Report 230, Life Sciences and Agriculture Experiment Station, University of Maine at Orono, Orono, ME, USA, 1980.

Zhao, J., Daigneault, A., Wei, X., Salcido, E., and Weiskittel, A.: Forest type, landowner practices, and climate shape tree species diversity in Maine, USA, Forest Ecology and Management, 593, 122919, https://doi.org/10.1016/j.foreco.2025.122919, 2025.

Zhu, Z., Wang, S., and Woodcock, C. E.: Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images, Remote Sensing of Environment, 159, 269–277, https://doi.org/10.1016/j.rse.2014.12.014, 2015.