

GEDI L4A Footprint Level Aboveground Biomass Density, Golden Weeks, Version 1

Documentation Revision Date: 2022-02-15

Dataset Version: 1.1

Summary

This dataset contains Global Ecosystem Dynamics Investigation (GEDI) Level 4A (L4A) predictions of the aboveground biomass density (AGBD; in Mg/ha) and estimates of the prediction standard error within each sampled geolocated laser footprint. The footprints are located within the global latitude band observed by the International Space Station (ISS), nominally 51.6 degrees N and S, and reported for the mission weeks 19, 32, 34, and 38 (a.k.a., Golden Weeks). These weeks cover the range of instrument operating conditions important for calibration and validation of geolocation algorithms, and also include GEDI orbits that are coincident with underflights acquired by the LVIS (Land, Vegetation, and Ice Sensor) airborne lidar instrument. The GEDI instrument consists of three lasers producing a total of eight beam ground transects, which instantaneously sample eight ~25 m footprints spaced approximately every 60 m along-track. The GEDI beam transects are spaced approximately 600 m apart on the Earth's surface in the cross-track direction, for an across-track width of ~4.2 km. Footprint AGBD was derived from parametric models that relate simulated GEDI Level 2A (L2A) waveform relative height (RH) metrics to field plot estimates of AGBD. Height metrics from simulated waveforms associated with field estimates of AGBD from multiple regions and plant functional types (PFT) were compiled to generate a calibration dataset for models representing the combinations of world regions and PFTs (i.e., deciduous broadleaf trees, evergreen broadleaf trees, evergreen needleleaf trees, deciduous needleleaf trees, and the combination of grasslands, shrubs, and woodlands).

Reported with the AGBD estimates for each of the eight beams are the associated uncertainty metrics, quality flags, and model inputs including the scaled and transformed GEDI L2A RH metrics, and other information about the GEDI L2A waveform for this selected algorithm setting group. Also provided are model inputs for each of the eight beams including footprint geolocation variables and land cover input data including PFTs and the world region identifiers. Additional model outputs include the AGBD predictions for each of the six GEDI L2A algorithm setting groups with AGBD in natural and transformed units and associated prediction uncertainty for each GEDI L2A algorithm setting group. Providing these ancillary data products will allow users to evaluate and select alternative algorithm setting groups. Also provided are outputs of parameters and variables from the L4A models used to generate AGBD predictions that are required as input to the GEDI04_B algorithm to generate 1 km gridded products.

There are 386 data files in HDF5 (*.h5) format included in this dataset and four companion files that provide additional details regarding the product model development and variable descriptions.

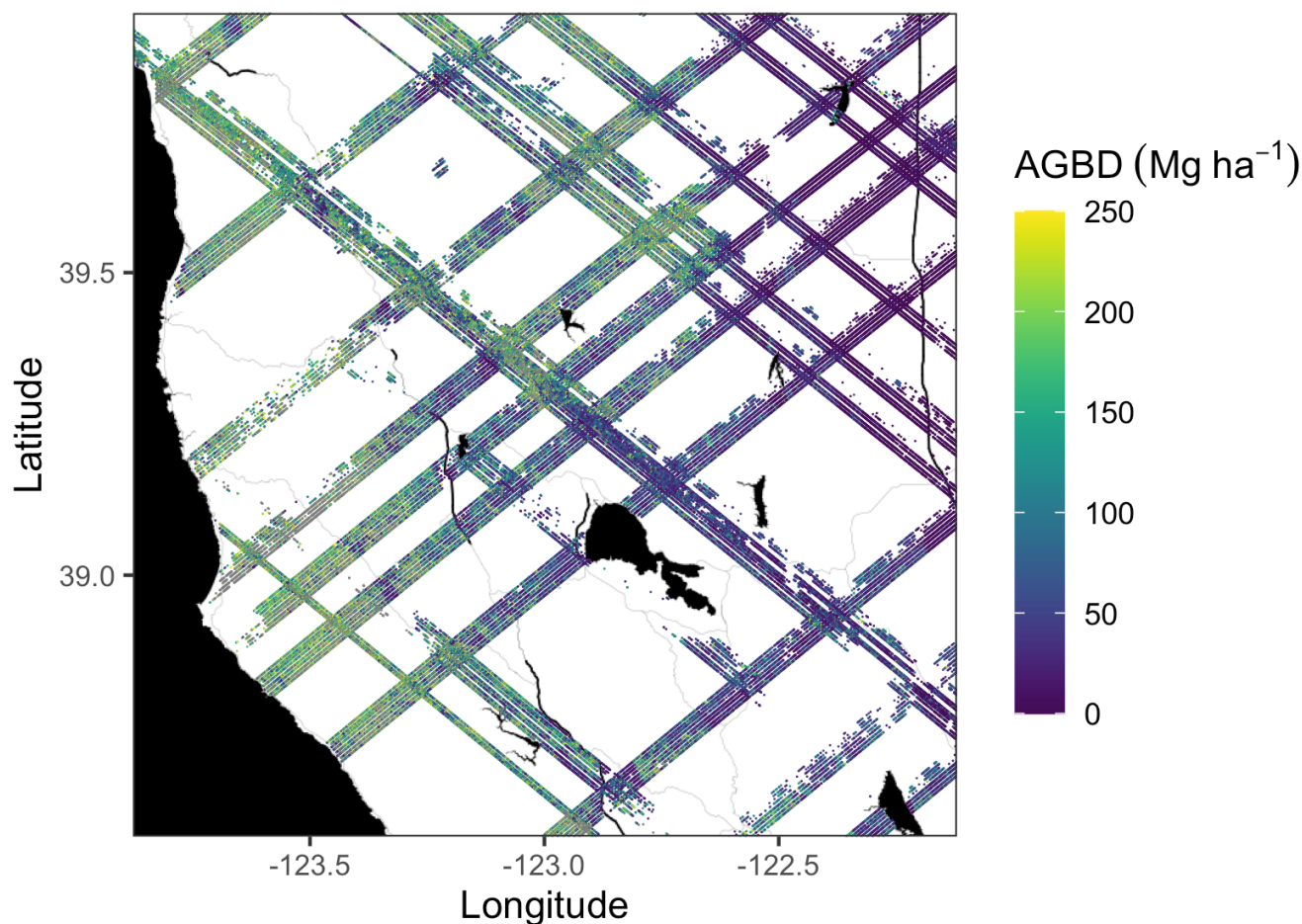


Figure 1. Example subset of aboveground biomass density (AGBD; Mg ha⁻¹) predictions from the GEDI Level 4A footprint product over northern California, U.S., spanning April to July 2019. GEDI footprints are spaced 60 m along-track and 600 m across-track.

Citation

Dubayah, R.O., J. Armston, J.R. Kellner, L. Duncanson, S.P. Healey, P.L. Patterson, S. Hancock, H. Tang, M.A. Hofton, J.B. Blair, and S.B. Luthcke. 2022. GEDI L4A Footprint Level Aboveground Biomass Density, Golden Weeks, Version 1. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAAC/2028>

Table of Contents

1. Dataset Overview
2. Data Characteristics
3. Application and Derivation
4. Quality Assessment
5. Data Acquisition, Materials, and Methods
6. References
7. Dataset Revisions

1. Dataset Overview

This dataset contains Global Ecosystem Dynamics Investigation (GEDI) Level 4A (L4A) predictions of the aboveground biomass density (AGBD; in Mg/ha) and estimates of the prediction standard error within each sampled geolocated laser footprint. The footprints are located within the global latitude band observed by the International Space Station (ISS), nominally 51.6 degrees N and S, and reported for the mission weeks 19, 32, 34, and 38 (a.k.a., Golden Weeks). These weeks cover the range of instrument operating conditions important for calibration and validation of geolocation algorithms, and also include GEDI orbits that are coincident with underflights acquired by the LVIS (Land, Vegetation, and Ice Sensor) airborne lidar instrument. The GEDI instrument consists of three lasers producing a total of eight beam ground transects, which instantaneously sample eight ~25 m footprints spaced approximately every 60 m along-track. The GEDI beam transects are spaced approximately 600 m apart on the Earth's surface in the cross-track direction, for an across-track width of ~4.2 km. Footprint AGBD was derived from parametric models that relate simulated GEDI Level 2A (L2A) waveform relative height (RH) metrics to field plot estimates of AGBD. Height metrics from simulated waveforms associated with field estimates of AGBD from multiple regions and plant functional types (PFT) were compiled to generate a calibration dataset for models representing the combinations of world regions and PFTs (i.e., deciduous broadleaf trees, evergreen broadleaf trees, evergreen needleleaf trees, deciduous needleleaf trees, and the combination of grasslands, shrubs, and woodlands).

Reported with the AGBD estimates for each of the eight beams are the associated uncertainty metrics, quality flags, and model inputs including the scaled and transformed GEDI L2A RH metrics and other information about the GEDI L2A waveform for this selected algorithm setting group. Also provided are model inputs for each of the eight beams including footprint geolocation variables and land cover input data including PFTs and the world region identifiers. Additional model outputs include the AGBD predictions for each of the six GEDI L2A algorithm setting groups with AGBD in natural and transformed units and associated prediction uncertainty for each GEDI L2A algorithm setting group. Providing these ancillary data products will allow users to evaluate and select alternative algorithm setting groups. Also provided are outputs of parameters and variables from the L4A models used to generate AGBD predictions that are required as input to the GEDI04_B algorithm to generate 1 km gridded products.

Project: Global Ecosystem Dynamics Investigation

The Global Ecosystem Dynamics Investigation (GEDI) produces high resolution laser ranging observations of the 3D structure of the Earth. GEDI's precise measurements of forest canopy height, canopy vertical structure, and surface elevation greatly advance our ability to characterize important carbon and water cycling processes, biodiversity, and habitat. GEDI was funded as a NASA Earth Ventures Instrument (EVI) mission. It was launched to the International Space Station in December 2018 and completed initial orbit checkout in April 2019.

Related Publication

Dubayah, R., J.B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, G. Hurtt, J. Kellner, S. Luthcke, J. Armston, H. Tang, L. Duncanson, S. Hancock, P. Jantz, S. Marselis, P.L. Patterson, W. Qi, and C. Silva. 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing* 1:100002. <https://doi.org/10.1016/j.srs.2020.100002>

Duncanson, L., J.R. Kellner, J. Armston, R. Dubayah, D.M. Minor, S. Hancock, et al. 2022. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sensing of Environment* 270:112845. <https://doi.org/10.1016/j.rse.2021.112845>

Related Datasets

Dubayah, R.O., J. Armston, J.R. Kellner, L. Duncanson, S.P. Healey, P.L. Patterson, S. Hancock, H. Tang, M.A. Hofton, J.B. Blair, and S.B. Luthcke. 2021. GEDI L4A Footprint Level Aboveground Biomass Density, Version 1. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAAC/1907>

- Version 1 release of GEDI L4A Footprint Level Aboveground Biomass Density dataset containing data derived from observations during GEDI mission weeks 19 to 90 (2019-04-18 to 2020-09-02). Now superseded by Version 2 and available upon request.

Dubayah, R.O., J. Armston, J.R. Kellner, L. Duncanson, S.P. Healey, P.L. Patterson, S. Hancock, H. Tang, J. Bruening, M.A. Hofton, J.B. Blair, and S.B. Luthcke. 2021. GEDI L4A Footprint Level Aboveground Biomass Density, Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAAC/1986>

- Version 2 release of GEDI L4A Footprint Level Aboveground Biomass Density dataset containing data derived from observations during GEDI mission weeks 19 to 138 (2019-04-18 to 2021-08-05).

Level 1B, Level 2A, and Level 2B data from GEDI are available from the Land Processes Distributed Active Archive Center at <https://lpdaac.usgs.gov/>.

Acknowledgments

This work was funded by NASA contract #NNL 15AA03C to the University of Maryland for the development and execution of the GEDI mission (Dubayah, Principal Investigator). We thank the NASA Terrestrial Ecology Program and Hank Margolis for supporting the GEDI mission, and the University of Maryland for providing independent financial support. We thank Jamis Bruening, Suzanne Marselis, David Minor, and Carlos E. Silva for contributing to the development and management of the GEDI Forest Structure and Biomass Database. We gratefully acknowledge the GEDI Science Team and numerous collaborators who generously contributed field estimates of AGBD, stem maps, and airborne lidar data. These people include Katharine Abernethy, Hans-Erik Andersen, Paul Aplin, Timothy R. Baker, Nicolas Barbier, Jean Francois Bastin, Pascal Boeckx, Jan Bogaert, Luigi Boschetti, Peter Brehm Boucher, Doreen S. Boyd, Patrick Burns, David F.R.P. Burslem, Sofia Calvo-Rodriguez, Jérôme Chave, Robin L. Chazdon, David B. Clark, Deborah A. Clark, Warren B. Cohen, David A. Coomes, Piermaria Corona, K.C. Cushman, Mark E. J. Cutler, James William Dalling, Michele Dalponte, Sergio de-Miguel, Songqiu Deng, Peter Woods Ellis, Barend Erasmus, Michael Falkowski, Temilola Fatoyinbo, Patrick A. Fekety, Alfredo Fernández-Landa, Antonio Ferraz, Rico Fischer, Adrian G. Fisher, Antonio García-Abril, Terje Gobakken, Scott J. Goetz, Jonathan A. Greenberg, Jorg M. Hacker, Matt Hansen, Marco Heurich, Ross A. Hill, Sören Holm, Chris Hopkinson, Chengquan Huang, Huabing Huang, Stephen P. Hubbell, Andrew T. Hudak, George Hurtt, Andreas Huth, Benedikt Imbach, Patrick Jantz, Kathryn Jeffery, Masato Katoh, Elizabeth Kearsley, Natascha Kljun, Nikolai Knapp, Kamil Král, Martin Krůček, Nicolas Labrière, Seung-kuk Lee, Simon L. Lewis, Marcos Longo, Richard M. Lucas, Scott Luthcke, Russell Main, Jose A. Manzanera, Suzanne Marselis, Rodolfo Vásquez Martínez, Renaud Mathieu, Victoria Meyer, Paul Montesano, Felix Morsdorf, Erik Næsset, Laven Naidoo, Reuben Nilus, Michael J. O'Brien, David A. Orwig, Geoffrey Parker, Paul Patterson, Christopher Philipson, Oliver L. Phillips, Jan Pisek, Jim Pontius, John R. Poulsen, Wenlu Qi, Christoph Rüdiger, Svetlana Saarela, Sassan Saatchi, Arturo Sanchez-Azofeifa, Nuria Sanchez-Lopez, Crystal B. Schaff, Marc Simard, Andrew Kerr Skidmore, Göran Ståhl, Krzysztof Stereńczak, Chiara Torresan, Rubén Valbuena, Hans Verbeeck, Tomas Vrska, Konrad Wessels, Joanne C. White, and Carlo Zraggen.

2. Data Characteristics

Spatial Coverage: Global within a latitude extent of -52 to 52 degrees

Spatial Resolution: Footprints with ~25 m in diameter

Temporal Coverage: 2019-04-18 to 2019-04-24 (mission week 19); 2019-07-18 to 2019-07-24 (mission week 32); 2019-08-01 to 2019-08-07 (mission week 34); and 2019-08-29 to 2019-09-04 (mission week 38)

Temporal Resolution: One-time estimate

Study Area: Latitude and longitude are given in decimal degrees.

Site	Westernmost Longitude	Easternmost Longitude	Northernmost Latitude	Southernmost Latitude
Global	-180	180	52.46	-52.20

Data File Information

There are 386 data files in HDF5 (*.h5) format included in this dataset. Each file provides multiple datasets/groups for each of the eight beams with valid data (0000, 0001, 0010, 0011, 0101, 0110, 1000, 1011).

The files are named GEDI04_A_YYYYDDHHMMSS_O_T_P_R_V.h5 (e.g., GEDI04_A_2019244182445_O04082_T01002_O2_001_01), where:

- GEDI04_A = product short name representing GEDI Level 4A data,
- YYYYDDHHMMSS = date and time of acquisition in Julian day of year, hours, minutes, and seconds format,
- O = orbit number,
- T = track number,
- P = positioning and pointing determination system (PPDS) type (00 is "predict", 01 is "rapid", 02 and higher is "final"),
- R = release number 001, representing the SOC SDS (software) release that corresponds to the version number of this dataset,
- V = production version, and
- h5 = file extension, HDF5 format.

Table 1. File names and descriptions.

File Name	Description
Data Files	
GEDI04_A_YYYYDDHHMMSS_O_T_P_R_V.h5	Each contains information and data in METADATA and BEAMXXXX groups and three ANCILLARY group datasets.
Companion Files	
GEDI_ATBD_L4A_v1.0.pdf	Algorithm Theoretical Basis Document (ATBD) for GEDI L4A Footprint Aboveground Biomass Density Product (current dataset).
GEDI_L4A_AGB_Density_GW.pdf	A PDF version of this user guide.

File Name	Description
GEDI_L4A_Common_Queries.pdf	Common data questions and answers on how to use and interpret the GEDI L4A product. This information is also provided in Section 5 of this user guide.
GEDI_L4A_Product_Data_Dictionary.pdf	Data product dictionary that provides detailed information about each variable included in the data files.

File Organization

Each GEDI04_A granule contains information and data in METADATA and BEAMXXXX groups and three ANCILLARY group datasets.

The METADATA group contains data identification information. See companion file GEDI_L4A_Product_Data_Dictionary.pdf for details.

The BEAMXXXX root group (Table 2) contains the AGBD prediction, associated uncertainty metrics, quality flags, and model inputs including the scaled and transformed GEDI02_A RH metrics and other information about the waveform for the selected algorithm setting group.

- The BEAMXXXX/geolocation group (Table 3) contains the input variables elevation, latitude, longitude, sensitivity, and other information for each algorithm setting group.
- The BEAMXXXX/land_cover_data group (Table 4) contains land cover data extracted from external data sources, including Landsat tree cover, Landsat water persistence, a modified version of MCD12Q1 V006 PFT, the world region identifier, urban proportion derived from the DLR TanDEM-X Global Urban Footprint (GUF) dataset, and leaf-off and leaf-on flags derived from the VNP22Q2 land surface phenology product. The PFT and world region identifiers used in L4A are described further under "Data acquisition, materials, and methods".
- The BEAMXXXX/agbd_prediction group (Table 5) contains ancillary information, AGBD predictions in natural and transformed units, and associated prediction uncertainty for predictions for each of the six GEDI L2A algorithm setting groups. Providing these data allows the user to evaluate and select alternative algorithm setting groups.

The ANCILLARY/model_data group (Table 6) provides parameters and variables from the L4A models used to generate predictions. All the model parameters and uncertainty estimates (e.g. variance-covariance matrix of the model parameters) required as input to the GEDI04_B algorithm to generate 1 km gridded products are provided.

ANCILLARY/pft_lut (Table 7) and ANCILLARY/region_lut (Table 8) are look-up tables that link a numeric value from gridded PFT or world region to a descriptive text name.

What is the algorithm setting group?

Investigators applied a sophisticated post-processing algorithm to the received waveforms from the GEDI instrument to detect weaker waveform signals. The "algorithm setting group" defines the specific set of parameters used in an algorithm run. There are six (i.e., 1, 2, 3, 4, 5, 6, and 10¥) defined groups.

Each algorithm run's output and externally-set parameters are available in the L2A data product, within the 'rx_processing_a<n>' subgroup. For details refer to Hofton and Blair, 2020.

In the L4A data products, the /geolocation group and the agbd_prediction group report footprint data for each "algorithm setting group". The variables are *_aN, where N is 1, 2, 3, 4, 5, 6, or 10¥. In the BEAMXXXX root group, the reported AGBD prediction value is for the selected algorithm setting group. The selected "algorithm setting group" is contained in the selected_algorithm variable. The selected AGBD value is reported in the agbd_prediction group dataset.

¥ Note that a value of 10 indicates algorithm setting group 5 has been used, but that the lowest detected mode is likely a noise detection. When this occurs, a higher mode has been used to calculate RH metrics (Hofton and Blair, 2020).

Variables in the L4A footprint data files: Data are inputs from L2A (Source L2A) and outputs of the GEDI04_A algorithm, descriptors, and quality flags. Data files are provided for each beam.

Table 2. Variables in the Aboveground Biomass Density (agbd) group. These variables include the AGBD prediction, associated uncertainty metrics, quality flags, the scaled and transformed GEDI02_A RH metrics, and other information about the waveform for the selected algorithm setting group. Note: "-" indicates the variable is unitless and input variables from the GEDI02_A data product are marked with "L2A" as the source.

Variables	Units (Source)	Description
agbd	Mg/ha	Aboveground biomass density (Mg/ha)
agbd_pi_lower	Mg/ha	Lower prediction interval (see alpha attribute for the level)
agbd_pi_upper	Mg/ha	Upper prediction interval (see alpha attribute for the level)
agbd_se	Mg/ha	Aboveground biomass density (Mg/ha) prediction standard error
agbd_t	-	Model prediction in fit units
agbd_t_se	-	Model prediction standard error in fit units (needed for calculation of custom prediction intervals)
algorithm_run_flag	-	The L4A algorithm is run if this flag is set to 1. This flag selects data that have sufficient waveform fidelity for AGBD estimation.

Variables	Units (Source)	Description
beam	- (L2A)	Beam identifier
channel	- (L2A)	Channel identifier
degrade_flag	- (L2A)	Flag indicating degraded state of pointing and/or positioning information
delta_time	s (L2A)	Time delta since Jan 1 00:00 2018.
elev_lowestmode	m (L2A)	Elevation of center of lowest mode relative to reference ellipsoid
l2_quality_flag	-	Flag identifying the most useful L2 data for biomass predictions
l4_quality_flag	-	Flag simplifying selection of most useful biomass predictions
lat_lowestmode	degrees (L2A)	Latitude of center of lowest mode
lon_lowestmode	degrees (L2A)	Longitude of center of lowest mode
master_frac	s (L2A)	Master time, fractional part. master_int+master_frac is equivalent to /BEAMXXXX/delta_time.
master_int	s (L2A)	Master time, integer part. Seconds since master_time_epoch. master_int+master_frac is equivalent to /BEAMXXXX/delta_time.
predict_stratum	-	Character ID of the prediction stratum name for the 1 km cell
predictor_limit_flag	-	Prediction stratum identifier. Predictor value is outside the bounds of the training data (0=in bounds; 1=lower bound; 2=upper bound)
response_limit_flag	-	Prediction value is outside the bounds of the training data (0=in bounds; 1=lower bound; 2=upper bound)
selected_algorithm	- (L2A)	selected_algorithm
selected_mode	- (L2A)	ID of mode selected as lowest non-noise mode
selected_mode_flag	- (L2A)	Flag indicating status of selected_mode
sensitivity	- (L2A)	Maximum canopy cover that can be penetrated considering the SNR of the waveform
shot_number	- (L2A)	Shot number
solar_elevation	degrees (L2A)	Solar elevation angle
surface_flag	- (L2A)	Indicates elev_lowestmode is within 300m of Digital Elevation Model (DEM) or Mean Sea Surface (MSS) elevation
xvar	-	Predictor variables (offset and transformation have been applied)

Table 3. Variables in the Geolocation group. This group contains elevation, latitude, longitude and other information for each algorithm selection group (1, 2, 3, 4, 5, 6, and 10). Note: “-” indicates the variable is unitless and input variables from the GEDI02_A data product are marked with “L2A” as the source.

Variables	Units (Source)	Description
elev_lowestmode_aN	m (L2A)	Elevation of center of lowest mode relative to reference ellipsoid
lat_lowestmode_aN	degrees (L2A)	Latitude of center of lowest mode
lon_lowestmode_aN	degrees (L2A)	Longitude of center of lowest mode
sensitivity_aN	- (L2A)	Maximum canopy cover that can be penetrated considering the SNR of the waveform
shot_number	- (L2A)	Shot number
stale_return_flag	- (L2A)	Flag from digitizer indicating the real-time pulse detection algorithm did not detect a return signal above its detection threshold within the entire 10 km search window.

Table 4. Variables in the Landcover Group. This group contains land cover data extracted from external data sources, including Landsat tree cover, Landsat water persistence, a modified version of MCD12Q1 V006 PFT, the world region identifier, the TanDEM-X global urban footprint classification, and leaf-off and leaf-on flags. The PFT and world region identifier used in L4A are described further in Section 5 of this document. Note: “-” indicates the variable is unitless and input variables from the GEDI02_A data product are marked with “L2A” as the source.

Variables	Units (Source)	Description
landsat_treecover	percent (L2A)	Tree cover in the year 2010, defined as canopy closure for all vegetation taller than 5 m in height (Hansen et al., 2013) and encoded as a percentage per output grid cell.
landsat_water_persistence	percent	The percent UMD GLAD Landsat observations with classified surface water between 2018 and 2019 (>80=represent permanent water, <10=permanent land).
leaf_off_doy	days	GED1 1 km EASE 2.0 grid leaf-off start day-of-year derived from the NPP VIIRS Global Land Surface Phenology Product.
leaf_off_flag	-	GED1 1 km EASE 2.0 grid flag derived from leaf_off_doy, leaf_on_doy, and pft_class, indicating if the observation was recorded during leaf-off conditions in deciduous needleleaf or broadleaf forests and woodlands (1=leaf-off, 0=leaf-on).
leaf_on_cycle	-	Flag that indicates the vegetation growing cycle for leaf-on observations (0=leaf-off conditions, 1=cycle 1, 2=cycle 2).
leaf_on_doy	-	GED1 1 km EASE 2.0 grid leaf-on start day- of-year derived from the NPP VIIRS Global Land Surface Phenology product.
pft_class	-	GED1 1 km EASE 2.0 grid Plant Functional Type (PFT) derived from the MODIS MCD12Q1 V006 product. Values follow the Land Cover Type 5 Classification scheme.
region_class	-	GED1 1 km EASE 2.0 grid world continental regions (0=Water, 1=Europe, 2=North Asia, 3=Australasia, 4=Africa, 5=South Asia, 6=South America, 7=North America).
shot_number	- (L2A)	Shot number
urban_focal_window_size	pixels	The focal window size used to calculate urban_proportion (3=3x3 pixel window size, 5=5x5 pixel window size).
urban_proportion	percent (TanDEM-X)	The percentage proportion of land area within a focal area surrounding each shot that is urban land cover. Urban land cover was derived from the DLR 12 m resolution TanDEM-X Global Urban Footprint Product.

Table 5. Variables in the Aboveground Biomass Prediction group. This group contains ancillary information, AGBD predictions in natural and transformed units, and associated prediction uncertainty for each algorithm setting group. Providing these data allows the user to evaluate and select alternative algorithm setting groups. Note: “-” indicates the variable is unitless and input variables from the GEDI02_A data product are marked with “L2A” as the source.

Variables	Units (Source)	Description
agbd_aN	Mg/ha	Above ground biomass density; Geolocation latitude lowestmode (_aN=a1, 2, 3, 4, 5, 6, and a10).
agbd_pi_lower_aN	Mg/ha	Above ground biomass density lower prediction interval (_aN=a1, 2, 3, 4, 5, 6, and a10).
agbd_pi_upper_aN	Mg/ha	Above ground biomass density upper prediction interval (_aN=a1, 2, 3, 4, 5, 6, and a10).
quality_flag_aN	-	Flag identifying the most useful L2 data for biomass prediction (_aN=a1, 2, 3, 4, 5, 6, and a10).
predictor_limit_flag_aN	-	Predictor value is outside the bounds of the training data (_aN=a1, 2, 3, 4, 5, 6, and a10).
agbd_se_aN	Mg/ha	Aboveground biomass density (Mg/ha) prediction standard error (_aN=a1, 2, 3, 4, 5, 6, and a10).
selected_mode_aN	-	ID of mode selected as lowest non-noise mode (_aN=a1, 2, 3, 4, 5, 6, and a10).
selected_mode_flag_aN	-	Flag indicating status of selected mode (_aN=a1, 2, 3, 4, 5, 6, and a10).
xvar_aN	-	Predictor variables (_aN=a1, 2, 3, 4, 5, 6, and a10).
agbd_t_aN	Mg/ha	Aboveground biomass density model prediction in transform space (_aN=a1, 2, 3, 4, 5, 6, and a10).
agbd_t_pi_lower_aN	Mg/ha	Lower prediction interval in transform space (_aN=a1, 2, 3, 4, 5, 6, and a10).

Variables	Units (Source)	Description
agbd_t_pi_upper_aN	Mg/ha	Upper prediction interval in transform space (_aN=a1, 2, 3, 4, 5, 6, and a10).
agbd_t_se_aN	-	Model prediction standard error in fit units (_aN=a1, 2, 3, 4, 5, 6, and a10).
algorithm_run_flag_aN	-	Algorithm run flag-this algorithm is run if this flag is set to 1. This flag selects data that have sufficient waveform fidelity for AGBD estimation (_aN=a1, 2, 3, 4, 5, 6, and a10).
l2_quality_flag_aN	-	Flag identifying the most useful L2 data for biomass predictions (_aN=a1, 2, 3, 4, 5, 6, and a10).
l4_quality_flag_aN	-	Flag simplifying selection of most useful biomass predictions (_aN=a1, 2, 3, 4, 5, 6, and a10).
response_limit_flag_aN	-	Predictor value is outside the bounds of the training data (_aN=a1, 2, 3, 4, 5, 6, and a10).
shot_number	-	Unique identifier used to link observations between groups and between data products. The shot number format is OOOOBBRRGNNNNNNNN, where OOOOO is the orbit number, BB is the beam number, RR is reserved for future, G is the sub-orbit number, and NNNNNNNN is the shot number within the beam.
agbd_t	-	Model prediction in fit units.
agbd_t_se	-	Model prediction standard error in fit units (needed for calculation of custom prediction intervals).
algorithm_run_flag	-	The L4A algorithm is run if this flag is set to 1. This flag selects data that have sufficient waveform fidelity for AGBD estimation.
beam	-(L2A)	Beam identifier.
channel	-(L2A)	Channel identifier.
degrade_flag	-(L2A)	Flag indicating degraded state of pointing and/or positioning information.
delta_time	s (L2A)	Time delta since Jan 1 00:00 2018.
elev_lowestmode	m (L2A)	Elevation of center of lowest mode relative to a reference ellipsoid.

Table 6. Variables in the Ancillary Group: Model data. This group provides parameters and variables from the L4A models used to generate predictions. All the model parameters and uncertainty estimates (e.g. variance-covariance matrix of the model parameters) required as input to the GEDI04_B algorithm are also provided. Note: "-" indicates the variable is unitless and input variables from the GEDI02_A data product are marked with "L2A" as the source.

Variables	Units (Source)	Description
predict_stratum	-	Prediction stratum (e.g., DBT_Af=Deciduous Broadleaf Tree, Africa)
model_group	-	Model group (1=all predictors considered, 2=no RH metrics below RH50, 3=forced inclusion of RH98, 4=forced inclusion of RH98 and no RH metrics below RH50)
model_name	-	Model name (prediction stratum used for the fit data)
model_id	-	Model rank used for the prediction stratum
bias_correction_name	-	Back-transform bias correction method (Snowdon, Baskerville)
bias_correction_value	-	Back-transform bias correction value
dof	-	Degrees of freedom
fit_stratum	-	Fit stratum
par	-	Model parameters (coefficients)
npar	-	Number of model parameters (coefficients)
predictor_id	-	Predictor identifier
predictor_max_value	-	Maximum value of predictor in transform space used to train the model
response_max_value	Mg/ha	Maximum value of Mg/ha used to train the model
rh_index	-	Index of RH metric to use as a predictor

Variables	Units (Source)	Description
rse	-	Residual Standard Error
vcov	-	Variance-covariance matrix of model parameters
x_transform	-	Predictor transform (e.g., sqrt, log, none)
y_transform	-	Response transform (e.g., sqrt, log)

Table 7. Variables in the Ancillary/pft_lut group. This group provides look-up tables that link a numeric value from gridded PFTs to a descriptive text name. Note: “-” indicates the variable is unitless and input variables from the GEDI02_A data product are marked with “L2A” as the source.

Variables	Units (Source)	Description
pft_class	- (MCD12Q1)	MCD12Q1 Type 5 plant functional type (PFT) class
pft_name	-	L4A PFT strata

Table 8. Variables in the Ancillary/region_lut group. This group provides look-up tables that link a numeric value from the gridded world region to a descriptive text name. Note: “-” indicates the variable is unitless and input variables from the GEDI02_A data product are marked with “L2A” as the source.

Variables	Units (Source)	Description
region_class	-	L4A geographical region identifier
region_name	-	L4A geographical region strata

3. Application and Derivation

Most previous efforts have developed site-specific or regional relationships between AGBD and remote sensing measurements (Drake et al., 2002). In contrast, GEDI requires models and algorithms designed to perform well throughout the entire observation domain of the ISS. Locally developed or regional relationships between AGBD and height are unlikely to perform well at locations outside the limited geographic extent of training data unless procedures are developed specifically to ensure transferability beyond the extent of calibration measurements. The GEDI L4A algorithm and product currently address two important components of transferability: (1) geographic transferability, meaning that the models can be extrapolated to locations outside the geographic extent of training data; and (2) transferability from simulated to recorded GEDI waveforms.

4. Quality Assessment

The GEDI Forest Structure and Biomass Database (FSBD) contained 31,414 simulated GEDI waveforms co-located with field plot estimates of AGBD. After excluding projects that are not analysis-ready or otherwise inappropriate for GEDI (e.g., variable radius plots), the unfiltered GEDI04_A calibration dataset contained 12,140 simulated GEDI waveforms. Quality control filters designed to flag observations that are likely to be erroneous (e.g., incongruence between height and AGBD) or that do not meet the requirements of waveform simulator were then applied. The filtered GEDI04_A calibration dataset used to develop the first release of the GEDI04_A data product contained 8,587 simulated waveforms from 21 countries on all continents within the GEDI domain.

To quantify geographic transferability candidate models were evaluated within sets of 5-degree grid cells that contain simulated GEDI waveforms with coincident field data. The approach sets aside data from one grid cell for testing and trains the model using data within the remaining grid cells. This model was used to predict AGBD within the held-out grid cell, and the process was repeated for all grid cells within each stratum for all models under consideration.

Refer to Kellner et al. (2021) for further details on the uncertainty/calibration analysis applied.

5. Data Acquisition, Materials, and Methods

The GEDI instrument is aboard the International Space Station (ISS) and its mission aims to characterize ecosystem structure and dynamics to enable improved quantification and understanding of the Earth’s carbon cycle and biodiversity. GEDI is led by the University of Maryland in collaboration with NASA Goddard Space Flight Center. GEDI science data algorithms and products are created by the GEDI Science Team.

The GEDI instrument produces high-resolution laser ranging observations of the 3-dimensional structure of the Earth. GEDI was launched on December 5, 2018, and is attached to the ISS. GEDI collects data globally at the highest resolution and densest sampling of any light detection and ranging (lidar) instrument in orbit to date. The GEDI instrument consists of 3 lasers producing a total of 8 beam ground transects, which consist of ~25 m footprint samples spaced approximately every 60 m along-track. The GEDI beam transects are spaced approximately 600 m apart on the Earth’s surface in the cross-track direction, for an across-track width of ~4.2 km.

Footprint AGBD is derived from linear parametric models that relate GEDI L2A waveform relative height metrics to aboveground biomass estimates from co-located field plots. The GEDI approach to footprint model selection is data-driven. Candidate models are stratified by plant functional type (PFT) and continental region, with natural logarithm or square root transformations on the response and predictor variables. The GEDI footprint models represent the following combination of PFT’s—deciduous broadleaf trees, evergreen broadleaf trees, evergreen and deciduous needleleaf trees, and combinations of woodlands, grasslands, and shrubs.

GEDI footprint AGBD is an L4A data product (GEDI04_A). Models to produce GEDI04_A were developed using field estimates of AGBD colocated with simulated GEDI waveforms derived from discrete-return airborne lidar (Blair and Hofton, 1999; Hancock et al., 2019). The justification for using simulated GEDI waveforms is that few locations on the land surface are associated with field estimates of AGBD that could be used to train GEDI models. Because GEDI is a sampling mission and most field plots are small, GEDI data will not intersect most of these locations during the mission life. Simulated GEDI waveforms are processed to GEDI02_A equivalent RH metrics, which are defined as the percentage of the received laser waveform intensity that is less than a given height, where height is computed relative to the elevation of the lowest mode in the waveform (Fig. 2).

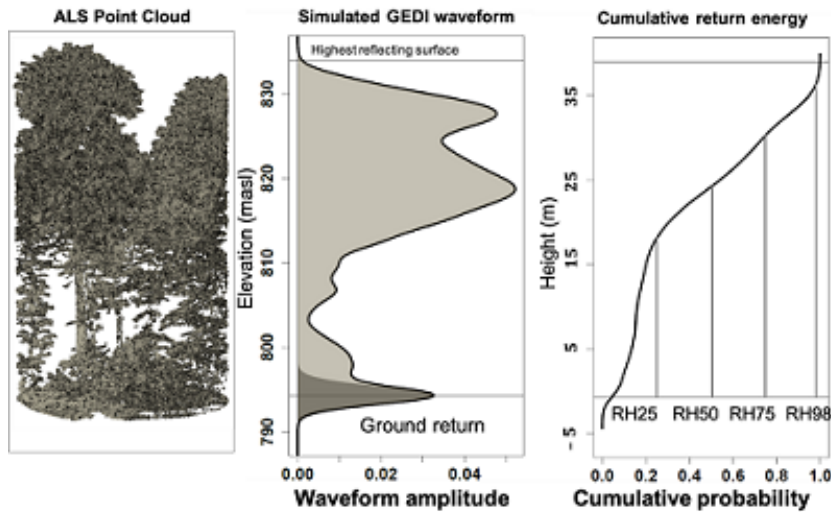


Figure 2. Relative height (RH) metrics were calculated as the height relative to ground elevation under which a certain percentage of waveform energy has been returned. RH50, for example, is the height relative to the ground elevation below which 50% of waveform energy has been returned.

The GEDI approach to developing footprint AGBD models considers multiple candidates stratified by world region and PFT with different functional forms. The models were developed using a quality-filtered calibration dataset that contains 8,587 simulated waveforms in 21 countries. These data were contributed by numerous researchers and standardized into the GEDI FSBD, which is a living data archive that grows over time as new datasets are assimilated and improvements are made to existing records.

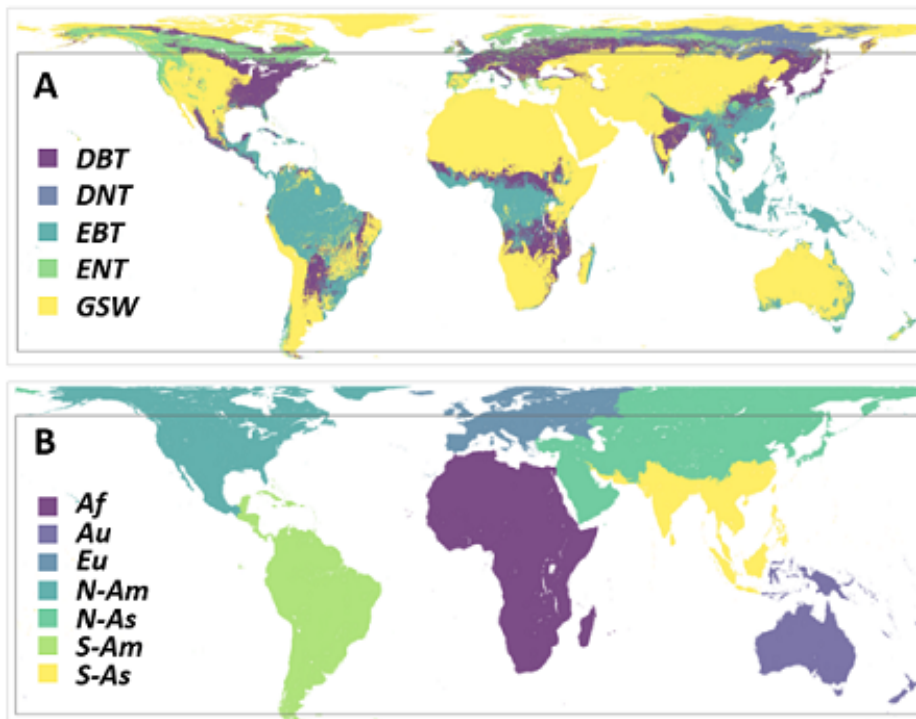


Figure 3. Global stratification by five combinations of error-corrected and infilled MODIS MCD12Q1 PFT (A) and world region (B) to produce GEDI footprint AGBD estimators. The box inset is the GEDI observation domain of 51.6 degrees N to S latitude. DBT (deciduous broadleaf trees), DNT (deciduous needleleaf trees), EBT (evergreen broadleaf trees), ENT (evergreen needleleaf trees), GSW (grasses, shrubs, and woodlands). Af (Africa), Au (Australia and Oceania), Eu (Europe), N-Am (North America north of southern Mexico), N-As (North Asia), S-Am (South America, Central America, southern Mexico, and the Caribbean), S-As (South Asia).

GEDI04_A world region includes the geologically defined continents of Africa and Europe. The South America world region is the continent of South America, Central America and the Caribbean islands, and geological North America south of southern Mexico. The Australia and Oceania world region is geological Australia and the island regions north of Australia on the east side of the Wallace line, which defines the floral and faunal boundary between Australia and Asia during the Pleistocene (Mayr, 1944). The islands of Micronesia, Melanesia, and Polynesia are associated with the

Australia and Oceania world region regardless of political affiliation. The North American world region includes geological North America north of southern Mexico. The continent of Asia was divided into north and south regions that approximately correspond to temperate and tropical forests. GEDI04_A PFTs are assigned using an error-corrected and infilled 1km grid derived from the Type 5 classification in the MODIS MCD12Q1 V006 data product (Friedl et al., 2010; 2002). These are deciduous broadleaf trees (DBT; class 4), deciduous needleleaf trees (DNT; class 3), evergreen broadleaf trees (EBT, class 2), evergreen needleleaf trees (ENT, class 1), and grasses, shrubs, and woodlands (GSW, classes 5 and 6).

On-orbit predictions of AGBD are made using the GEDI02_A elevation and height metric data product as input. The algorithms used by GEDI for generating these are described in the ATBD for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products (Hofton and Blair, 2020). The L4A product contains the information necessary to reproduce the AGBD prediction for individual GEDI shots from L2A data; the algorithm setting group selection used in Release 2 GEDI02_A data is applied to these data on a per footprint basis.

The GEDI04_A algorithm is described in detail in the ATBD for L4 GEDI aboveground biomass density (Kellner et al., 2021), and the development of GEDI04_A models will be described in forthcoming publications. The algorithm generates a predicted value of AGBD in units of megagrams per hectare for every valid GEDI02_A waveform. The algorithm uses the latitude and longitude of the lowest mode to lookup the PFT from a modified version of MCD12Q1 V006 PFT classification and a world region grid. It then gets the selected estimator for the given combination of PFT and world region and predicts AGBD after scaling and transforming GEDI02_A RH metrics. Prediction intervals and the standard error of the prediction are generated and written files.

Common Queries on How to Use and Interpret the GEDI04_A Data Product

This section is also provided in the companion file GEDI_L4A_Common_Queries.pdf.

How are the GEDI04_A biomass estimates geolocated?

The GEDI04_A product uses the ground position as the location of each shot and AGBD estimate (*elev_lowestmode*, *lat_lowestmode*, *lon_lowestmode*). Additional waveform ranging points are available in the GEDI02_A product (e.g., *elev_highestreturn*, *lat_highestreturn*, *lon_highestreturn*) and may be joined to GEDI04_A using the *shot_number* dataset.

Note that the Release 1 GEDI04_A product is derived from the Release 1 GEDI02_A product (PGEVersion 1), therefore has the same geolocation. It is not straightforward to link Release 1 and 2 data granules because the *shot_number* format changed, and the number of shots in a granule changed as a result of the switch to sub-orbit granules and removal of laser off periods in Release 2 (see the GEDI01_B and GEDI02_A User Guides for details). Release 1 (PGEVersion 1) and 2 (PGEVersion 3) GEDI02_A data product files are both available through LPDAAC.

What quality metrics and flags should I use to filter the data?

AGBD is predicted for every shot where it is possible to run the GEDI04_A algorithm, as indicated by the *algorithm_run_flag* dataset (see Table 1). The GEDI04_A product provides multiple quality flags and metrics that may be used to subset the predictions to the most useful observations for a particular application or region.

The *I2_quality_flag* encapsulates a number of GEDI02_A quality metrics to identify land surface shots with waveforms of high fidelity for AGBD estimation. The *I4_quality_flag* identifies shots that may be considered as samples of the population for which the applied models are representative. For example, GEDI04_A models for deciduous forests are only calibrated using GEDI waveforms simulated from leaf-on ALS data, therefore we can only apply the derived models to on-orbit GEDI waveforms acquired under similar conditions.

The *I2_quality_flag* uses a beam sensitivity threshold of 0.9 to match what is used for the Level 2 products. The *I4_quality_flag* uses a beam sensitivity threshold of 0.95, which was selected based on analysis of GEDI02_A and GEDI04_A on-orbit data. Beam sensitivity is an estimate of the maximum canopy cover that can be penetrated considering the signal-to-noise ratio of the waveform. For dense tropical forests, users may consider raising the beam sensitivity threshold (e.g., 0.98) to minimize measurement error in the RH metrics. In future releases, quality filtering will be improved by using the beam sensitivity together with the expected level of canopy cover for each shot.

Some users may wish to also evaluate the *predictor_limit_flag* and *response_limit_flag*. These identify shots with RH metrics or AGBD predictions, respectively, that are outside the observed range of values used to train the GEDI04_A models. Care should be taken when using such observations.

What are the units of *xvar*, and why doesn't *xvar* match the relative height metrics in a corresponding GEDI02_A file?

The variables called *xvar* in the BEAMXXXX group and *xvar_aN* in the BEAMXXXX/agbd_prediction group are the scaled and transformed RH metrics used to generate the AGBD prediction for a given estimator and prediction stratum. GEDI_04A estimators are linear statistical models with a square root or natural logarithm transformation on the response or predictor variables. The appropriate transformation for the given estimator has been applied to GEDI_02A RH metrics to generate *xvar*, and is indicated by the *x_transform* and *y_transform* variables in the ANCILLARY/model_data compound dataset. This transformation is applied after adding *predictor_offset* to the RH metrics. We add *predictor_offset* because RH metrics can be negative when a large percentage of waveform energy is within the ground return. Because the square root and natural logarithm of a negative number are undefined, adding a large positive constant is necessary. For example, if a given estimator used a square root transformation, *predictor_offset* = 100, and the true RH metric had a value of 20, the number in *xvar* would be:

$$xvar = \sqrt{(20+100)}$$

What is the relationship between *rh_index*, *predictor_id*, and *par* in the ANCILLARY/model_data compound dataset?

The vector *par* contains coefficients of the linear model used to predict AGBD, where the first element is the intercept and subsequent elements are slope coefficients. The vector *rh_index* is the height percentile associated with the given RH metric. The variable *predictor_id* provides a mapping between *rh_index* and *par*. For example, if *predictor_id* is:

$$predictor_id = [1, 2, 3, 0]$$

and *rh_index* is:

$$rh_index = [50, 98, 50, 70, 0]$$

the associated estimator (ignoring transformations) would be:

$AGBD = \text{par}[0] + \text{par}[1] \times RH50 + \text{par}[2] \times RH98 + \text{par}[3] \times RH50 \times RH70$

Note that when the same *predictor_id* is associated with two *rh_index* values, it indicates that the product of two RH metrics was used in the given linear model. Note also that *par*[0] is always the intercept term.

How can I derive prediction intervals at a different confidence level?

The GEDI L4A product provides the standard error of the prediction and the lower/upper prediction intervals for every estimate. The default confidence level used for these intervals is 90%; however, some users may wish to specify their own confidence level. The general formula of a prediction interval for a new observation is:

$\text{estimate} \pm (\text{standard error} \times t\text{-multiplier})$

where the *estimate* is the sample prediction in transform space (*agbd_t* in Table 1) and standard error is the standard error of the prediction in transform space (see *agbd_t_se* in Table 1). The *t-multiplier* ($t_{1-\alpha/2, \text{dof}}$) can be derived using standard libraries in R or Python and depends on: (1) the degrees-of-freedom for the applied model (*dof*), which is provided in the L4A product (*ANCILLARY/model_data/dof*); and (2) the t-distribution probability (α) which is specified by the BEAMXXXX/agbd_prediction group attribute *alpha* and may be modified by the user. For example, an *alpha* value of 0.1 is used for a 90% confidence level and 0.05 for a 95% confidence level. Note that the prediction intervals described above are in transform space and need to be back-transformed to place estimates in units of aboveground biomass density. A correction also needs to be applied to account for bias introduced by transformation of the response variable (*agbd*). For example, if *ANCILLARY/model_data/y_transform* is “*sqr*” and *ANCILLARY/model_data/bias_correction_name* is “*Snowdon*”, then

$\text{agbd} = \text{agbd_t}^2 \times \text{ANCILLARY/model_data/bias_correction_value}$

If *ANCILLARY/model_data/y_transform* is “*exp*” and *ANCILLARY/model_data/bias_correction_name* is “*Baskerville*”, then

$\text{agbd} = \exp(\text{agbd_t}) \times \exp(\text{ANCILLARY/model_data/bias_correction_value})$

6. References

- Blair, J.B., and M.A. Hofton. 1999. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters* 26:2509–2512. <https://doi.org/10.1029/1999GL010484>
- Drake, J.B., R.O. Dubayah, R.G. Knox, D.B. Clark, and J.B. Blair. 2002. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment* 81:378–392. [https://doi.org/10.1016/S0034-4257\(02\)00013-5](https://doi.org/10.1016/S0034-4257(02)00013-5)
- Dubayah, R., J.B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, G. Hurtt, J. Kellner, S. Luthcke, J. Armston, H. Tang, L. Duncanson, S. Hancock, P. Jantz, S. Marselis, P.L. Patterson, W. Qi, and C. Silva. 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing* 1:100002. <https://doi.org/10.1016/j.srs.2020.100002>
- Friedl, M.A., D.K. McIver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf. 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment* 83:287–302. [https://doi.org/10.1016/S0034-4257\(02\)00078-0](https://doi.org/10.1016/S0034-4257(02)00078-0)
- Friedl, M.A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* 114:168–182. <https://doi.org/10.1016/j.rse.2009.08.016>
- Hancock, S., J. Armston, M. Hofton, X. Sun, H. Tang, L.I. Duncanson, J.R. Kellner, and R. Dubayah. 2019. The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. *Earth and Space Science* 6:294–310. <https://doi.org/10.1029/2018EA000506>
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R. and Kommareddy, A., 2013. High-resolution global maps of 21st-century forest cover change. *science*, 342(6160), pp.850-853. <https://doi.org/10.1126/science.1244693>
- Hofton, M.A., and J.B. Blair. 2020. Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products. Goddard Space Flight Center, Greenbelt, MD. https://doi.org/10.5067/DOC/GEDI/GEDI_WF_ATBD.001
- Kellner, J.R., J. Armston, and L. Duncanson. 2021. Algorithm Theoretical Basis Document for GEDI footprint aboveground biomass density. See the companion file *GEDI_ATBD_L4A_v1.0.pdf*.
- Mayr, E., 1944. Wallace's Line in the Light of Recent Zoogeographic Studies. *The Quarterly Review of Biology* 19:1–14. <https://www.jstor.org/stable/280856>

7. Dataset Revisions

Version	Release Date	Description
2.0	2021-12-15	In this release, the algorithm setting group selection was modified for Evergreen Broadleaf Trees in South America to reduce false-positive errors resulting from the selection of waveform modes aboveground elevation as the lowest mode. The region class has been updated to correct boundary errors and to include the islands of Micronesia and Polynesia in the Australia and Oceania region class. Also, the granules are in suborbits. In Version 1, one orbit was contained in one file. It is not straightforward to link Version 1 and Version 2 data granules because the variable <i>shot_number</i> format changed, and the number of shots in a granule changed as a result of the switch to sub-orbit granules and removal of the laser off periods in Version 2.

Version	Release Date	Description
1.1	2022-02-15	This dataset consists only of the "golden weeks" (i.e., mission weeks 19, 32, 34, and 38) data from the GEDI L4A Version 1 dataset.
1.0	2019-09-09	Initial release of the GEDI L4A data. Superseded and available only upon request.

