

GEDI L4C Footprint Level Waveform Structural Complexity Index, Version 3

Documentation Revision Date: 2026-09-02

Dataset Version: 3

Summary

This dataset contains Global Ecosystem Dynamics Investigation (GEDI) Level 4C (L4C) Version 3 predictions of the Waveform Structural Complexity Index (WSCl) and estimates of prediction intervals for each footprint estimate at 95% confidence. In this version, the granules are in sub-orbits. The algorithm setting group selection used for GEDI04_C is the same as in the GEDI02_A product. 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 period 2019-04-17 to present. 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 WSCl was derived from XGBoost regression models relating simulated GEDI Level 2A (L2A) waveform relative height (RH) metrics to a 3D structural complexity metric calculated from matched Airborne laser Scanning (ALS) point clouds. Four global WSCl models were trained on a plant functional type (PFT) basis (i.e., deciduous broadleaf trees, evergreen broadleaf trees, evergreen needleleaf trees, and the combination of grasslands, shrubs, and woodlands). For each of the eight beams, additional data are reported with the WSCl estimates, including the associated uncertainty metrics, quality flags, and other information about the GEDI L2A waveform for this selected algorithm setting group. Additional model outputs include WSCl predictions for each of the four GEDI L2A algorithm setting groups and associated prediction intervals. Providing these ancillary data products will allow users to evaluate and select alternative algorithm setting groups. The data are provided in HDF5 format.

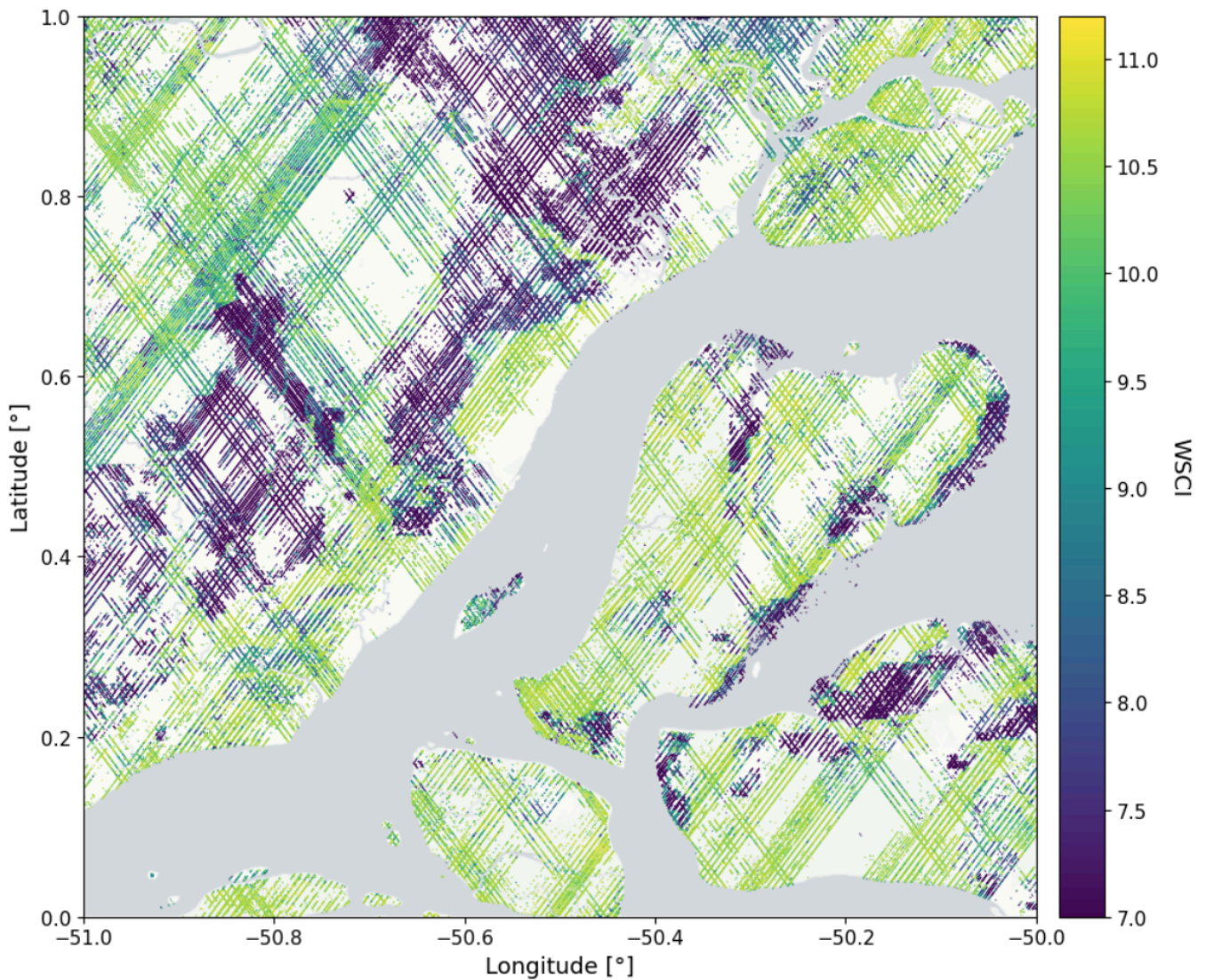


Figure 1. Example subset of the Waveform Structural Complexity Index (WSCl) predictions from the GEDI Level-4C footprint product over the eastern Amazon.

Citation

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 GEDI Level 4C (L4C) predictions of the WSCI and estimates of prediction intervals at 95% confidence within each sampled geolocated laser footprint. The footprints are located within the global latitude band observed by the ISS, nominally 51.6 degrees N and S and reported for the period 2019-04-17 to present. 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 WSCI estimates were derived from machine learning models that relate GEDI Level 2A (L2A) relative height (RH) metrics to measures of 3D structural complexity from GEDI intersected Airborne Laser Scanning (ALS) plots. The training dataset consisted of more than 800,000 collocated GEDI/ALS footprints, used to train four global WSCI models representing different PFTs (i.e., deciduous broadleaf trees, evergreen broadleaf trees, evergreen needleleaf trees and the combination of grasslands, shrubs, and woodlands).

Reported with the WSCI estimates for each of the eight beams are the associated uncertainty metrics, quality flags and other information about the GEDI L2A waveform for this selected algorithm setting group. Also provided are footprint geolocation variables and land cover input data including PFTs and the land cover type identifiers. Additional model outputs include the WSCI predictions for each of the four GEDI L2A algorithm setting groups with associated prediction intervals for each GEDI L2A algorithm setting group. Providing these ancillary data products will allow users to evaluate and select alternative algorithm setting groups.

The Version 3 GEDI04_C product is derived from the Version 3 GEDI02_A product. The WSCI models and methodology are unchanged from Version 2; the changes in this version are inherited from the GEDI L2A Version 3 update and from an updated quality-flag recipe that is consistent across the GEDI Level-2 and Level-4 footprint products. In particular, footprint predictions and ancillary outputs are provided for the four GEDI L2A algorithm setting groups retained in L2A Version 3 (setting groups 1, 2, 5, and 10; groups 3, 4, and 6 were removed), and a new quality flag (l4c_quality_flag_rel3) applies the Version 3 quality recipe while the Version 2 recipe is retained as l4c_quality_flag_rel2.

Project: Global Ecosystem Dynamics Investigation (GEDI)

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>

de Conto, T., J. Armston, and R. Dubayah. 2024. Characterizing the structural complexity of the Earth's forests with spaceborne lidar. *Nature Communications* 15:8116. <https://doi.org/10.1038/s41467-024-52468-2>

Related Datasets

De Conto, T., J. Armston, and R.O. Dubayah. 2025. GEDI L4C Gridded Waveform Structural Complexity Index, Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2470>

De Conto, T., J. Armston, and R.O. Dubayah. 2026. GEDI L4C Global Waveform Structural Complexity Index (WSCI) Fusion Product, Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2474>

Level 1 and Level 2 data from GEDI are available from the Land Processes Distributed Active Archive Center at <https://www.earthdata.nasa.gov/data/projects/gedi>.

Acknowledgements

This work was funded with a NASA Earth Ventures Instrument (EVI) supporting the GEDI mission (contract NNL15AA03C), and FINESST (grant NNH21ZDA001N).

2. Data Characteristics

Spatial Resolution: ~25 m footprints in diameter

Temporal Coverage: 2019-04-17 to present

Temporal Resolution: One-time estimates

Study Area: Latitude and longitude are given in decimal degrees

Site	Westernmost Longitude	Easternmost Longitude	Northernmost Longitude	Southernmost Latitude
Global	-180	180	56	-53

Data File Information

The data files are in HDF5 format (.h5). There is one companion file, **GEDI_L4C_WSCI_Data_Dictionary_P001_V3.pdf**, which provides a data dictionary for the variables in the data files. This data dictionary is also included in this guide document.

Data file naming convention:

The files are named *GEDI04_C_YYYYDDHHMMSS_O[orbit_number]_[granule_number]_T[track_number]_[PPDS_type]_[release_number]_[production_version]_V[version_number].h5* (e.g., GEDI04_C_2019349001415_O05699_04_T03266_02_003_01_V003.h5), where:

- *GEDI04_C* = product short name representing GEDI Level 4C data,
- *YYYYDDHHMMSS* = date and time of acquisition in Julian day of year, hours, minutes, and seconds format,
- *[orbit_number]* = orbit number,
- *[granule_number]* = sub-orbit granule (or file) number,
- *[track_number]* = track number,
- *[PPDS_type]* = positioning and pointing determination system (PPDS) type (00 is "predict", 01 is "rapid", 02 and higher is "final"),
- *[release_number]* = release number, representing the SOC SDS (software) release used to generate this L4C dataset,
- *[production_version]* = granule production version, e.g., a particular data granule (or file) may have been regenerated multiple times, and
- *[version_number]* = L4C dataset production version (003), corresponding to the ORNL DAAC's dataset version number.

Table 1. Data Dictionary

Dimension Variable	Description
MT	Number of Shots

Group: /		
short_name	(Attribute)	GEDI04_C
Group: /METADATA/DatasetIdentification		
abstract	(Attribute)	The GEDI04_C standard data product contains predictions of the waveform structural complexity index (WSCI) within each laser footprint
characterSet	(Attribute)	utf8
creationDate	(Attribute)	File creation date
credit	(Attribute)	The software that generates the WSCI product was implemented at the Department of Geographical Sciences at the University of Maryland (UMD)
fileName	(Attribute)	Original file name
language	(Attribute)	eng
originatorOrganizationName	(Attribute)	GEDI Science Data Processing System and University of Maryland
PGEVersion	(Attribute)	Product generating executive SDPS release ID
purpose	(Attribute)	The purpose of the WSCI dataset is to provide an estimate of forest structural complexity, similarly to what is measured by airborne LiDAR in the 3D space, using GEDI waveforms
shortName	(Attribute)	GEDI04_C
spatialRepresentationType	(Attribute)	along-track
status	(Attribute)	onGoing
topicCategory	(Attribute)	geoscientificInformation
uuid	(Attribute)	Universally unique identifier (UUID) for this file
VersionID	(Attribute)	SDPS DAAC release ID

gedi_wsci_githash	(Attribute)	Git commit hash of the software used to create the WSCI file		
Group: /BEAMXXXX				
Label	Datatype (Dimensions)	long_name	Units source	Description
algorithm_run_flag	UINT8 MT	Algorithm run flag	-	The WSCI algorithm is run if this flag is set to 1. This flag selects data which have sufficient waveform fidelity for WSCI estimation.
beam	UINT16 MT	Beam	- L2A	Beam identifier
channel	UINT8 MT	Channel	- L2A	Channel identifier
degrade_include_flag	UINT8 MT	Degrade inclusion flag	- L2A	Flag indicating acceptable pointing and positioning quality for inclusion. 1 = include (attitude_degrade_flag in [0,1,2,3,4,6] and trajectory_degrade_flag in [0,3,8]). Should be used together with elev_highestreturn_outlier_flag.
delta_time	FLOAT64 MT	Delta time	s L2A	Time delta since Jan 1 00:00 2018.
elev_lowestmode	FLOAT32 MT	Elevation of the lowest mode	m L2A	Elevation of center of lowest mode relative to reference ellipsoid
elev_highestreturn_outlier_flag	UINT8 MT	Elevation highest return outlier flag	- L4B	Flag indicating shot is from a section of an orbit with L2A canopy elevation metrics classified as large outliers relative to reference data. Should be used together with degrade_include_flag.
fhd_normal	FLOAT64 MT	Foliage Height Diversity	m L2B	Foliage height diversity index calculated by vertical foliage profile normalized by total plant area index
l2_algrunflag	UINT8 MT	L2 algorithm run flag	- L2A	Flag indicating the algorithm ran successfully for the selected algorithm. Copied from L2A.
l2a_quality_flag_rel3	UINT8 MT	L2A quality flag (release 3)	- L2A	Pre-computed L2A v3 quality flag for the selected algorithm. Copied from L2A. Set to 0 when not available in the source L2A file.
lat_lowestmode	FLOAT64 MT	Latitude of the lowest mode	degrees L2A	Latitude of center of lowest mode
lon_lowestmode	FLOAT64 MT	Longitude of the lowest mode	degrees L2A	Longitude of center of lowest mode
master_frac	FLOAT64 MT	Fraction component of shot time	s L2A	Master time, fractional part. master_int+master_frac is equivalent to /BEAMXXXX/delta_time.
master_int	UINT32 MT	Integer component of shot time	s L2A	Master time, integer part. Seconds since master_time_epoch. master_int+master_frac is equivalent to /BEAMXXXX/delta_time
selected_algorithm	UINT8 MT	Selected algorithm setting group	- L2A	Selected algorithm setting group
sensitivity	FLOAT32 MT	Beam sensitivity	- L2A	Maximum canopy cover that can be penetrated considering the SNR of the waveform
shot_number	UINT64 MT	Shot number	- L2A	Shot number
solar_elevation	FLOAT32 MT	Solar elevation	degrees L2A	Solar elevation angle
surface_flag	UINT8 MT	Surface flag	- L2A	Indicates elev_lowestmode is within 300m of Digital Elevation Model (DEM) or Mean Sea Surface (MSS) elevation
wsci	FLOAT32 MT	Waveform Structural Complexity Index	-	Predicted 3D canopy entropy from the corresponding Plant Functional Type (PFT) model

wsci_pi_lower	FLOAT32 MT	Waveform Structural Complexity Index lower prediction interval	-	Lower prediction interval at 95% confidence
wsci_pi_upper	FLOAT32 MT	Waveform Structural Complexity Index upper prediction interval	-	Upper prediction interval at 95% confidence
l4c_quality_flag_rel2	UINT8 MT	L4C quality flag (release 2 recipe)	-	Flag simplifying selection of most useful L4C predictions using the v2 quality recipe. Identical to wsci_quality_flag in the GEDI04_C v2 release.
l4c_quality_flag_rel3	UINT8 MT	L4C quality flag (release 3 recipe)	-	Flag simplifying selection of most useful L4C predictions using the v3 quality recipe: l2a_quality_flag_rel3 == 1, urban_proportion <50, worldcover_class not in [0, 80], and rh100 <150. Set to 0 when the L2A v3 quality flag is not available in the source L2A file.
wsci_xy	FLOAT32 MT	Horizontal Structural Complexity	-	Predicted WSCI horizontal term over the XY plane within the footprint
wsci_xy_pi_lower	FLOAT32 MT	Horizontal Structural Complexity lower prediction interval	-	Lower prediction interval at 95% confidence
wsci_xy_pi_upper	FLOAT32 MT	Horizontal Structural Complexity upper prediction interval	-	Upper prediction interval at 95% confidence
wsci_z	FLOAT32 MT	Vertical Structural Complexity	-	Predicted WSCI vertical term along the Z axis within the footprint
wsci_z_pi_lower	FLOAT32 MT	Vertical Structural Complexity lower prediction interval	-	Lower prediction interval at 95% confidence
wsci_z_pi_upper	FLOAT32 MT	Vertical Structural Complexity upper prediction interval	-	Upper prediction interval at 95% confidence

Group: /BEAMXXXX/geolocation

Label	Datatype (Dimensions)	long_name	Units source	Description
elev_lowestmode_aN	FLOAT32 MT	Elevation of the lowest mode	m L2A	Elevation of center of lowest mode relative to reference ellipsoid
lat_lowestmode_aN	FLOAT64 MT	Latitude of the lowest mode	degrees L2A	Latitude of center of lowest mode
lon_lowestmode_aN	FLOAT64 MT	Longitude of the lowest mode	degrees L2A	Longitude of center of lowest mode
sensitivity_aN	FLOAT32 MT	Beam sensitivity	- L2A	Maximum canopy cover that can be penetrated considering the SNR of the waveform
shot_number	UINT64 MT	Shot number	- L2A	Shot number
stale_return_flag	UINT8 MT	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. The pulse location of the previous shot was used to select the telemetered waveform.

Group: /BEAMXXXX/land_cover_data

Label	Datatype (Dimensions)	long_name	Units source	Description
-------	--------------------------	-----------	-----------------	-------------

landsat_treecover	FLOAT64 MT	Landsat tree canopy cover	percent L2A	Tree cover in the year 2010, defined as canopy closure for all vegetation taller than 5m in height (Hansen et al., 2013). Encoded as a percentage per output grid cell.
landsat_water_persistence	UINT8 MT	25 m Landsat water persistence	percent L2A	The percent UMD GLAD Landsat observations with classified surface water between 2018 and 2019. Values > 80 usually represent permanent water while values < 10 represent permanent land.
leaf_off_flag	UINT8 MT	1 km VIIRS leaf-off flag	- L4A	GEDI 1 km EASE 2.0 grid flag indicating if the observation was recorded during leaf-off conditions in deciduous needleleaf or broadleaf forests and woodlands. 1 = leaf-off and 0 = leaf-on.
phenology_phase	UINT8 MT	VIIRS phenological phase	- L2A	VIIRS phenological phase of GEDI shot (1 = onset greenness increase, 2 = mid-greenup, 3 = onset greenness maximum, 4 = onset greenness decrease, 5 = mid-senescence, 6 = onset greenness minimum, 255 = null)
phenology_year	UINT16 MT	VIIRS source year	- L2A	VIIRS product year used to determine phenological phase
pft_class	UINT8	1 km MODIS Plant Functional Type class	- L2A	GEDI 1 km EASE 2.0 grid Plant Functional Type (PFT) derived from the MODIS MCD12Q1v006 Product. Values follow the Land Cover Type 5 Classification scheme.
region_class	UINT8 MT	1 km geographic region class	- L2A	GEDI 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	UINT64 MT	Shot number	- L2A	Shot number
urban_focal_window_size	UINT8 MT	Urban focal window size	pixels L4A	The focal window size used to calculate urban_proportion. Values are 3 (3x3 pixel window size) or 5 (5x5 pixel window size).
urban_proportion	UINT8 MT	25 m TDX urban percentage	percent L4A	The percentage proportion of land area within a focal area surrounding each shot that is urban land cover. Urban land cover is derived from the DLR 12 m resolution TanDEM-X Global Urban Footprint Product.
worldcover_class	UINT8 MT	ESA WorldCover 2021 v200	- L2A	Land cover class from the European Space Agency WorldCover 10 m 2021 v200 product (https://doi.org/10.5281/zenodo.7254221), copied from the source GEDI02_A granule. A value of 0 indicates no data (e.g. ocean, outside the land-only product's coverage).

Group: /BEAMXXXX/wsci_prediction

Label	Datatype (Dimensions)	long_name	Units source	Description
algorithm_run_flag_aN	UINT8 MT	Algorithm run flag	-	The algorithm is run if this flag is set to 1. This flag selects data which have sufficient waveform fidelity for WSCI estimation.
l2_algrunflag_aN	UINT8 MT	L2 algorithm run flag	- L2A	Per-algorithm flag indicating the algorithm ran successfully. Copied from L2A geolocation/l2_algrunflag_aN.
l2a_quality_flag_rel3_aN	UINT8 MT	L2A quality flag per algorithm (release 3)	- L2A	Pre-computed L2A v3 quality flag per algorithm. Copied from L2A geolocation/l2a_quality_flag_rel3_aN. Set to 0 when not available.
shot_number	UINT64 MT	Shot number	- L2A	Shot number
wsci_aN	FLOAT32 MT	Waveform Structural Complexity Index	-	Waveform Structural Complexity Index
wsci_pi_lower_aN	FLOAT32 MT	Waveform Structural Complexity Index lower prediction interval	-	Lower prediction interval at 95% confidence
wsci_pi_upper_aN	FLOAT32 MT	Waveform Structural	-	Upper prediction interval at 95% confidence

		Complexity Index upper prediction interval		
l4c_quality_flag_rel2_aN	UINT8 MT	L4C quality flag per algorithm (release 2 recipe)	-	Per-algorithm flag simplifying selection of most useful L4C predictions using the v2 quality recipe. Identical to wsci_quality_flag_aN in the GEDI04_C v2 release.
l4c_quality_flag_rel3_aN	UINT8 MT	L4C quality flag per algorithm (release 3 recipe)	-	Per-algorithm flag simplifying selection of most useful L4C predictions using the v3 quality recipe: l2a_quality_flag_rel3_aN == 1, urban_proportion <50, worldcover_class not in [0, 80], and rh100 <150. Set to 0 when the L2A v3 quality flag is not available.
wsci_xy_aN	FLOAT32 MT	Horizontal Structural Complexity	-	Horizontal Structural Complexity
wsci_xy_pi_lower_aN	FLOAT32 MT	Horizontal Structural Complexity lower prediction interval	-	Lower prediction interval at 95% confidence
wsci_xy_pi_upper_aN	FLOAT32 MT	Horizontal Structural Complexity upper prediction interval	-	Upper prediction interval at 95% confidence
wsci_z_aN	FLOAT32 MT	Vertical Structural Complexity	-	Vertical Structural Complexity
wsci_z_pi_lower_aN	FLOAT32 MT	Vertical Structural Complexity lower prediction interval	-	Lower prediction interval at 95% confidence
wsci_z_pi_upper_aN	FLOAT32 MT	Vertical Structural Complexity upper prediction interval	-	Upper prediction interval at 95% confidence

3. Application and Derivation

The WSCI is the first product aiming at describing 3D structural complexity consistently on a near-global scale. Recent studies mapping forest structural complexity from lidar were limited to sites where airborne or terrestrial laser scanning point clouds were available. To expand estimates of 3D complexity to a global scale, relationships between GEDI and over 800,000 collocated ALS point cloud samples were developed across the globe where 3D complexity was measured. Those paired GEDI/ALS samples were used to develop four global models to predict structural complexity on a PFT basis, using all GEDI RH percentiles from each footprint to predict 3D canopy entropy (CE_{xyz}), a surrogate for structural complexity measured from ALS point clouds, using XGBoost regression. The modeling framework relied on a rigorous grid search spatial cross-validation regime to ensure (1) optimal hyperparameter tuning, (2) model generalization (i.e. robustness against overfitting) and (3) geographical transferability of all models.

4. Quality Assessment

The analysis used 816,276 collocated GEDI/ALS samples from 229 ALS surveys distributed around the world. Those sites were split into training and calibration sets, containing the samples from 80% and 20% of the sites, respectively. The training set was used to fit the WSCI models through XGBoost regression, applying grid search hyperparameter tuning with 5-fold spatial cross-validation. The best model was chosen based on the optimal combination of hyperparameters that minimized the root mean squared error (RMSE) loss function in the validation folds, while maintaining a difference of less than 5% between training and validation folds in terms of RMSE and R^2 , thus enforcing model generalization. Since each fold contained data from different geographical locations, the validation folds measured the model performance on sites unseen during model training, granting geographical transferability of the selected models. The calibration set was used to train conformal predictors (Vovk et al., 2005) for each WSCI model, enabling uncertainty estimation for every model prediction based on paired GEDI/ALS samples never seen during training of the XGBoost models. Uncertainty estimates are provided as prediction intervals at 95% confidence for every footprint.

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.

The GEDI WSCI models were trained on a database of 816,276 collocated GEDI/ALS samples over from all major continental regions, with a fraction of 24%, 32%, 24% and 20% of those footprints registered in deciduous broadleaf trees (DBT), evergreen broadleaf trees (EBT), evergreen needleleaf trees (ENT), and combinations of woodlands, grasslands, and shrubs (GSW), respectively. The ALS data used in the model development are part of the GEDI Calibration and Validation (Cal/Val) database and were provided by research partners or downloaded from open data initiatives, such as the NEON (NEON, 2023) in the United States, the INPE-EBA (Ometto et al., 2023) in Brazil. With more ALS data sets becoming available allied to the GEDI mission extension, expected to stay operational until the end of the life cycle of the International Space Station, we plan to expand the WSCI training database and keep updating its underlying models.

Prior to training the models, the GEDI footprints were collocated to ALS point clouds to correct for geolocation error in the Cal/Val processing, ensuring that both datasets were matched as precisely as possible and represented the exact same plot over the land surface. 3D Canopy Entropy (CExyz) measurements were then extracted from 25-m diameter ALS point clouds, matching the GEDI footprints in position and area. The CExyz metric was developed by Liu et al. (2022), and leverages entropy planes on the horizontal and vertical directions of a forest plot to account for 3D structural complexity of the forest canopy, capturing variations in features related to tree density and vertical layering simultaneously.

Footprint WSCI is derived from extreme gradient boosted trees (XGBoost) regression models (Chen and Guestrin, 2016) relating GEDI L2A relative height metrics to CExyz measured from ALS point clouds. The XGBoost models were trained on a PFT basis, assigned to each GEDI footprint in the L2A product and extracted from an error-corrected and infilled 1-km grid derived from the Type 5 classification in the MODIS MCD12Q1 V061 data product (Friedl and Sulla-Menashe, 2022). WSCI estimates are derived from four different models trained on footprints registered over DBT, EBT, ENT and GSW

All available GEDI RH metrics were used as input in the regression models, allowing the XGBoost algorithm to choose the best predictors of CExyz in a data driven fashion. RH percentiles within close ranges from each other are highly correlated, and although the XGBoost is an efficient machine learning method robust to overfitting, a rigorous regime of hyperparameter tuning and spatial cross-validation was applied to guarantee model generalization and geographical transferability.

On-orbit predictions of WSCI 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 L4C product contains estimates from all algorithm group settings available in L2A, which apply slightly different methods to calculate the RH metrics, thus affecting WSCI estimates, accompanied by prediction intervals and quality flags for every footprint.

How are the GEDI04_C estimates geolocated?

The GEDI04_C product uses the ground position as the location of each shot and WSCI 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_C using the shot_number dataset.

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

WSCI was predicted for every shot where the GEDI02_A algorithm was successfully run and RH metrics are provided, as indicated by the algorithm_run_flag dataset. The GEDI04_C 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 l2_quality_flag_rel3, inherited from the GEDI02_A Version 3 product, encapsulates a number of GEDI02_A quality metrics - including beam sensitivity thresholds (higher over tropical evergreen forests) - to identify land surface shots with waveforms of high fidelity for WSCI estimation. The l4c_quality_flag_rel3 applies more strict filters on top of l2_quality_flag_rel3 to identify shots with a higher likelihood of occurring over vegetated land surfaces, matching the domain for which the WSCI models were developed. More specifically, the l4c_quality_flag_rel3 applies additional filters for urban surfaces (urban_proportion) and excludes open water and other non-land surfaces using the ESA worldcover 10m 2021 v200 product (Zanaga et al., 2022), provided as the worldcover_class dataset in the land_cover_data group of each laser beam in the L4C product. For continuity with the previous version, the Version 2 quality recipe is also retained as l4c_quality_flag_rel2 (formerly named wsci_quality_flag).

What are the units and range for the WSCI estimates?

The WSCI estimates are equivalent to CExyz measured on ALS point clouds of 25-m diameter, matching the GEDI's footprint size. The WSCI estimates 3D entropy, which does not have a physically meaningful unit, but rather ranges from low to high values. Globally, WSCI ranges approximately from 7 to 11.4 on forest landscapes, with WSCI <8 corresponding to barely vegetated areas, where little to no forest structural complexity is observed. On the other extreme, WSCI >10.5 represents high structural complexity, such as tall, dense, undisturbed forests in the tropics. Again, these values are given on a global context, and the WSCI ranges and their interpretations may change on local contexts or when comparing forest types of similar structure (e.g. within a given canopy height interval). Therefore, which WSCI values represent low or high structural complexity depends on the local context of each study and should be assessed accordingly.

6. References

- Chen, T., and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. <https://doi.org/10.1145/2939672.2939785>
- de Conto, T., J. Armston, and R. Dubayah. 2024. Characterizing the structural complexity of the Earth's forests with spaceborne lidar. Nature Communications 15:8116. <https://doi.org/10.1038/s41467-024-52468-2>
- De Conto, T., J. Armston, and R.O. Dubayah. 2025. GEDI L4C Gridded Waveform Structural Complexity Index, Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2470>
- De Conto, T., J. Armston, and R.O. Dubayah. 2026. GEDI L4C Global Waveform Structural Complexity Index (WSCI) Fusion Product, Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2474>
- 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.,and D. Sulla-Menashe. 2022. MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061. NASA Land Processes Distributed Active Archive Center. <https://doi.org/10.5067/MODIS/MCD12Q1.061>

Hansen, M.C., P.V. Potapov, R. Moore, M. Hancher, S.A. Turubanova, A. Tyukavina, D. Thau, S.V. Stehman, S.J. Goetz, T.R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C.O. Justice, and J.R.G. Townshend. 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342:850-853. <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

Liu, X., Q. Ma, X. Wu, T. Hu, Z. Liu, L. Liu, Q. Guo, and Y. Su. 2022. A novel entropy-based method to quantify forest canopy structural complexity from multiplatform lidar point clouds. *Remote Sensing of Environment* 282:113280. <https://doi.org/10.1016/j.rse.2022.113280>

NEON (National Ecological Observatory Network). 2023. Discrete return LiDAR point cloud. NEON (National Ecological Observatory Network). <https://doi.org/10.48443/xxby-5a18>

Ometto, J., E.B. Gorgens, F.R. de Sousa Pereira, L. Sato, M.L.R. Assis, R. Cantinho, M. Longo, A.D. Jacon, and M. Keller. 2023. L1A - Discrete airborne LiDAR transects collected by EBA in the Brazilian Amazon (Roraima e Amapá) (Version v20230301). Zenodo. <https://doi.org/10.5281/zenodo.7689693>

Vovk, V., A. Gammerman, and G. Shafer (eds.). 2005. Conformal prediction. In *Algorithmic Learning in a Random World* (pp. 17–51). Springer US. https://doi.org/10.1007/0-387-25061-1_2

Zanaga, D., R. Van De Kerchove, D. Daems, W. De Keersmaecker, C. Brockmann, G. Kirches, J. Wevers, O. Cartus, M. Santoro, S. Fritz, M. Lesiv, M. Herold, N.-E Tsendbazar, P. Xu, F. Ramoino, and O. Arino. 2022. ESA WorldCover 10 m 2021 v200 (Version v200). Zenodo. <https://doi.org/10.5281/ZENODO.7254221>

7. Dataset Revisions

Version	Release Date	Description
3	2026-08-31	The Version 3 GEDI04_C product is derived from the Version 3 GEDI02_A product. The WSCI models and methodology are unchanged from Version 2; the changes in this version are inherited from the GEDI L2A Version 3 update and from an updated quality-flag recipe that is consistent across the GEDI Level-2 and Level-4 footprint products. In particular, footprint predictions and ancillary outputs are provided for the four GEDI L2A algorithm setting groups retained in L2A Version 3 (setting groups 1, 2, 5, and 10; groups 3, 4, and 6 were removed), and a new quality flag (l4c_quality_flag_rel3) applies the Version 3 quality recipe while the Version 2 recipe is retained as l4c_quality_flag_rel2.
2.1	2026-05-20	This release adds new files for mission weeks 328 through 343 (2025-03-20 through 2025-07-09).
2.1	2025-09-23	This release adds new files for mission weeks 281 through 327 (2024-04-26 through 2025-03-19). No acquisitions occurred while the GEDI instrument was in storage on the International Space Station (ISS) from March 2023 to April 2024.
2.1	2025-01-27	In this update, 23 granules from 6 orbits were renamed to reflect corrected reference ground track (RGT) numbers. No science data was affected. Specifically, granules with track number 01461 (T01461) from orbit 09360 (O09360) were renamed to track number 01614 (T01614). For example, the granule GEDI04_C_2020220033908_O09360_02_T01461_02_001_01_V002.h5 was renamed to GEDI04_C_2020220033908_O09360_02_T01614_02_001_01_V002.h5. Additionally, the following tracks were renamed: track T11164 from orbit O16529 was renamed to T11011, track T08166 from orbit O16530 was renamed to T09436, track T06592 from orbit O16532 was renamed to T10861, track T08016 from orbit O16533 was renamed to T09286, and track T10863 from orbit O16534 was renamed to T06441.