Science of Remote Sensing 10 (2024) 100150

Contents lists available at ScienceDirect

Science of Remote Sensing

journal homepage: www.sciencedirect.com/journal/science-of-remote-sensing

Towards global spaceborne lidar biomass: Developing and applying boreal forest biomass models for ICESat-2 laser altimetry data

A. Neuenschwander [a,*], L. Duncanson [b], P. Montesano [c], D. Minor [b], E. Guenther [a], S. Hancock [d], M.A. Wulder [e], J.C. White [e], M. Purslow [d], N. Thomas [c], A. Mandel [f], T. Feng [b], J. Armston [b], J.R. Kellner [g], H.E. Andersen [h], L. Boschetti [i], P. Fekety [j], A. Hudak [k], J. Pisek [l], N. Sánchez-López [k], K. Stereńczak [m]

[a] Center for Space Research, University of Texas at Austin, USA
[b] Department of Geography, University of Maryland, USA
[c] Biospheric Sciences, NASA Goddard Space Flight Center, USA
[d] School of Geosciences, University of Edinburgh, UK
[e] Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, Victoria, BC, Canada
[f] Development Seed, Washington DC, USA
[g] Institute at Brown for Environment and Society, Brown University, USA
[h] USDA Forest Service, Pacific Northwest Research Station, USA
[i] Department of Forest Rangeland and Fire Sciences, University of Idaho, USA
[j] Natural Resource Ecology Laboratory, Colorado State University, USA
[k] USDA Forest Service, Rocky Mountain Research Station, USA
[l] Tartu Observatory, University of Tartu, Estonia
[m] Forest Research Institute, Poland

ARTICLE INFO

Keywords:
ICESat-2
Biomass density
Boreal forest
Lidar

ABSTRACT

Space-based laser altimetry has revolutionized our capacity to characterize terrestrial ecosystems through the direct observation of vegetation structure and the terrain beneath it. Data from NASA's ICESat-2 mission provide the first comprehensive look at canopy structure for boreal forests from space-based lidar. The objective of this research was to create ICESat-2 aboveground biomass density (AGBD) models for the global entirety of boreal forests at a 30 m spatial resolution and apply those models to ICESat-2 data from the 2019–2021 period. Although limited in dense canopy, ICESat-2 is the only space-based laser altimeter capable of mapping vegetation in northern latitudes. Along each ICESat-2 orbit track, ground and vegetation height is captured with additional modeling required to characterize biomass. By implementing a similar methodology of estimating AGBD as GEDI, ICESat-2 AGBD estimates can complement GEDI's estimates for a full global accounting of aboveground carbon. Using a suite of field measurements with contemporaneous airborne lidar data over boreal forests, ICESat-2 photons were simulated over many field sites and the impact of two methods of computing relative height (RH) metrics on AGBD at a 30 m along-track spatial resolution were tested; with and without ground photons. AGBD models were developed specifically for ICESat-2 segments having land cover as either Evergreen Needleleaf or Deciduous Broadleaf Trees, whereas a generalized boreal-wide AGBD model was developed for ICESat-2 segments whose land cover was neither. Applying our AGBD models to a set of over 19 million ICESat-2 observations yielded a 30 m along-track AGBD product for the pan-boreal. The ability demonstrated herein to calculate ICESat-2 biomass estimates at a 30 m spatial resolution provides the scientific underpinning for a full, spatially explicit, global accounting of aboveground biomass.

* Corresponding author.
E-mail address: amy@csr.utexas.edu (A. Neuenschwander).

https://doi.org/10.1016/j.srs.2024.100150
Received 25 March 2024; Received in revised form 2 July 2024; Accepted 14 July 2024
Available online 16 July 2024
2666-0172/Crown Copyright © 2024 Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

1. Introduction

Boreal forests comprise a critical component of the global carbon cycle and quantifying their vertical structure is therefore an important information need for science, policy, and management (Gauthier et al., 2015). The boreal biome is characterized by large, closed canopy trees in the southern extent and sparse, short trees in the northern latitudes (Larsen, 1980). Recently this ecosystem has experienced large warming trends and the vegetation is generally subject to disturbances such as fires and insects (Girona et al., 2023). Vegetation structure is regarded as the amount of above ground biomass (AGB) and how that biomass is arranged vertically and horizontally (Oliver and Larson, B.A., 1996). The quantification of forest vegetation structure over large spatial extents requires mapping in three dimensions, with laser altimetry as the optimal tool for measuring forest structure (Lefsky et al., 2002). Space-based laser altimetry provides an opportunity to estimate AGB across all landscapes. Although lidar is the optimal technology for measuring the three-dimensional structure of global forests, these data lack the required spatial coverage to generate wall-to-wall, management and science relevant (sub-hectare scale) biomass map products. The systematic sampling of vegetation structure from ICESat-2 data provides the necessary measurements to establish a baseline of biomass in the northern latitudes (Markus et al., 2017; Neuenschwander and Pitts, 2019) as GEDI's northern latitude extent is 51.5°.

To date, high spatial resolution biomass maps have been largely unavailable across the boreal system due to a lack of active remote sensing data sensitive to forest structure at the spatial scales of boreal forest processes (~900 m^2 to 1 ha). Further, the remoteness of many of these environs complicates collection of airborne lidar data or the implementation of resource management focused forest inventories. To address known limitations, earlier approaches have generated regional biomass maps using various combinations of optical data (e.g. MODIS, Landsat), sparse spaceborne lidar (ICESat/GLAS) (Nelson, 2010), and models informed using airborne lidar transects (Margolis et al., 2015; Matasci et al., 2018; Neigh et al., 2015).

ICESat/GLAS (Geoscience Laser Altimeter System) was the first space-based laser altimeter mission dedicated to mapping surface heights globally and it operated from 2003 to 2009 (Schutz et al., 2005). After the first laser on ICESat failed within the first month, the two remaining lasers were operated in a campaign style which consisted of 33-day observation periods three times per year. However, operating the GLAS lasers in campaign mode significantly reduced the amount of ICESat data collected (Neuenschwander et al., 2008). ICESat's single beam design and orbit resulted in large groundtrack spacing, with almost 49 km between tracks at the equator. For this reason, ICESat was off-pointed to a series of "Targets of Opportunity'' in the mid-latitudes (i.e. non-cryospheric regions) to increase the spatial sampling density. GLAS was a full waveform lidar system operating at 1064 nm with a nominal footprint diameter of 70 m (Neuenschwander et al., 2008). The large diameter of the ICESat/GLAS footprint limited the usability of data in areas with topographic slope more than 10–15° due to the difficulty discriminating the ground and canopy signal in the returned waveform (Duncanson et al., 2010; Hilbert and Schmullius, 2012; Lee et al., 2011). In addition, Nelson (2010) found that ICESat heights are unreliable in boreal forests when canopy heights were less than 7 m indicating the difficulty of using large footprint waveforms (i.e. 70 m diameter) to characterize short stature vegetation. These factors with ICESat data led to the development of multistage and statistically informed approaches (Margolis et al., 2015). More recently, the use of optical imagery from earth observation satellites such as Landsat or Sentinel-2 has demonstrated capability for producing spatially extensive and temporally dense mapping of a variety of forest structural attributes (including height, canopy cover, biomass, and volume; Matasci et al., 2018).

Learning from the ICESat mission, NASA's ICESat-2 (Ice, Cloud, and Land Elevation Satellite-2) mission, launched in 2018 and was designed to increase the spatial coverage over all surfaces (Markus et al., 2017; Neumann et al., 2019). Remotely sensed estimates of vertical structure from an active sensor, such as NASA's ICESat-2 laser altimeter, offer novel and informative spatial predictors with both the required spatial coverage and fine spatial grain necessary to inform modeling and mapping efforts over large (boreal-wide) areas. Instead of one beam, ICESat-2 was designed with six beams with the laser energy split between a high and low energy level to accommodate highly variable surface reflectance. Furthermore, the ICESat-2 satellite is systematically off-pointed in the mid-latitudes every 91-day orbit cycle to increase the spatial sampling such that the spacing between ground-tracks at the equator is less than 2 km after two years on orbit. The 92° orbit of the ICESat-2 satellite allows for surface height measurements to be collected globally (88° S to 88° N). In this research, our goal is to generate aboveground biomass models using ICESat-2 derived heights in the same fashion as those derived from the NASA Global Ecosystem Dynamics Investigation (GEDI) mission (Dubayah et al., 2020). The GEDI mission utilizes simulated lidar metrics from airborne lidar along with field measurements to derive biomass models (Dubayah et al., 2022; Kellner et al., 2023).Accurate estimation of aboveground forest biomass stocks is required to assess the impacts of land use changes such as deforestation and subsequent regrowth on concentrations of atmospheric CO2. The Global Ecosystem Dynamics Investigation (GEDI) is a lidar mission launched by NASA to the International Space Station in 2018. GEDI was specifically designed to retrieve vegetation structure within a novel, theoretical sampling design that explicitly quantifies biomass and its uncertainty across a variety of spatial scales. In this paper we provide the estimates of pan-tropical and temperate biomass derived from two years of GEDI observations. We present estimates of mean biomass densities at 1 km resolution, as well as estimates aggregated to the national level for every country GEDI observes, and at the sub-national level for the United States. For all estimates we provide the standard error of the mean biomass. These data serve as a baseline for current biomass stocks and their future changes, and the mission's integrated use of formal statistical inference points the way towards the possibility of a new generation of powerful monitoring tools from space (Dubayah et al., 2022). These predictions are calibrated against a large set of global field data representing multiple forest types. By implementing the same approach to biomass modeling as those used for GEDI, the boreal biomass models developed in this research create a path toward a seamless and cohesive estimate of global biomass for the time period of 2019–2021. Herein, the methodology and results of producing the ICESat-2 aboveground biomass models for the circumpolar boreal forests and an ICESat-2 along-track biomass product is presented.

1.1. Vegetation structure captured from the ICESat-2 mission

The primary science objective of ICESat-2 data is to collect critical surface height measurements needed to understand changes to the earth's surface by providing an accurate 3D representation of the earth's surface from 88° S to 88° N latitude (Neumann et al., 2019). Since ICESat-2 is in a nearly polar orbit, it is the only space-based laser altimeter that can collect measurements over northern latitudes, including boreal forests. Over non-polar surfaces (roughly 70° and below), the satellite is intentionally pointed away from the reference ground track every 91 days to decrease the distance between adjacent ground tracks and increase the spatial sampling. The more than four years of ICESat-2 data collected to date have demonstrated unprecedented spatial detail in surface profiles over a variety of surfaces, including land ice shelves, sea ice, the ocean (near-shore bathymetry and ocean surface topography) and vegetated land (forest canopy height and topography) (Magruder et al., 2021).

The instrument onboard ICESat-2 is the Advanced Topographic Laser Altimeter System (ATLAS) which uses a photon counting detection system and a 532 nm (green) laser to actively map the surface elevation (Markus et al., 2017). What sets the ATLAS instrument apart from other space-based laser ranging systems is the photon-counting detection technology, which uses a detection modality sensitive to individual photons. A benefit of utilizing a photon counting system is that the laser can operate at 10 kHz laser repetition rate and, when combined with the altitude and velocity of the satellite, results in a 70 cm separation for each laser shot on the surface in the along-track direction (Neuenschwander and Pitts, 2019). A limitation, however, is that only a handful of photons (typically less than 2) are detected from each outgoing laser shot over vegetated surfaces. ATLAS is equipped with a diffractive optical element that splits the laser into 6 beams organized as 3 beam pairs. Each beam pair is approximately 3.3 km apart and consists of a strong beam and weak beam separated by approximately 90 m in the cross-track direction. The energy and detection ratio between the strong and weak beam is approximately 4-to-1, however over vegetated regions that ratio is observed to be approximately 2.5:1 in non-snow covered conditions (Neuenschwander et al., 2022). The ATLAS system utilizes an Etalon 30 p.m. filter tuned at 532 nm to eliminate the majority of non-signal photons. Because ATLAS is sensitive at the photon level, solar background noise that passes through the filter can present a challenge in the surface finding algorithm in some situations. The number of detected photons associated with each outgoing laser pulse is a function of the laser wavelength, transmitted laser energy, surface reflectance, vegetation structure, solar conditions, and scattering and attenuation in the atmosphere. For measurements over vegetation, the number of signal photons detected per shot is distributed between canopy top, mid-canopy and terrain and has been observed to range between 0 and 2 photons in non-snow covered conditions (Neuenschwander et al., 2022). The return photon distribution is a function of the reflectance of both the ground and canopy (Neuenschwander and Magruder, 2016) while also dependent on the vegetation structure (Queinnec, 2022).

The ATL08 algorithm is designed to extract terrain and canopy heights from the ATL03 photon clouds. It produces the ATL08 geophysical data product at a 100 m step size in the along-track direction. Based on the distribution of signal photons, the algorithm estimates the ground surface and top of the canopy surface, and then subsequently labels the individual photons as either noise, ground, canopy, or top of canopy photons. Statistics calculated based on the labeled photons include terrain heights, canopy heights, canopy height metrics (e.g., RH50), as well as radiometric estimates for both terrain and canopy. The 100 m step size was chosen prior to launch to ensure that a sufficient number of labeled photons existed within each segment to confidently estimate both a terrain height and canopy height. As such, if there are not an adequate number of classified photons within a 100 segment, the terrain and canopy heights are not reported in that particular segment. The ATL08 canopy height is calculated as RH98 instead of using RH100 with the motivation to not include a stray errant noise photon in the height calculations. The combination of using RH98 along with a vertical sampling uncertainty present in photon counting data (Neuenschwander and Magruder, 2016), the ATL08 canopy heights are all but guaranteed to be biased low (Neuenschwander et al., 2020).

Although critical for building many biomass allometries, canopy cover is not a derived parameter on the current ATL08 data product due to the influence of surface reflectance on the radiometry (Neuenschwander et al., 2022). In Armston et al. (2013), the derivation for canopy gap (or 1 - canopy cover) was shown as falling along a line describing the integrated backscatter of ground against the integrated backscatter of canopy. Fig. 1 depicts the partitioned ground and canopy radiometry observed by ICESat-2 over boreal forest in Central Alaska which displays a similar relationship as to waveform lidar. In theory, canopy cover, whether derived from GEDI or ICESat-2 can be calculated as

$$cover = \frac{1}{1 + \frac{\rho_v}{\rho_g}\frac{R_g}{R_v}} \quad \text{Equation 1}$$

where the canopy and ground reflectance are, ρ_v and ρ_g, and R_v and R_g represent the integrated backscatter on GEDI. ICESat-2 is a photon

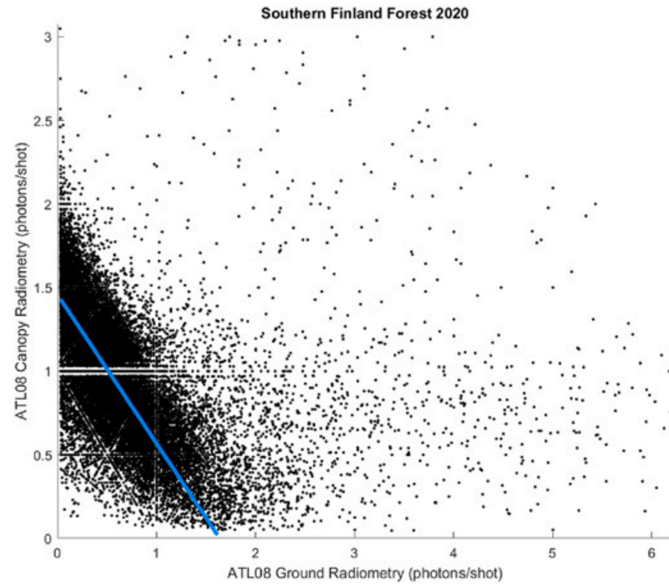

Fig. 1. ICESat-2 ground and vegetation radiometric profiles. The linear relation of the scatter is indicative of canopy cover where the x-intercept implies 0% cover and the y-intercept corresponds to 100% canopy cover.

sampling system, and like any lidar system, the number of detected photons (R_g and R_v) is not a direct measure of the ground and canopy reflectance (ρ_g and ρ_v). Instead, the number of detected photons is a function of the atmospheric transmission and surface reflectance, and the number of photons will vary with each orbit. Although photon counting sampling limits the signal to noise, canopy cover can be directly estimated in the same manner as other lidar systems. Fig. 1 depicts the radiometric profile of ICESat-2 and the linear relationship between R_g and R_v. The relationship between reflectances and canopy cover is an important element within the ICESat-2 simulator (Purslow et al., 2023).

2. Methodology and data

The objective of this research is to create ICESat-2 aboveground biomass density (AGBD) models for boreal forests globally at a 30 m spatial resolution and apply those models to ICESat-2 data from the 2019–2021 period (Fig. 2). Our approach is to relate field estimates of AGB to simulated ICESat-2 heights derived from airborne lidar and an ICESat-2 simulator (Purslow et al., 2023). The strategy for this research was to adopt the "crowd sourced" approach for collecting and applying field data as was implemented by the GEDI mission to develop the GEDI biomass models. Specifically, field data that contains a temporally coincident collection of airborne lidar data with the collection of the field measurements ensures that the field data never become stagnant. Field estimates of AGBD were then related to simulated ICESat-2 data derived from the coincident airborne lidar. The ICESat-2 biomass models were then applied to on-orbit ICESat-2 data remapped to 30 m canopy heights and canopy percentile relative heights. These ICESat-2 AGB predictions provided a global representation of biomass across all PFTs which aided in reducing regional uncertainties due to a lack of training data identified in previous studies (Duncanson et al., 2022; Kellner et al., 2023). All data processing was conducted on the ESA-NASA Multi-mission Algorithm and Analysis Platform (MAAP) (Albinet et al., 2020). This open science system allowed efficient parallel cloud-based computation. All associated code can be accessed in a public git repository https://github.com/lauraduncanson/icesat2_boreal.git.

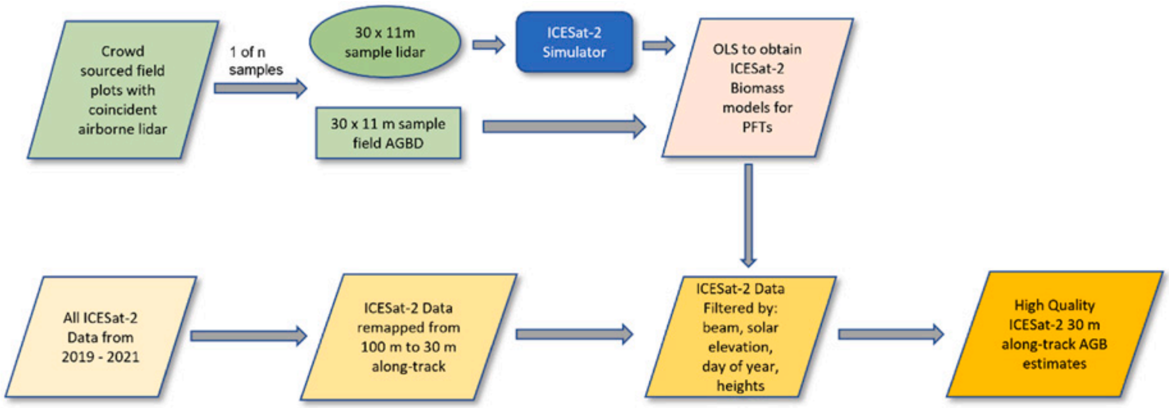

Fig. 2. Flow diagram of methodology used in this research.

2.1. Assembling field data

Coincident field and airborne lidar data were used to create the ICESat-2 based biomass models, following the GEDI approach as outlined by Duncanson et al. (2022) and Kellner et al. (2023). The field data consist of an international set of over 30 projects (listed in Supplemental Table 1) that were publicly or privately provided. Since the field data projects were not collected specifically for the purpose of ICESat-2 AGB modeling, these sites feature a variety of plot sizes and shapes, numbers of plots, and airborne lidar instruments. All field data were required to be north of 50° latitude, geolocated, and have coincident airborne lidar data. Furthermore, as many field sites consisted of several plots, an additional requirement was that individual field plots be no smaller than 10 × 10 m and no larger than 30 × 30 m in size. Additionally, each field plot needed AGBD aggregated to the plot level or contain individual tree diameter measurements such that AGBD could be directly estimated. The distribution and the number of equivalent 30 m along-track training samples based on the field data are shown in Fig. 3. The training data sites used in this project are distributed across plant functional types (derived from MODIS Land Cover (MCD12Q1) Type 1) throughout the boreal forest domain including: Deciduous Broadleaf Trees, DBT; Evergreen Broadleaf Trees, EBT; Evergreen Needleleaf Trees, ENT; Mixed Forest, MF; and Grassland Shrubland Woodland, GSW. Because the northern extents of boreal forests are open and have shorter canopy heights, we included grassland/savanna/shrubland/woodland as a PFT in our initial modeling. For this project, Deciduous Needleleaf Trees were not represented in our field data as those forests typically occur in Eastern Russia where field data is difficult to obtain due to political and logistic reasons. Although we originally explored fitting models to all represented PFTs (e.g., grass woodland/shrubland or mixed forest), we

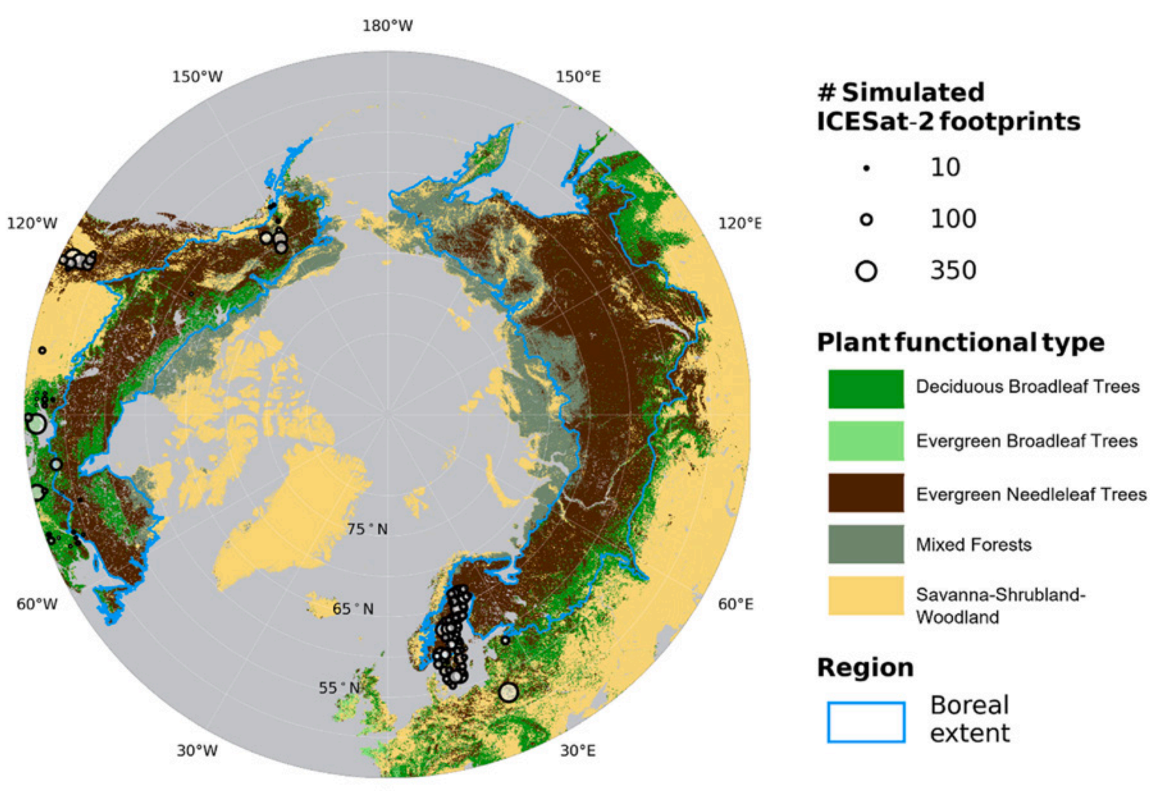

Fig. 3. Map of simulated ICESat-2 footprints used in analysis, showing the distribution for different PFTs across the domain: Deciduous Broadleaf Trees (DBT), Evergreen Broadleaf Trees (EBT), Evergreen Needleleaf Trees (ENT), Mixed Forest (MF), and Grassland/Savanna Shrubland Woodland (GSW).

selected ENT and DBT as the two primary models as the majority of the field data were located within those PFTs. For land cover that did not fall within either the ENT or DBT class, we created a generalized pan-boreal model comprised of all landcover types. As more training datasets become available in the boreal, further stratification may be possible for future versions of the product. Generalized allometric models for each site (see Supplemental Table 1) were used to estimate individual tree biomass. Plot level AGBD (~30 × 11 m^2 plot size) was calculated by summing all individual tree AGBD above a minimum diameter at breast height threshold of 10 cm.

2.2. Simulating ICESat-2 from airborne lidar

The ICESat-2 biomass model creation followed the method developed for the GEDI biomass product as described in Duncanson et al. (2022) and Kellner et al. (2023). The method for this effort used a lidar simulator (Hancock et al., 2019; Neuenschwander and Magruder, 2016; Purslow et al., 2023) to generate ICESat-2-like data from airborne laser scanning (ALS) data corresponding to the area covered by the field plots. The generation of simulated ICESat-2 data overcomes limitations and challenges of matching field plots to satellite lidar, both in terms of spatial resolution (ICESat-2 segments versus plots) and spatial co-location. Furthermore, the simulated ICESat-2 data ensures temporal matching to the field data. The ICESat-2 simulator follows the method described by Neuenschwander and Magruder (2016), using ALS point-cloud data to produce 20.5 m diameter pseudo-waveforms over ground plots, by scaling the canopy and ground signals to match the observed ratio of canopy and ground reflectance. A simulated ICESat-2 footprint, centered on each field plot, was extracted from the airborne lidar point cloud. A 20.5 m diameter footprint was used in the simulator to match the area of a 30 × 11 m ICESat-2 segment (i.e. ~330 m^2). Next, each pseudo-waveform was randomly sampled vertically N times based on the expectation value for mean photon count over boreal forest to produce simulated ICESat-2 photons representative of an ICESat-2 laser shot. The process was replicated 70 cm in location from the previous location within the airborne lidar data, matching the along-track sample rate of ICESat-2. The motivation for analyzing the ICESat-2 at 30 m instead of the nominal ATL08 resolution of 100 m is threefold. First, the canopy height increments at 30 m along-track resolution are closer in scale to GEDI footprint level canopy heights. Second, the 30 m resolution closely aligns with the pixel resolution of Landsat (Crawford et al., 2023) or Harmonized Landsat Sentinel (HLS) (Claverie et al., 2018) data products. Finally, this resolution approximately matches the spatial resolution of most of the available training datasets in the boreal zone.

Once the simulated ICESat-2 photons were extracted, metrics such as canopy height and the relative height (RH) metrics were calculated, generating data analogous to that of ICESat-2 for each plot. The simulator has been validated in full-waveform mode (Hancock et al., 2019) and photon-counting mode (Purslow et al., 2023) and found to have a negligible bias (<50 cm) and small RMSE (<3 m) for RH metrics, as long as the ratio of canopy and ground reflectance is correctly derived (Purslow et al., 2023). A bias in the reflectance ratio causes a bias in simulated RH metrics. This ratio was determined following the method of (Armston et al., 2013), with adaptations to account for the nature of a photon-counting lidar described in Purslow et al. (2023). Note that these results showed the accuracy of the simulator at generating synthetic ICESat-2 metrics, rather than the accuracy of ICESat-2 metrics themselves.

For the simulations used in this study, a reflectance ratio (vegetation/ground) of 1.21 was used, derived from the Wind River site in Purslow et al. (2023). It should be noted that this ratio is likely to vary between sites, but without a reflectance ratio map derived from ICESat-2 it is not possible to vary the ratio between sites. As such, this reflectance ratio may cause some spatially varying bias. To quantify the potential impact of varying reflectance ratio on the modeled AGB, AGB was estimated for the full range of observed reflectance ratios (1.2–0.75) and the standard deviation of the AGB estimates across these reflectance ratios was calculated. This analysis showed a standard deviation in AGB of 5.75% across all footprints tested, rising to a median standard deviation of 8% at 40% canopy cover due to the instability of RH60 at this cover level (Hancock et al., 2019). This pre-analysis test confirmed that the impact of the ground-canopy reflectance ratio on the overall biomass estimates is small.

2.3. Linking field AGB with simulated ICESat-2

The simulated 30 m ICESat-2 RH metrics were used as predictors in a suite of PFT-based Ordinary Least Squares (OLS) models, where plot-based estimates of AGB serve as reference data. To create an ICESat-2 derived AGBD product that can potentially be used with GEDI biomass estimates, the same model form was adopted. Following the approach used for GEDI data (Duncanson et al., 2022; Kellner et al., 2023) the geographic domain of the boreal was spatially divided by PFT (Evergreen Needleleaf (ENT), Deciduous Broadleaf (DBT)) or a general pan-boreal for vegetation that does not fit the ENT or DBT forest type (e.g. mixed forests). An exhaustive set of models were fit and the best performing models based on geographic transferability were selected (Duncanson et al., 2022; Kellner et al., 2023). While all 10th percentile RH metrics were originally considered (e.g. RH10, RH20, RH30 … RH98), the lowest RH metric adopted for the final GEDI models was RH50 (Duncanson et al., 2022; Kellner et al., 2023), as the lower RH metrics were less reliably simulated for GEDI (Hancock et al., 2019), and therefore may have introduced systematic errors into the AGB products. This systematic error does not appear to hold for the lower RH metrics based on ICESat-2 simulations (Purslow et al., 2023), and future ICESat-2 AGB models will consider lower RH metrics. A lookup table derived from land cover provided on the ATL08 product was used to choose which strata each on-orbit ICESat-2 shot fell into, and therefore the associated OLS model to be applied. Models were assessed both by cross validation of model fits, and by geographic cross validation (see Duncanson et al., 2022 for details). Geographic cross validation aims to assess the transferability of models to data outside the geographic domain of the training data by removing geographic cells within a fit strata from the model fit, applying the model to the out of bag data, and iterating through all 1 × 1° geographic cells in the domain. Geographic cross validation is important to ensure the model is not falsely inflating the predictive power due to spatial autocorrelation (Ploton et al., 2020). This process is also known as a leave-one-out approach. The model, trained without data from a particular geographic cell, was used to predict values for that geographic cell. The final model for each PFT was selected based on minimizing systematic error (bias) and %RMSE (Duncanson et al., 2022; Kellner et al., 2023). This approach ensures the model is not biased toward a specific geographic area. The chosen models were used to create the ICESat-2 30 m AGB data product.

2.4. Compiling ICESat-2 estimates at 30 m

To better match the footprint size of GEDI and to more closely match the average plot size in the available field dataset, the ICESat-2 ATL08 photon labels were mapped to the ATL03 geolocated photons to produce 30 m along-track segments. The photon indices provided on the ATL08 data product allowed the terrain and canopy height metrics at a 30 m resolution in the along-track direction to be calculated in the same manner that they are calculated for the 100 m transect on the official ATL08 data product. The standard ATL08 data product imposes a series of quality checks to ensure that the estimated terrain and canopy heights are valid. When creating the 30 m ATL08 data, the calculation of the 30 m product was restricted to only photons that are part of a valid ATL08 record, thus keeping the quality check measures from the standard 100 m ATL08 data product in place. An example of the 30 m and 100 m maximum canopy heights for a 3 km track of ATL03 photon level data is shown in Fig. 4. The photons are color coded by the label assigned from

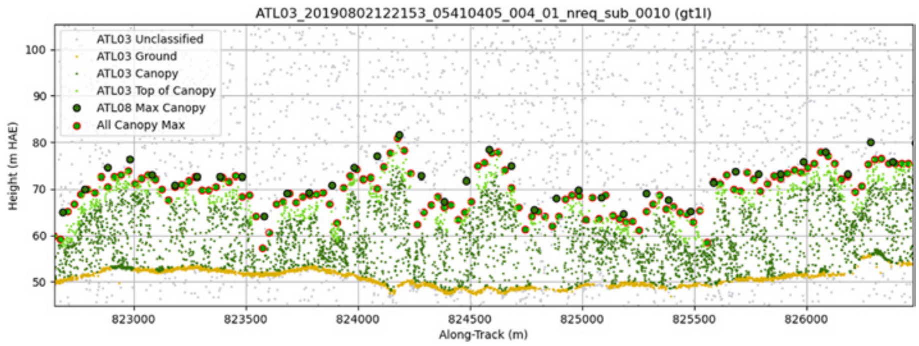

Fig. 4. Along-track profile of ICESat-2 ATL03 data emphasizing the close correspondence between 30 m (light green large dots) and 100 m (dark green large dots) canopy height estimates. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the ATL08 algorithm.

In addition to calculating the terrain and canopy height, the relative canopy height metrics were also calculated at 30 m and include: RH10, RH20, RH25, RH30, RH40, RH50, RH60, RH70, RH75, RH80, RH85, RH90, RH95, RH98 and RH99. On the official ATL08 data product (Neuenschwander and Pitts, 2019) the RH metrics are defined solely upon the photons labeled as canopy and top of canopy. For the purposes of this analysis, redefining the RH metrics as the relative height of all signal photons, thus including the labeled ground photons into the calculations was tested. This shift in the RH calculation would allow for a more direct comparison of metrics and AGBD models calculated from waveform lidar systems such as GEDI, which is based on the total waveform energy. By incorporating the ground photons into the calculation of the RH metrics, an element of vegetation cover is inherent in the relative heights. Since canopy cover is not directly calculated on the ATL08 data product (Neuenschwander et al., 2022), it was hypothesized that the RH metrics with ground photons would potentially be more sensitive to AGBD variations related to canopy cover. For example, a sparse forest with a RH50 value of 20 m may have a very different AGBD than a dense forest with the same canopy height. Conversely, however, including ground photons in the calculation would result in RH metrics that are sensitive to surface reflectance and thus the brightness of the underlying ground would potentially impact AGBD estimates.

2.5. Filtering ICESat-2 data

The aboveground biomass models created for this project are derived from simulated ICESat-2 data, which represent the optimal acquisition conditions. For this reason, a series of filters and thresholds when creating the actual ICESat-2 along-track biomass estimates to eliminate ICESat-2 data collected under non-optimal conditions was utilized. To create the along-track ICESat-2 biomass estimates, the ATL03 (geolocated photon cloud) and ATL08 (Land and Vegetation) release 005 (rel005) data for latitudes above 50° north from the years 2019 through 2021 was used. By creating a 30 m estimate of terrain and canopy height, several flags and parameters from the ATL08 data product were used to remove potentially low-quality returns. These flags and filters included the multiple scattering flag (msw_flag), a snow presence flag, a solar elevation value (solar_elev), and the segment land cover. The multiple scattering flag is derived from ATL09 atmospheric data product and flags potential atmospheric attenuation. The snow flag is extracted from the NOAA daily snow cover product (Estilow et al., 2015) and indicates whether the ground within an ATL08 100 m segment is covered with snow or ice or is considered snow-free. The solar elevation parameter can be used to identify night from day acquisitions. The ATLAS strong beams under night acquisitions yielded better canopy height estimates compared to day acquisitions for two studies in Scandinavia (Neuenschwander et al., 2020; Varvia et al., 2022). To calculate above ground biomass estimates, ATL08 data with the following criteria (Table 1) were selected: ICESat-2 strong beams only, acquired during the growing season (June through September), with solar elevations less than 5°, over snow free land (snow flag set to 1), and with an msw_flag equal to 0 indicating clear skies and no observed atmospheric scattering. Including ICESat-2 data having a solar elevation up to 5° allowed us to increase the number of samples for the most northern latitudes during our temporal window (June through September), while still selecting data collected with a low amount of solar background noise.

Table 1
Filter fields, and parameter and threshold values used to derive a model-ready set of 30 m ICESat-2 observations.

ATL08 Field description	Value (parameter or threshold value used to include observations)
Beam	Strong (Beams 1, 3, 5)
Multiple scattering flag (msw_flg)	0 (no atmospheric scattering detected)
Solar elevation angle (solar_elev)	$\leq 5^{\circ}$
Snow flag (segment_snow)	≤ 1 (no ice, no snow)
Seasonality	June through September
Copernicus land cover	See Table 3

With these basic filters applied to the data, the data were further filtered the ICESat-2 data based upon a maximum allowed canopy height for each land cover classification as listed in Table 2. The threshold heights were selected to represent a maximum expected height for a given land cover type within the boreal domain. The ATL08 data product incorporates the Copernicus 2019 land cover values (Buchhorn et al., 2020) on Release 005 for each 100 m segment. Applying a maximum canopy height filter eliminated those ATL08 segments where canopy heights were mislabeled due to low-lying clouds, fog, or background noise. Although these filters are helpful to eliminate errant estimated heights, the potential for incorrect canopy heights from ICESat-2 remains.

The AGB models developed using the field plot and simulated ICESat-2 data with the two different sets of RH metrics were then applied to the filtered ICESat-2 data to produce an along-track, 30 m ICESat-2 data product.

3. Results

3.1. Canopy height metrics

Two sets of biomass models were fit to the 30 m ATL08 RH metrics as predictors using the ICESat-2 based RH metrics and the RH metrics with ground photons included in the calculation. A comparison between the two sets of RH metrics is shown in Fig. 5. With the ground photons not included in the RH calculation, the ICESat-2 RH metrics represent a higher canopy height for all RH metrics except for RH100, where the

Table 2
Height thresholds and AGB models applied to each ICESat-2 ATL08 land cover class.

Canopy Height Threshold (m)	Land cover Type	AGB model applied
60	Closed forest, evergreen needleleaf	ENT (boreal-wide)
60	Closed forest, deciduous needleleaf	DBT (boreal-wide)
60	Closed forest, evergreen broadleaf	DBT (boreal-wide)
60	Closed forest, deciduous broadleaf	DBT (boreal-wide)
60	Closed forest, mixed	Boreal-wide model
60	Closed forest, unknown	Boreal-wide model
50	Open forest, evergreen needleleaf	ENT (boreal-wide)
50	Open forest, deciduous needleleaf	DBT (boreal-wide)
50	Open forest, evergreen broadleaf	DBT (boreal-wide)
50	Open forest, deciduous broadleaf	DBT (boreal-wide)
50	Open forest, mixed	Boreal-wide model
50	Open forest, unknown	Boreal-wide model
20	Shrubs	Boreal-wide model
10	Herbaceous	Boreal-wide model
10	Herbaceous Wetland	Boreal-wide model
5	Moss and lichen	Boreal-wide model
5	Bare/sparse vegetation	Boreal-wide model

Table 3
Results from ICESat-2 biomass OLS predictions using simulated ICESat-2 (**no** ground photons included in RH metric calculation). Statistics are calculated by geographic cross validation, which differs from the statistics on Fig. 7 which were calculated by normal cross validation.

ICESat-2 land cover strata AGBD model	R^2	% RMSE	ME (Mg/ha)	Transform	Predictors
DBT	0.71	41.17	10.04	sqrt-sqrt	RH50, RH98
ENT	0.70	64.90	6.31	sqrt-sqrt	RH70, RH98
Generalized pan-boreal	0.70	62.42	7.75	sqrt-sqrt	RH60, RH98

value is the same for both sets of RH metrics.

Another way to visualize the impact of including the ground photons in the RH calculations is captured in Fig. 6. Here, the RH metrics are plotted as a function of the canopy cover ratio based upon the ICESat-2 photons where

$$cover = \frac{nphotons_{canopy}}{nphotons_{ground} + nphotons_{canopy}} \qquad \text{Equation 2}$$

Viewing the data in this manner, the influence of the ground photons is significant in the lower RH metrics until a certain threshold of canopy cover is attained. For example, 30% of the ICESat-2 photons must be canopy photons before the RH70 heights represent a height greater than zero. For this reason, only RH metrics at RH50 or above for our biomass models were tested as the lower RH values would be heavily dominated by ground photons in much of the relatively sparse boreal ecosystem. The ICESat-2 based RH metrics (i.e., with no ground photons) are more sensitive to segments that have low canopy cover, or rather, a low number of photons reflected from the canopy.

3.2. Modeled AGBD from simulated ICESat-2

The OLS model fits between field-estimated AGB and simulated 30 m ICESat-2 RH metrics with and without ground photons are shown in Fig. 7. The OLS models are stratified by PFT. A boreal-wide model was also fit to the entire domain. In the case of the simulated ICESat-2 RH metrics when the ground photons were included (Table 4), the model accuracy statistics for all strata (PFTs and pan-boreal) were comparable to the models without ground photons (Table 3), with slightly higher % RMSE values. The statistics reported in Tables 3 and 4 are the results of the cross validation. ME is the mean error, taken as a bias term, which is the absolute average of binned errors in AGB strata. Ideally ME would be zero for a totally unbiased model. As with many AGBD models, overestimations of low AGBD and underestimations of high AGBD were found, particularly for the DBT strata. All models were forced to use maximum height as a predictor, but flexibly selected lower RH metrics for prediction. The models built with ground photons selected RH60, RH70 and RH98 as predictors, with the models built using ground-free RH metrics selecting lower predictors which are likely comparable to their slightly higher RH metric counterparts (i.e., RH50 with no ground photons may be comparable to RH60 with ground photons included, as ground photons decrease the height values). All models were square-root square-root models, which followed the selected model form for GEDI's tropical and temperate AGBD models (Duncanson et al., 2022; Kellner et al., 2023).

The inclusion of ground photons has a significant influence on the AGBD predictions. When comparing the predicted AGBD via normal cross validation (Fig. 7), it is apparent that models using RH metrics with ground photons for both ENT forests and pan-boreal models occasionally predict zero AGBD when reference AGBD is not zero (including when reference AGBD >200 Mg/ha). Conversely, ground-free models (ICESat-2 style RH metrics) occasionally predict high AGBD when the reference AGBD is zero or near-zero. Within the models with ground photons included, this phenomenon can be explained as the influence of the ground photons on the RH metrics. Areas with a low canopy cover (i.e., cover <40%), or high surface reflectance (e.g. bright bare earth) would likely have an RH60 height of zero within the ground photon-inclusion models. Thus, these models will have difficulty predicting AGBD based on only photons from vegetation when vegetation is very sparse. Therefore, while the RH98 would be similar for each style of metrics, the lower RH metrics (RH60 and RH70) would likely report a ground elevation in lower canopy cover areas. This would result in under estimates of AGBD in sparse vegetation when the RH metrics include ground photons. Conversely, the ICESat-2 style model developed without the ground photons included in the RH metrics occasionally overpredicts AGBD when there is no to little biomass in the reference data. For example, a single tall tree in a plot may have a relatively tall height, but the remainder of the plot will be populated only by ground photons. Calculating RH metrics without inclusion of these ground photons would yield high RH metric values for the tall tree, while plot-level AGBD remains low. This would indeed result in overestimates of AGBD in areas of sparse cover. Finally, ground surface reflectance would influence RH metrics from changes due to surface reflectance phenology (i.e. snow cover, moisture, or sun angle) which in turn would lead to variable RH metrics without a true change in vegetation structure.

3.3. ICESat-2 canopy heights and biomass for the pan-boreal region

The goal of this work was to apply the aboveground biomass density models to the set of remapped 30 m ATL08 data for the 2019–2021 period. Following the filtering process described earlier, the 30 m ATL08 canopy heights (RH98) were mapped across the entire pan-boreal region shown in Fig. 8. In total, more than 19 million (19,262,962) 30 m

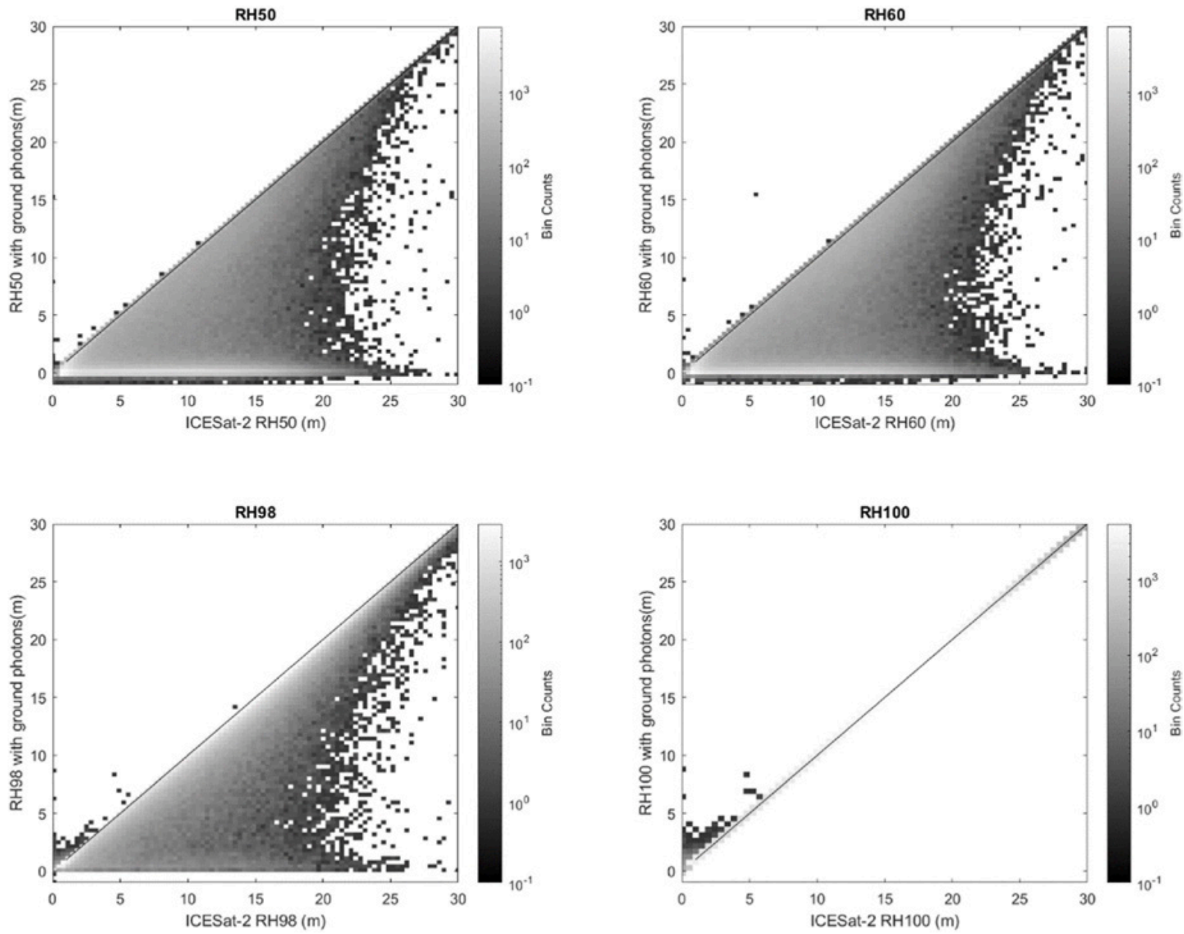

Fig. 5. ICESat-2 relative height metrics for canopy compared to RH metrics with ground photons included. RH metrics are calculated from ICESat-2 photons filtered by strong beam, forest, and night conditions. As the RH percentage increases the more similar the metrics become.

ICESat-2 samples were used to create this product although only 1% of the randomly sampled data points are plotted. Areas shown as light gray fall outside of the boreal domain and dark gray areas represent regions with no or sparse amounts of forested ICESat-2 samples. The data plotted in Fig. 9 depict the distribution of forest canopy heights spatially across the boreal zone from ICESat-2, however this is not a wall-to-wall product. The plotted canopy heights agree with our expected height distributions with the taller forests occurring closer to the southern extent (50–60° north) of the boreal zone with tree heights decreasing with increasing latitude. This product represents the first space-based canopy height product derived solely from laser altimetry for boreal forest.

Fig. 9 represents the distribution of the AGBD models across the boreal domain applied in this project. As observed in this figure, a majority of the ICESat-2 samples fall under the ENT model boreal wide (shown as aqua), whereas the DBT model (shown as purple) was applied primarily to forests at the southern extent of the boreal biome. ICESat-2 samples falling over short open forest, frequently found closer to the northern limit of the boreal biome (shown as yellow), were estimated with the general pan-boreal model.

Fig. 10 represents the 30 m along-track mean aboveground biomass density estimates using the biomass models developed herein. Biomass density was calculated with one of the three models: ENT, DBT, or the generalized pan-boreal based upon the land cover class within each 30 m segment. The highest canopy heights and biomass density occurs in eastern Europe and Western Russia. This along-track product represents the first space-based estimate of aboveground biomass density derived solely from laser altimetry for the boreal forest region, however, this product does not provide a full accounting of boreal biomass as the data only represent the ICESat-2 tracks, which are samples and not wall to wall coverage. As evident in Fig. 10, Eurasia contains more AGBD than North America and observations between 100 and 200 Mg/Ha are more prevalent in Eurasian boreal forests than North American boreal forests.

Fig. 11 represents the histograms of aboveground biomass density stratified by continent for our filtered 30 m ICESat-2 dataset. In our sample set, the majority of samples over North American boreal forests contain less than 50 Mg/ha of aboveground biomass, whereas the majority of biomass estimates over Eurasian forests were less than 100 Mg/ha. The histograms of our 30 m AGBD, however, show that the distribution of values is not Gaussian and thus the reported statistics (i.e. mean, standard deviation, etc.) do not convey a full description of the values.

4. Discussion

The objective of this research was to develop ICESat-2 based biomass density models for boreal forests and apply them to a set of filtered 30 m ICESat-2 canopy heights across the entire pan-boreal region. The outcome of this research is a set of high-quality aboveground biomass density estimates from ICESat-2 that can be further used to estimate the total aboveground biomass density for boreal forests. Although the ICESat-2 data used in this study were highly filtered based on several criteria (i.e., time of day, beam strength, day of year, and presumed cloud-free observations), there still might be errant values associated with the derived canopy heights due to low-lying clouds/fog or blunders in the ATL08 ground finding algorithm. The process of creating our aboveground biomass density estimates stem from the availability of

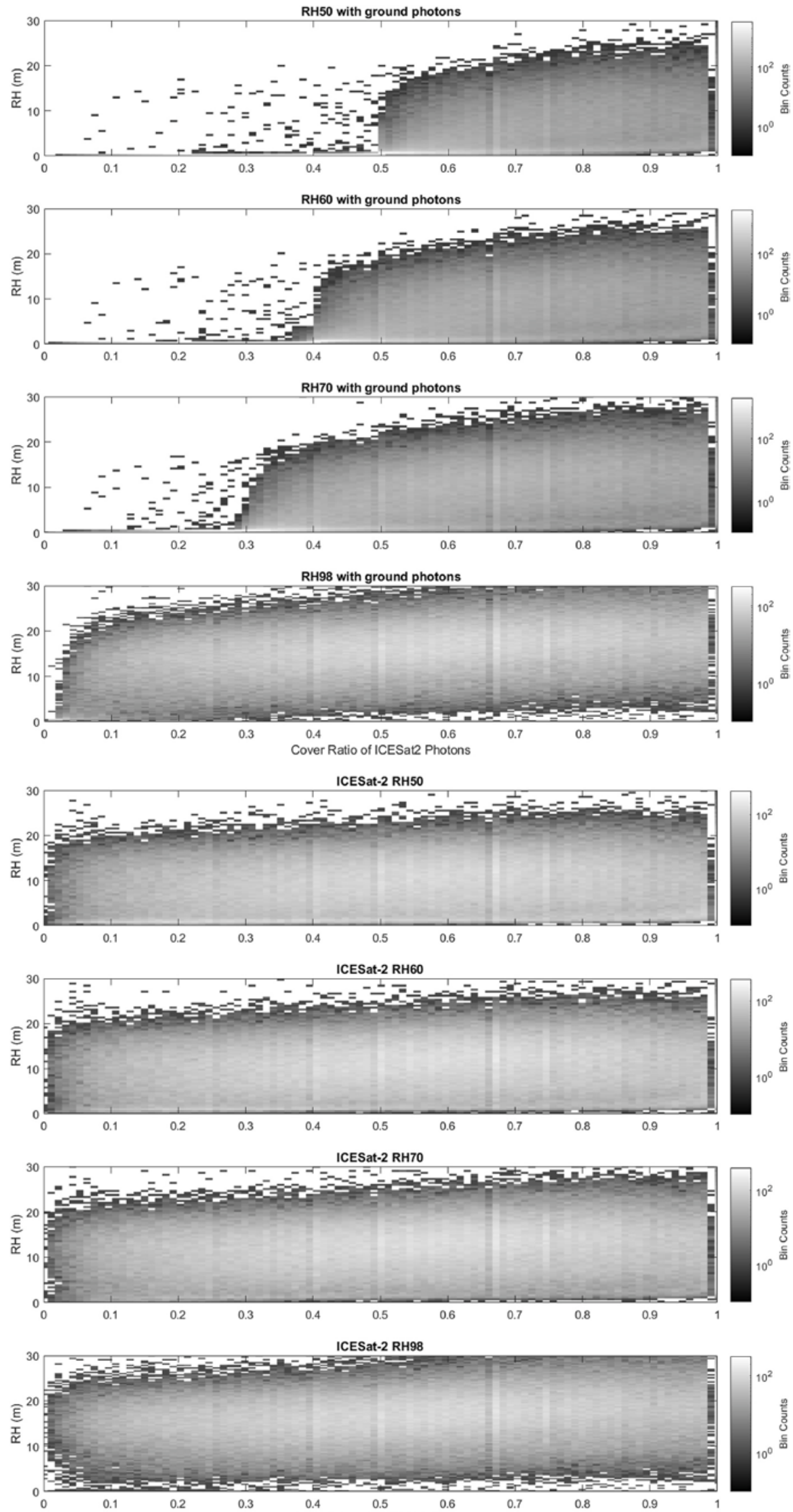

Fig. 6. A comparison of RH canopy height metrics (RH50, RH60, RH70, and RH98) plotted against an ICESat-2 based cover ratio with ground photons (Top Panel) and without ground photons (Bottom Panel).

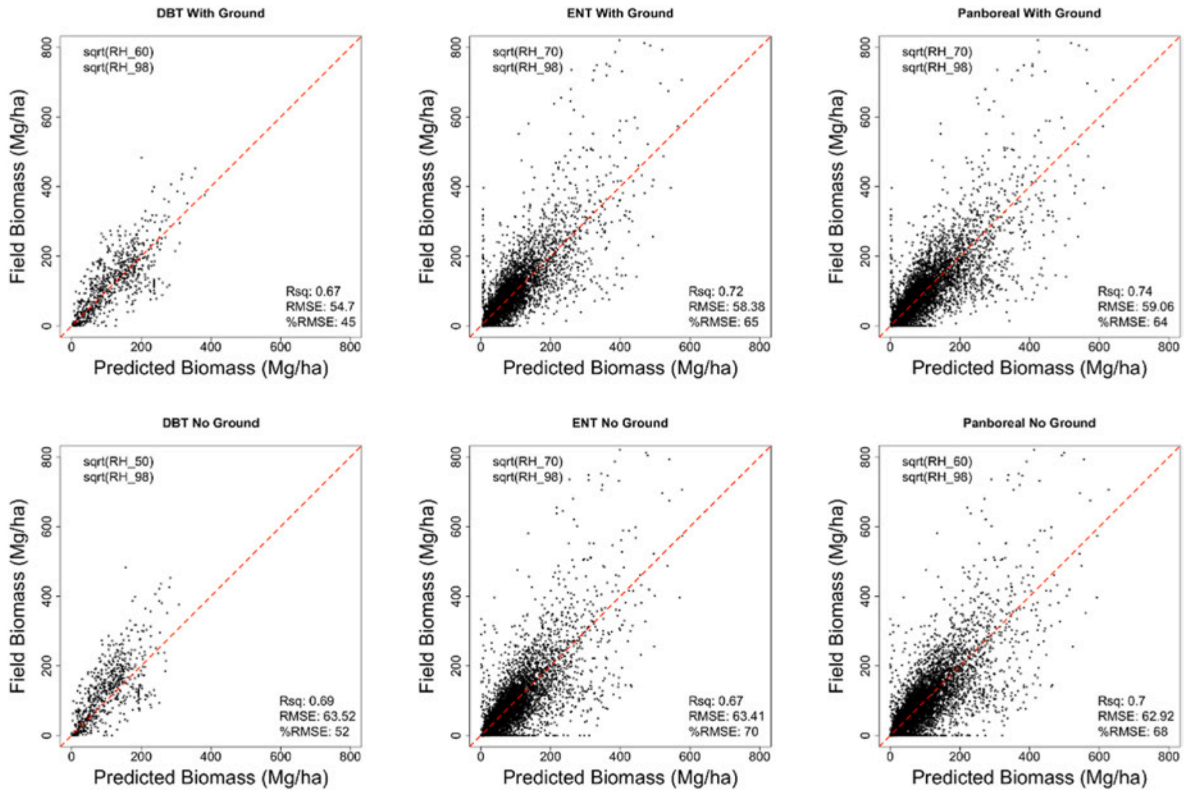

Fig. 7. Scatter plots of observed versus predicted above ground biomass density estimates (in Mg/ha) using canopy height metrics from simulated ICESat-2 data, with (top row) and without (bottom row) ground photons included in metric calculations.

Table 4
Results from ICESat-2 biomass OLS predictions using simulated ICESat-2 (**with** ground photons included in RH metric calculation). Statistics are calculated by geographic cross validation, which differs from the statistics on Fig. 7 which were calculated by normal cross validation.

ICESat-2 land cover strata AGBD model	R^2	% RMSE	ME (Mg/ha)	Transform	Predictors
DBT	0.67	43.73	9.96	sqrt-sqrt	RH60, RH98
ENT	0.70	65.24	6.18	sqrt-sqrt	RH70, RH98
Generalized pan-boreal	0.71	62.75	6.90	sqrt-sqrt	RH70, RH98

field data with coincident airborne lidar across the pan-boreal. Most of the available open field data used in this study were collected over Canada, Alaska, Sweden, Norway, and Finland with notable absence of data over Russia. If additional field data with coincident lidar data were available, particularly over Russian forests, the aboveground biomass models could be developed specific to continent-based plant functional types (i.e. Eurasian DNT, ENT or North American DBT). The abundance of field data over Evergreen Needleleaf (ENT) forests far exceeded that of Deciduous Broadleaf (DBT) forests or other forest types. The majority of DBT field plots were collected over North America, thus Eurasian Deciduous Broadleaf forests are skewed toward North American tree characteristics and may not be as representative of Eurasian DBT forests. Although ENT field plots were available across both continents, the majority of ENT field data were collected from Scandinavian forests. Aboveground biomass estimates for ICESat-2 data that did not have a land cover value of ENT or DBT were calculated using a general boreal model that included all provided field data.

4.1. AGBD estimates built from ICESat-2 and GEDI-like predictor sets

The results presented herein indicate that models developed using ICESat-2 RH metrics calculated without ground photons were comparable to models developed using RH metrics that included ground photons. Differences in the predictors for each ICESat-2 AGBD model varied by land cover strata (See Table 3). The predictors for the model when ground photons were included in Deciduous Broadleaf were RH60 and RH98, whereas the ICESat-2 predictors were RH50 and RH98. Overall, the inclusion of the ground photons in the models resulted in a selection of higher RH metrics for model development. The inclusion of the ground photons also resulted in the more frequent prediction of zero biomass, presumably because ground photons dominate the height distribution in areas with very sparse forests with few trees. In contrast, models developed using RH metrics without the inclusion of ground photons over-predicted AGB in low biomass areas wherein single or sparse tree cover results in high RH metric values. This over-prediction in low biomass segments was observed in the OLS model fits for the simulated ICESat-2 data. The models based on RH metrics without ground photons, however, are also potentially under-estimating AGBD in areas with tall trees and denser cover as this trend was observed in both the OLS models as well as the AGBD transform distribution. The inclusion of an ICESat-2 derived canopy cover estimate would be expected to improve AGBD predictions, however, due to the impact of atmospheric transmission and surface reflectance on the number of detected surface photons, canopy cover cannot be accurately retrieved from the ATL03 data product without a true mapping of the ground and vegetation reflectance (Purslow et al., 2023). Ultimately, the motivation for using the ICESat-2 models (i.e. without the ground photons) was based on two primary reasons. Based solely upon the OLS models, the model including the ground photons actually performed slightly better than the ICESat-2 based models that did not include ground photons (Tables 3 and 4). However, the results of the geographic cross-validation

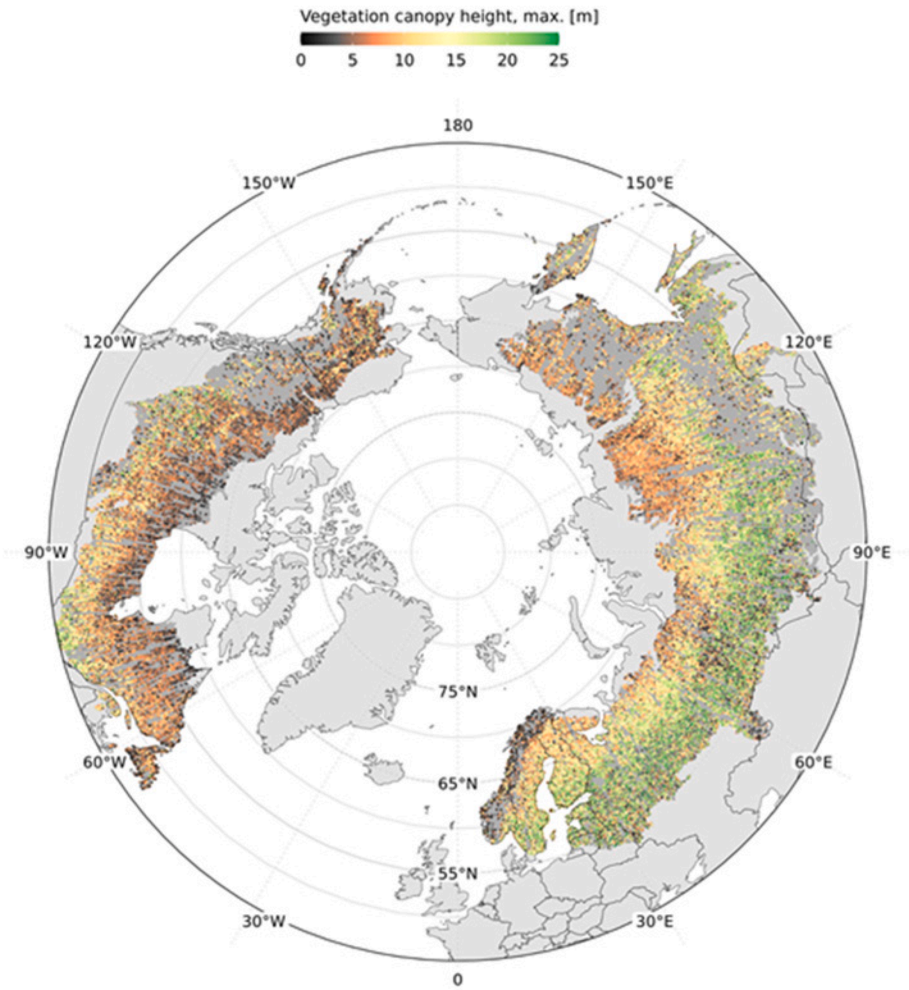

Fig. 8. 30 m along-track canopy heights (RH98) from ICESat-2 across the pan-boreal study area.

indicate that the models without ground photons performed slightly better than the models with ground photons, which may be due to large-scale variations in canopy and ground reflectance causing a spatially varying bias in RH metrics with ground photons if the canopy/ground reflectance ratio is not accounted for (Purslow et al., 2023). The results of the geographic cross validation method received greater weight in our consideration of which models to apply to the 30 m ICESat-2 data, as these results provide an indication of which models were more transferable across the full boreal domain.

4.2. Sampling of AGBD with ICESat-2 reveals details across forest gradients

The result of applying the biomass density models to the 30 m resampled ICESat-2 data is a set of high quality (filtered) aboveground biomass density estimates derived solely from a space-based laser altimeter over boreal forests. Remapping ATL08 to 30 m provided a more detailed account of canopy heights and vegetation structure, which better represents structural heterogeneity across forest gradients that are not well characterized in coarser resolution data. Across the broad expanses of boreal forests, particularly in open forests where the subtleties of forest extent and height can be missed in carbon accounting, this increase of resolution to obtain vegetation structure could hold significant implications for carbon stock estimation.

The next phase of this research is to create a full boreal biomass product through the imputation of the ICESat-2 AGBD estimates against optical imagery. A wall-to-wall biomass product developed for ICESat-2 from the 2020 time period will provide a critical piece of information to quantify the amount of carbon flux associated with forest disturbance for these regions. Over boreal forests, the impact of disturbances (particularly fires) can have a stronger influence on carbon flux due to the long regrowth and recovery time of forests (Rogers et al., 2020). The canopy height and aboveground biomass density products created in this research provide the opportunity to create wall-to-wall predictions across the entire pan-boreal region. Space-based laser altimetry systems such as ICESat-2 and GEDI are sampling tools, and do not provide wall-to-wall outputs. However, these data may be combined with wall-to-wall optical earth observation data from sensors such as Landsat and Sentinel-2 or SAR to generate spatially-explicit maps of AGBD at a spatial resolution ($\leq$30 m) that is relevant for science and management.

5. Conclusions

Space-based lidar has transformed our ability to describe terrestrial ecosystems by directly observing vegetation structure and the terrain beneath it. Advanced information about the structure, distribution, and biomass of global forests provides critical ecological insights and opportunities for sustainable forest management and enhancing forest conservation and ecosystem services. However, the uncertainty of biomass estimates in forests, particularly boreal forests, are largely due to their increasing heterogeneity over the landscape as well as impacts of edaphic, environmental, and disturbance conditions. The objective of

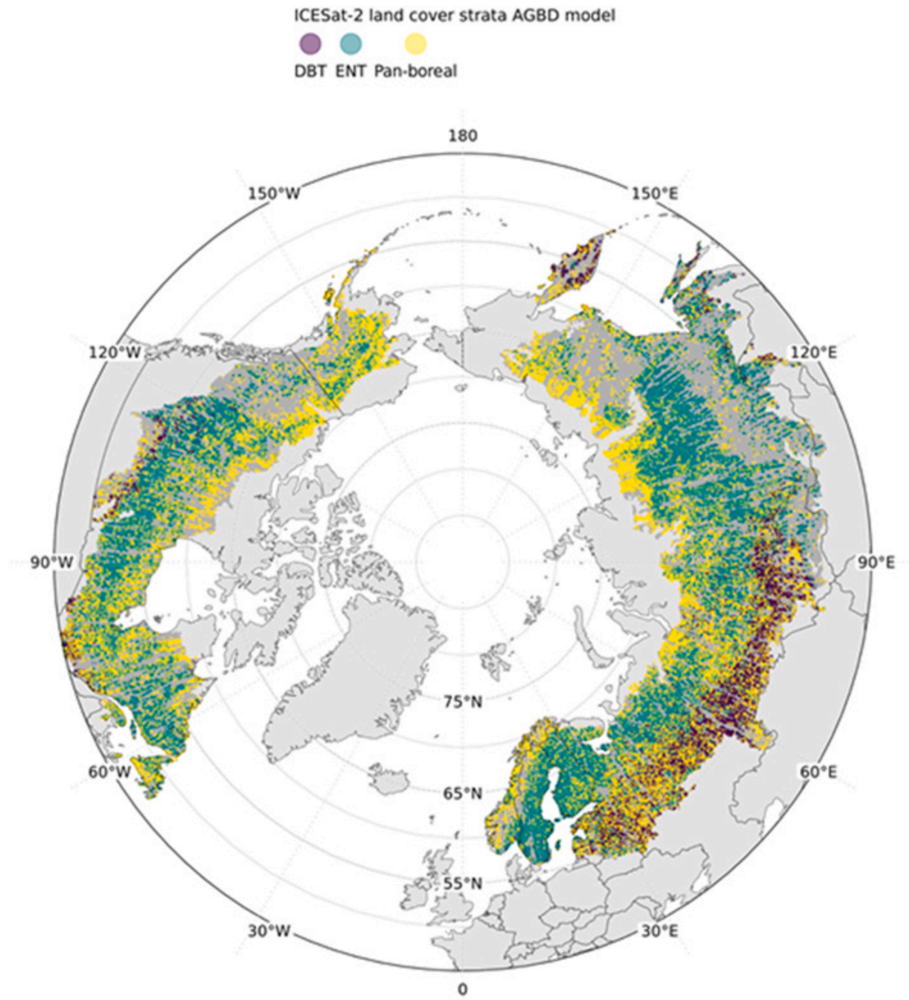

Fig. 9. Distribution of AGBD models applied to the filtered set of ATL08 30 m observations across the study domain.

this research was to create ICESat-2 aboveground biomass density (AGBD) models for boreal forests representing the 2020 era (2019 through 2021) at a 30 m resolution. ICESat-2 is currently the only space-based laser altimeter capable of mapping boreal vegetation and by implementing a similar methodology of estimating AGBD as was done for GEDI, the expectation is that ICESat-2 biomass estimates can complement GEDI's estimates there by enabling a full global accounting of carbon. The AGBD models were created by simulating ICESat-2 metrics over field sites using a suite of field measurements and contemporaneous airborne lidar data. ICESat-2 biomass models were fit specifically for Deciduous Broadleaf and Evergreen Needleleaf forests as well as a boreal-wide model. Based on the results observed in this research, it is anticipated that ICESat-2 biomass models may overpredict AGBD in areas of low cover and underpredict AGBD in areas of high cover until an estimate of canopy cover can be directly calculated from the data product. Some of the errors of over- and under-prediction of biomass are potentially due to mislabeled photons on the data product, however, the majority are likely due to the limitation of ICESat-2's under sampling of vegetation vertical structure. The ICESat-2 data were resampled from 100 m to 30 m based on the individual photon classifications and then filtered based on a series of criteria. Applying the biomass models to a set of over 19 million highly filtered ICESat-2 observations yielded a 30 m along-track aboveground biomass product for the pan-boreal. Remapping the ICESat-2 data to a 30 m along-track resolution allows for a more detailed representation of canopy structure that could have significant implications for total carbon estimates for boreal forests. This demonstrated ability to calculate ICESat-2 biomass estimates at a 30 m resolution sets the stage for a full global accounting of AGBD across the boreal domain. Future versions of this data product are anticipated with updates and improvements to the ATL08 ICESat-2 data product. As a complement to the open availability of the code developed for this project and available at https://github.com/lauraduncanson/icesat2_boreal.), the 30 m along-track canopy height and boreal AGBD product produced for this project can be obtained from the National Snow and Ice Data Center (NSIDC). (https://www.doi.org/10.5067/D2E2L5GK7ER3).

CRediT authorship contribution statement

A. Neuenschwander: Writing – review & editing, Writing – original draft, Validation, Funding acquisition. **L. Duncanson:** Writing – review & editing, Writing – original draft, Software, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **P. Montesano:** Writing – review & editing, Visualization, Software,

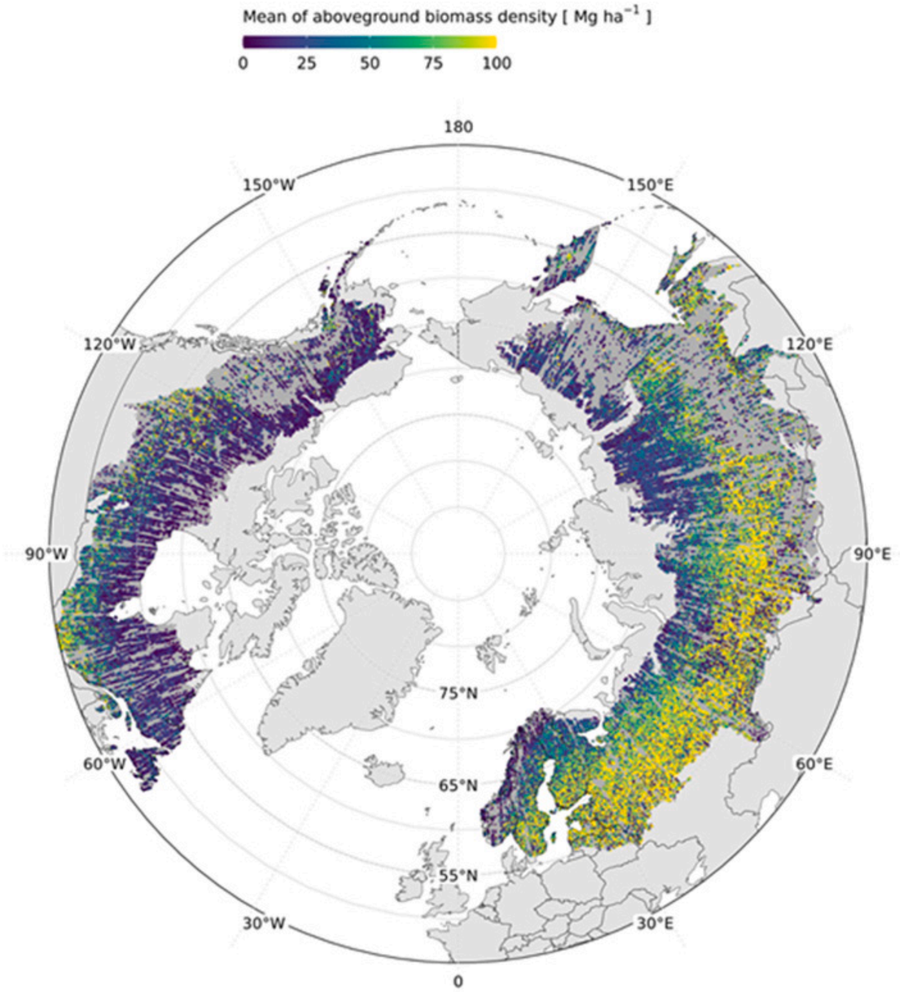

Fig. 10. 30 m along-track aboveground biomass density from ICESat-2 ATLAS data across the pan-boreal study area.

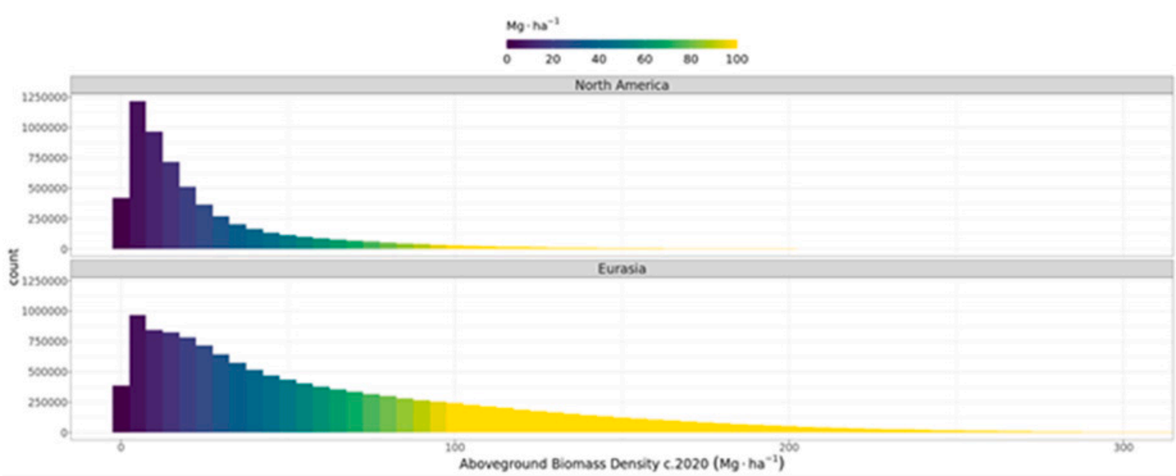

Fig. 11. Histograms of the aboveground biomass density in Mg/ha from ICESat-2 ATLAS observations across North America (top) and Eurasia (bottom).

Methodology, Investigation, Formal analysis. **D. Minor:** Project administration, Data curation. **E. Guenther:** Software, Formal analysis. **S. Hancock:** Writing – review & editing, Software, Investigation, Formal analysis. **M.A. Wulder:** Writing – review & editing, Methodology. **J.C. White:** Writing – review & editing, Methodology. **M. Purslow:** Investigation. **N. Thomas:** Methodology, Investigation. **A. Mandel:** Software, Resources. **T. Feng:** Investigation. **J. Armston:** Software, Conceptualization. **J.R. Kellner:** Conceptualization. **H.E. Andersen:** Data curation. **L. Boschetti:** Data curation. **P. Fekety:** Data curation. **A. Hudak:** Data curation. **J. Pisek:** Data curation. **N. Sánchez-López:** Data curation. **K. Stereńczak:** Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

Funding for this research was made possible by grants from the NASA ABoVE Program (grant number 80NSSC19M0117), NASA ICESat-2 Science Team (grant number 80NSSC20K0965), with support from the and NASA MAAP project. JP was supported by the Estonian Research Council grant PRG1405, SLTTO22345, and by Estonian Ministry of Education and Research, Centre of Excellence for Sustainable Land Use (TK232). Field and remote sensing data collection from Białowieża Forest were supported by Project LIFE+ ForBioSensing (contract number LIFE13ENV/PL/000048) and Poland's National Fund for Environmental Protection and Water Management (contract number 485/2014/WN10/OPNMLF/D). Thank you to the NASA MAAP Data team, and the platform team, especially George Chang, Sujen Shah, and Brian Satorious for their endless support to create the associated open science data products. The authors would also like to thank the collaborators and colleagues that helped collect, process, and share provided field and airborne data used in this research. This work largely built off several years of effort by the GEDI science team for the L4A model fitting exercise, and thus we thank all those who contributed to that effort, and particularly GEDI PI Ralph Dubayah, Sean Healey, Paul Patterson, Svetlana Saarela, Suzanne Marselis, and many others from the international community. All ICESat-2 data used in this study are Release 005 and are publicly available through the National Snow and Ice Data Center (https://nsidc.org/data/icesat-2).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.srs.2024.100150.

References

Albinet, C., Nouvellon, S., Frommknecht, B., Rutakaza, R., Daniel, S., Saüt, C., 2020. MAAP: the mission algorithm and analysis Platform: a new virtual and collaborative environment for the scientific community 19989. https://doi.org/10.5194/egusphere-egu2020-19989.

Armston, J., Disney, M., Lewis, P., Scarth, P., Phinn, S., Lucas, R., Bunting, P., Goodwin, N., 2013. Direct retrieval of canopy gap probability using airborne waveform lidar. Remote Sens. Environ. 134, 24–38. https://doi.org/10.1016/j.rse.2013.02.021.

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., Smets, B., 2020. Copernicus global land cover layers—collection 2. Rem. Sens. 12, 1044. https://doi.org/10.3390/rs12061044.

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

Crawford, C.J., Roy, D.P., Arab, S., Barnes, C., Vermote, E., Hulley, G., Gerace, A., Choate, M., Engebretson, C., Micijevic, E., Schmidt, G., Anderson, C., Anderson, M., Bouchard, M., Cook, B., Dittmeier, R., Howard, D., Jenkerson, C., Kim, M., Kleyians, T., Maiersperger, T., Mueller, C., Neigh, C., Owen, L., Page, B., Pahlevan, N., Rengarajan, R., Roger, J.-C., Sayler, K., Scaramuzza, P., Skakun, S., Yan, L., Zhang, H.K., Zhu, Z., Zahn, S., 2023. The 50-year Landsat collection 2 archive. Sci. Remote Sens. 8, 100103 https://doi.org/10.1016/j.srs.2023.100103.

Dubayah, R., Armston, J., Healey, S.P., Bruening, J.M., Patterson, P.L., Kellner, J.R., Duncanson, L., Saarela, S., Ståhl, G., Yang, Z., Tang, H., Blair, J.B., Fatoyinbo, L., Goetz, S., Hancock, S., Hansen, M., Hofton, M., Hurtt, G., Luthcke, S., 2022. GEDI launches a new era of biomass inference from space. Environ. Res. Lett. 17, 095001 https://doi.org/10.1088/1748-9326/ac8694.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P.L., Qi, W., Silva, C., 2020. The global ecosystem Dynamics investigation: high-resolution laser ranging of the earth's forests and topography. Sci. Remote Sens. 1, 100002 https://doi.org/10.1016/j.srs.2020.100002.

Duncanson, L., Kellner, J.R., Armston, J., Dubayah, R., Minor, D.M., Hancock, S., Healey, S.P., Patterson, P.L., Saarela, S., Marselis, S., Silva, C.E., Bruening, J., Goetz, S.J., Tang, H., Hofton, M., Blair, B., Luthcke, S., Fatoyinbo, L., Abernethy, K., Alonso, A., Andersen, H.-E., Aplin, P., Baker, T.R., Barbier, N., Bastin, J.F., Biber, P., Boeckx, P., Bogaert, J., Boschetti, L., Boucher, P.B., Boyd, D.S., Burslem, D.F.R.P., Calvo-Rodriguez, S., Chave, J., Chazdon, R.L., Clark, D.B., Clark, D.A., Cohen, W.B., Coomes, D.A., Corona, P., Cushman, K.C., Cutler, M.E.J., Dalling, J.W., Dalponte, M., Dash, J., de-Miguel, S., Deng, S., Ellis, P.W., Erasmus, B., Fekety, P.A., Fernandez-Landa, A., Ferraz, A., Fischer, R., Fisher, A.G., García-Abril, A., Gobakken, T., Hacker, J.M., Heurich, M., Hill, R.A., Hopkinson, C., Huang, H., Hubbell, S.P., Hudak, A.T., Huth, A., Imbach, B., Jeffery, K.J., Katoh, M., Kearsley, E., Kenfack, D., Kljun, N., Knapp, N., Král, K., Krůček, M., Labrière, N., Lewis, S.L., Longo, M., Lucas, R.M., Main, R., Manzanera, J.A., Martínez, R.V., Mathieu, R., Memiaghe, H., Meyer, V., Mendoza, A.M., Monerris, A., Montesano, P., Morsdorf, F., Næsset, E., Naidoo, L., Nilus, R., O'Brien, M., Orwig, D.A., Papathanassiou, K., Parker, G., Philipson, C., Phillips, O.L., Pisek, J., Poulsen, J.R., Pretzsch, H., Rüdiger, C., Saatchi, S., Sanchez-Azofeifa, A., Sanchez-Lopez, N., Scholes, R., Silva, C.A., Simard, M., Skidmore, A., Stereńczak, K., Tanase, M., Torresan, C., Valbuena, R., Verbeeck, H., Vrska, T., Wessels, K., White, J.C., White, L.J.T., Zahabu, E., Zgraggen, C., 2022. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. Remote Sens. Environ. 270, 112845 https://doi.org/10.1016/j.rse.2021.112845.

Duncanson, L.I., Niemann, K.O., Wulder, M.A., 2010. Estimating forest canopy height and terrain relief from GLAS waveform metrics. Remote Sens. Environ. 114, 138–154. https://doi.org/10.1016/j.rse.2009.08.018.

Estilow, T.W., Young, A.H., Robinson, D.A., 2015. A long-term Northern Hemisphere snow cover extent data record for climate studies and monitoring. Earth Syst. Sci. Data 7, 137–142. https://doi.org/10.5194/essd-7-137-2015.

Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A.Z., Schepaschenko, D.G., 2015. Boreal forest health and global change. Science 349, 819–822. https://doi.org/10.1126/science.aaa9092.

Girona, M.M., Morin, H., Gauthier, S., Bergeron, Y. (Eds.), 2023. Boreal Forests in the Face of Climate Change: Sustainable Management. Springer Nature. https://doi.org/10.1007/978-3-031-15988-6.

Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L.I., Kellner, J.R., Dubayah, R., 2019. The GEDI simulator: a large-footprint waveform lidar simulator for calibration and validation of spaceborne missions. Earth Space Sci. 6, 294–310. https://doi.org/10.1029/2018EA000506.

Hilbert, C., Schmullius, C., 2012. Influence of surface topography on ICESat/GLAS forest height estimation and waveform shape. Rem. Sens. 4, 2210–2235. https://doi.org/10.3390/rs4082210.

Kellner, J.R., Armston, J., Duncanson, L., 2023. Algorithm theoretical basis document for GEDI footprint aboveground biomass density. Earth Space Sci. 10, e2022EA002516 https://doi.org/10.1029/2022EA002516.

Larsen, J.A., 1980. The Boreal Ecosystem, pp. 1–512.

Lee, S., Ni-Meister, W., Yang, W., Chen, Q., 2011. Physically based vertical vegetation structure retrieval from ICESat data: validation using LVIS in White Mountain National Forest, New Hampshire, USA. Remote Sens. Environ., DESDynI VEG-3D Special Issue 115, 2776–2785. https://doi.org/10.1016/j.rse.2010.08.026.

Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar Remote Sensing for Ecosystem Studies: lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. Bioscience 52, 19–30. https://doi.org/10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2.

Magruder, L., Neumann, T., Kurtz, N., 2021. ICESat-2 early mission synopsis and observatory performance. Earth Space Sci. 8, e2020EA001555 https://doi.org/10.1029/2020EA001555.

Margolis, H.A., Nelson, R.F., Montesano, P.M., Beaudoin, A., Sun, G., Andersen, H.-E., Wulder, M.A., 2015. Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America. Can. J. For. Res. 45, 838–855. https://doi.org/10.1139/cjfr-2015-0006.

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., Popescu, S., Shum, C., Schutz, B.E., Smith, B., Yang, Y., Zwally, J., 2017. The ice, cloud, and land elevation satellite-2 (ICESat-2): science requirements, concept, and implementation. Remote Sens. Environ. 190, 260–273. https://doi.org/10.1016/j.rse.2016.12.029.

Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Zald, H.S.J., 2018. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. Remote Sens. Environ. 209, 90–106. https://doi.org/10.1016/j.rse.2017.12.020.

Neigh, C.S., Nelson, R.F., Ranson, K.J., Margolis, H., Montesano, P.M., Sun, G., Kharuk, V., Naesset, E., Wulder, M.A., Andersen, H.E., 2015. LiDAR-based Biomass Estimates, Boreal Forest Biome, Eurasia, 2005-2006. ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1278.

Nelson, R., 2010. Model effects on GLAS-based regional estimates of forest biomass and carbon. Int. J. Rem. Sens. 31, 1359–1372. https://doi.org/10.1080/01431160903380557.

Neuenschwander, A., Guenther, E., White, J.C., Duncanson, L., Montesano, P., 2020. Validation of ICESat-2 terrain and canopy heights in boreal forests. Remote Sens. Environ. 251, 112110 https://doi.org/10.1016/j.rse.2020.112110.

Neuenschwander, A., Magruder, L., Guenther, E., Hancock, S., Purslow, M., 2022. Radiometric assessment of ICESat-2 over vegetated surfaces. Rem. Sens. 14, 787. https://doi.org/10.3390/rs14030787.

Neuenschwander, A., Pitts, K., 2019. The ATL08 land and vegetation product for the ICESat-2 Mission. Remote Sens. Environ. 221, 247–259. https://doi.org/10.1016/j.rse.2018.11.005.

Neuenschwander, A.L., Magruder, L.A., 2016. The potential impact of vertical sampling uncertainty on ICESat-2/ATLAS terrain and canopy height retrievals for multiple ecosystems. Rem. Sens. 8, 1039. https://doi.org/10.3390/rs8121039.

Neuenschwander, A.L., Urban, T.J., Gutierrez, R., Schutz, B.E., 2008. Characterization of ICESat/GLAS waveforms over terrestrial ecosystems: implications for vegetation mapping. J. Geophys. Res. Biogeosciences 113. https://doi.org/10.1029/2007JG000557.

Neumann, T.A., Martino, A.J., Markus, T., Bae, S., Bock, M.R., Brenner, A.C., Brunt, K.M., Cavanaugh, J., Fernandes, S.T., Hancock, D.W., Harbeck, K., Lee, J., Kurtz, N.T., Luers, P.J., Luthcke, S.B., Magruder, L., Pennington, T.A., Ramos-Izquierdo, L., Rebold, T., Skoog, J., Thomas, T.C., 2019. The ice, cloud, and land elevation satellite – 2 mission: a global geolocated photon product derived from the advanced topographic laser altimeter system. Remote Sens. Environ. 233, 111325 https://doi.org/10.1016/j.rse.2019.111325.

Oliver, C.D., Larson, B.A., 1996. Forest stand Dynamics. Yale sch. Environ. Times 1.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 4540. https://doi.org/10.1038/s41467-020-18321-y.

Purslow, M., Hancock, S., Neuenschwander, A., Armston, J., Duncanson, L., 2023. Can ICESat-2 estimate stand-level plant structural traits? Validation of an ICESat-2 simulator. Sci. Remote Sens. 7, 100086 https://doi.org/10.1016/j.srs.2023.100086.

Queinnec, M., 2022. Enhancing the Characterization of Forest Structure and Resources for Forest Inventory and Monitoring Using Airborne and Spaceborne Single Photon Lidar. University of British Columbia. https://doi.org/10.14288/1.0422379.

Rogers, B.M., Balch, J.K., Goetz, S.J., Lehmann, C.E.R., Turetsky, M., 2020. Focus on changing fire regimes: interactions with climate, ecosystems, and society. Environ. Res. Lett. 15, 030201 https://doi.org/10.1088/1748-9326/ab6d3a.

Schutz, B.E., Zwally, H.J., Shuman, C.A., Hancock, D., DiMarzio, J.P., 2005. Overview of the ICESat mission. Geophys. Res. Lett. 32 https://doi.org/10.1029/2005GL024009.

Varvia, P., Korhonen, L., Bruguière, A., Toivonen, J., Packalen, P., Maltamo, M., Saarela, S., Popescu, S.C., 2022. How to consider the effects of time of day, beam strength, and snow cover in ICESat-2 based estimation of boreal forest biomass? Remote Sens. Environ. 280, 113174 https://doi.org/10.1016/j.rse.2022.113174.