Algorithm Theoretical Basis Document (ATBD)

for

GEDI L2B Footprint Canopy Cover and Vertical Profile Metrics

Hao Tang[1], John Armston[1]

1. University of Maryland, College Park, MD

Version 1.0
Release date: December, 6th, 2019
Goddard Space Flight Center, Greenbelt, MD

Authors:

Principal Investigator:

Abstract

Accurate measurements of vertical forest structure at a global scale are critically important to advance our knowledge of terrestrial ecology and biodiversity. NASA's Global Ecosystem Dynamics Investigation (GEDI) mission aims to fill current observation gaps by collecting the first high-resolution lidar observations of the 3D structure of the Earth and providing spatially dense samplings of forest structure between ~52° N and ~52° S. The GEDI instrument consists of 3 lasers producing a total of 8 beam ground transects that are spaced approximately 600 m apart on the Earth's surface in the cross-track direction. Each beam transect consists of ~25 m footprint samples approximately spaced every 60 m along track. The fundamental footprint observations made by the GEDI instrument are received waveforms of energy as a function of receive time. These are combined with laser pointing and positioning information for precise geolocation and post-processed to determine ranging points of reflecting surfaces with the waveform footprint. The waveforms provided in the L1B product and locations of reflecting surfaces within the footprint provided in the L2A product are then used to derive the directional gap probability profile and extract biophysical metrics from each GEDI waveform. These metrics include canopy cover, Plant Area Index (PAI), Plant Area Volume Density (PAVD) and Foliage Height Diversity (FHD). This ATBD presents the algorithm and approach used to determine these biophysical metrics within the GEDI waveforms.

Foreword

This document is the Algorithm Theoretical Basis Document for the GEDI L2B Footprint Canopy Cover and Vertical Profile Metrics. The GEDI Science Team assumes responsibility for this document and updates it, as required, as algorithms are refined. Reviews of this document are performed when appropriate and as needed updates to this document are made.

This document is a GEDI ATBD controlled document. Changes to this document require prior approval of the project. Proposed changes shall be noted in the change log, as well as incrementing the document version number.

Questions or comments concerning this document should be addressed to:

Hao Tang
2138 LeFrak Hall, Department of Geographical Sciences
University of Maryland, College Park MD 20742
htang@umd.edu
+1 (301) 405 3076

John Armston
1150 Lefrak Hall, Department of Geographical Sciences
University of Maryland, College Park MD 20742
armston@umd.edu
+1 (301) 405 8444

Change History Log

Revision Level	Description of Change	Date Approved
1.0	Initial version	Dec 6, 2019

Table of Contents

List of Tables

List of Figures

1.0 INTRODUCTION

1.1 GEDI Mission Overview

The Global Ecosystem Dynamics Investigation (GEDI) responds directly to observational priorities set by the National Academy of Sciences and NASA's Science Mission Directorate, which emphasizes the need for lidar vertical structure measurements to address key challenges in carbon cycling and biodiversity. The mission aims to answer three fundamental forest ecosystem science questions:

1. What is the carbon balance of the Earth's forests?
2. How will the land surface mitigate atmospheric CO_2 concentrations in the future?
3. How does forest structure affect habitat quality and biodiversity?

To address these science questions, the mission has identified four primary objectives:

1. Quantify the distribution of aboveground carbon stored in vegetation
2. Quantify the effects of vegetation disturbance and recovery on carbon storage
3. Quantify the potential for existing and new or regrowing forests to sequester carbon in the future
4. Quantify the spatial and temporal distribution of habitat structure and its influence on habitat quality and biodiversity

GEDI aims to fill current observation gaps by collecting the first high-resolution lidar observations of the 3D structure of the Earth with a geodetic-class, full-waveform light detection and ranging (lidar) laser system. It will provide spatially dense samplings of forest structure between ~52° N and ~52° S and is expected to produce about 10 billion cloud-free observations during its nominal 24-month mission length.

1.2 GEDI Data Products Overview

The GEDI data products are noted in Table 1. The GEDI Level 1 data products are developed in two separate products, a Level 1A (L1A) and a Level 1B (L1B) product. The GEDI L1A data product contains fundamental instrument engineering and housekeeping data as well as the raw waveform and geolocation information used to compute higher level data products. The GEDI L1B geolocated waveform data product, while similar to the L1A data product, contains specific data to support the computation of the higher level 2A and 2B data products. These L1B data include the corrected receive waveform, as well as the receive waveform geolocation information. The L1B data products provide end users with context for the higher L2 products as well as the ability for end users to apply their own waveform interpretation algorithms. The L2 products contain information derived from the geolocated GEDI return waveforms, including ground elevation, height and structure metrics and other waveform-derived metrics describing the imaged surface.

Table 1. GEDI data products

Product	Description	Data Latency	Archive Site
Level 1B	Geolocated Waveforms	4 months in monthly intervals	LPDAAC
Level 2A	Elevation and Height Metrics	4 months in monthly intervals	LPDAAC
Level 2B	Canopy Cover and Vertical Profile Metrics	4 months in monthly intervals	LPDAAC
Level 3	Gridded Land Surface Metrics	4 months in monthly intervals	ORNLDAAC
Level 4A	Footprint Aboveground Biomass	6 months after global sampling required to meet L1 requirement	ORNLDAAC
Level 4B	Gridded Aboveground Biomass	6 months after global sampling required to meet L1 requirement	ORNLDAAC
Level 4 Demonstrative	• Ecosystem model outputs • Enhanced height/biomass by fusion with TanDEM-X & Landsat • Habitat model outputs	-	-

1.3 GEDI Configuration

The GEDI instrument is a geodetic-class, full-waveform light detection and ranging (lidar) laser system comprised of 3 lasers producing a total of 8 beam ground tracks that are spaced approximately 600 m apart on the Earth's surface in the cross-track direction relative to the flight direction, and approximately 735 m of zonal (parallel to lines of latitude) spacing. Each beam transect consists of ~25 m footprint samples approximately spaced every 60 m along track. The "coverage" laser is split into two transects that are then each dithered producing four ground transects. The other two lasers are dithered only, producing two ground transects each. The configuration of the ground tracks is shown in Figure 1. A GPS system provides knowledge of where the GEDI instrument is in its orbit relative to the surface of the Earth, while star trackers give the orientation of the instrument. The ranging points from each footprint's waveform are geolocated to produce geolocation data groups ("geolocation" and "geophys_corr") provided in the L1 and L2 data products.

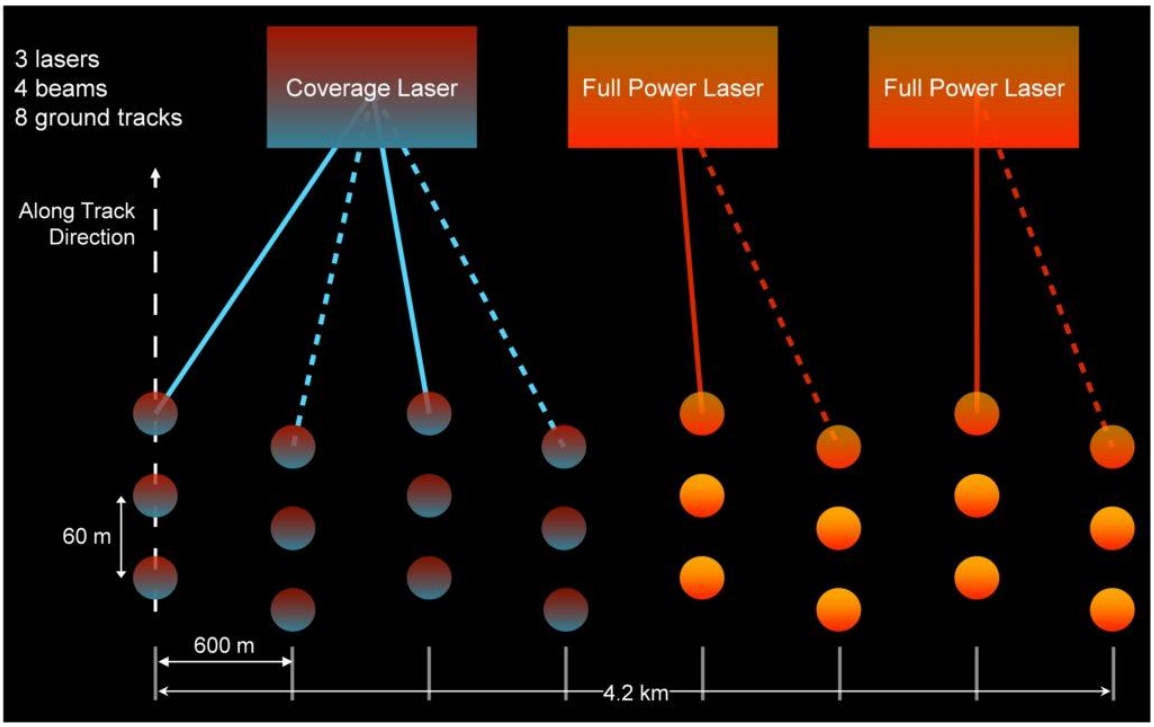

Figure 1. GEDI beam ground-track configuration

1.4 Document Overview and Objectives

This algorithm theoretical basis document (ATBD) for GEDI Level 2B footprint canopy cover and vertical profile metrics is designed to provide both: (1) a general theoretical overview of the algorithms, processing steps and procedures required to provide measurements of canopy profile metrics and (2) a detailed algorithm implementation and processing flow specifically designed for the GEDI mission level 2B products.

This topic is arranged in the following manner:

- **Section 1** presents a brief introduction and related documentation
- **Section 2** presents an overview of deriving GEDI vegetation profile metrics
- **Section 3** presents details of the retrieval algorithm for GEDI L2B products
- **Section 4** contains references
- An acronym glossary can be found at the end of this document

1.5 Historical Background

Vegetation structure, defined as the aboveground spatial arrangement of individual plant crowns, is closely linked to forest ecological functions and services. In the ecological literature, several key metrics describe three-dimensional distribution and complexity of the vegetation structure. Of particular importance are total canopy cover and its vertical profile, leaf area index (LAI) or plant area index (PAI) and its vertical profile, as well as

foliage height diversity (FHD). Accurate estimates of these variables are of critical importance in studies of global environmental change and terrestrial biodiversity. Tremendous efforts from the remote sensing community have resulted in vegetation structure products at different spatial and temporal resolutions (e.g. ~10 m to 10 km, daily to annually) since the 1980s, primarily using data collected from passive optical remote sensing platforms. To date there have been a large number of regional to global scale canopy cover and LAI data sets available, including MODIS (Myneni et al. 2002; Defries et al. 2000), Landsat (Ganguly et al. 2008; Hansen et al. 2013; Sexton et al. 2013) and many others (Weiss et al. 2007). These products have been widely used in studies of terrestrial ecosystems and have greatly improved our knowledge and understanding of global environment and biodiversity change (Mu et al. 2007; Zhao et al. 2005; Myneni et al. 2007; Hansen et al. 2013; Busch et al. 2015; Harris et al. 2012; Goetz et al. 2015; Pimm et al. 2014; Rose et al. 2015). However, three key issues remain unresolved: first, spectral signals of satellite images saturate over densely forested areas (Myneni et al. 2002; Wenze Yang et al. 2006); second, products derived from different sensors (or even from the same sensor) are inconsistent (Abuelgasim et al. 2006; Garrigues et al. 2008; Morisette et al. 2006; Fang et al. 2012; Sexton et al. 2016); and finally, measurements of vertical canopy structure are largely unavailable.

Lidar remote sensing offers a potential solution to these problems by providing precise measurements of three-dimensional forest structure (Lefsky et al. 2002; Dubayah & Drake 2000; Harding 2001). Both plot-level and landscape-scale products of canopy cover, LAI and their profiles have been derived from terrestrial laser scanning and airborne discrete return or waveform lidar (Armston et al. 2013; Calders et al. 2014; Tang et al. 2012; Tang, Brolly, et al. 2014; Hancock et al. 2014; Yang et al. 2013; F. Zhao et al. 2011; Hancock et al. 2011; Hancock et al. 2012; Parker et al. 2001; Falkowski et al. 2008; Wulder et al. 2007). These products cover all major biomes including dense tropical rainforests, and achieve high agreements with radiation-interception based techniques (e.g. LAI-2000 and hemispherical photons) and destructively sampled data. Attempts have also been made to generate similar data sets but over much broader geographical areas using data from spaceborne waveform lidar systems, such as ICESat (Tang, Dubayah, et al. 2014; Tang et al. 2016; Harding 2005; Garcia et al. 2012; Luo et al. 2013; Lee et al. 2011). However, further ecological applications of ICESat-derived products are still subject to limited data coverage and they are not optimized for forest-related studies. This is because ICESat is primarily designed to monitor ice dynamics in polar regions, resulting in sparsely distributed footprints over major forests at mid-low latitude (Abshire et al. 2005). The large footprint size of ICESat (~70 m in diameter) can introduce substantial measurement uncertainty of canopy structure, particularly over steep slopes (Duncanson et al. 2010; Pang et al. 2011). Yet no other canopy profile product exists globally.

Several lidar satellite missions have been proposed before GEDI, aiming to obtain reliable measurements of canopy profile metrics globally. These include the Vegetation Canopy Lidar (VCL) mission (Dubayah et al. 1997) and the Deformation, Ecosystem Structure, and Dynamics of Ice (DESDynI) mission (Dubayah et al. 2008). However, these missions did not get launched because of several issues in either immature technique or financial budget sequestration (F. Hall et al. 2011; Goetz 2011). Heritages from these missions

suggest that a waveform lidar system with ~25 m horizontal resolution and ~1 m vertical accuracy, is desired for accurately measuring vertical canopy structure (± 10%) in presence of moderate slopes (F. G. Hall et al. 2011). These requirements fit well into current design of GEDI for estimating canopy cover, LAI and FHD.

1.6 Algorithm Objectives

The algorithm specified in this document is designed to derive footprint level canopy cover and vertical profile over vegetated areas between ~52°N and ~52°S. The data product includes estimates of total canopy cover and PAI, vertical profiles of canopy cover and PAI, the vertical profile of Plant Area Volume Density and foliage height diversity. The GEDI Level 2A and 2B products will provide unprecedented dense spatial samplings of forest structure globally.

1.7 Cover and Vertical Profile Metrics

Canopy cover is a biophysical parameter widely used in terrestrial remote sensing to describe the spatially aggregated geometric properties of vegetation. Multiple definitions of canopy cover exist, depending on the applied measuring techniques (Fiala et al. 2006; Hansen et al. 2002; Hansen et al. 2003; Jennings et al. 1999; Korhonen et al. 2011; Rautiainen et al. 2005). The central issues in the definition are (1) whether the measurement is acquired at a specific viewing angle (mostly near-nadir) or over the entire hemisphere; and (2) whether a tree crown is treated as an opaque object including all small within-canopy gaps. In contrast to traditional passive optical sensors, lidar systems measure the forest mostly at a small zenith-viewing angle. An off-nadir angle of discrete return airborne lidar is typically < 15° and it is < 6° for GEDI to avoid inaccurate range measurements. While airborne lidar can help delineate individual tree crown with high spatial resolution and footprint density (i.e. counting small openings as part of tree cover), large footprint waveform lidar systems like GEDI can only provide estimates of canopy fractional cover over the laser-illuminated area. Thus, the GEDI derived canopy cover is the percent of the ground covered by the vertical projection of canopy material (i.e. leaves, branches and stems only). It is different from two other widely used cover types: canopy closure defined as "the proportion of the vegetation over a segment of the sky hemisphere at one point on the ground", or crown cover as "the percentage of the ground covered by a vertical projection of the outermost perimeter of the natural spread of the foliage of plants"(Jennings et al. 1999). The canopy cover profile is the horizontally-intercepted canopy elements at a given height, and is calculated as one minus gap distribution at that height level.

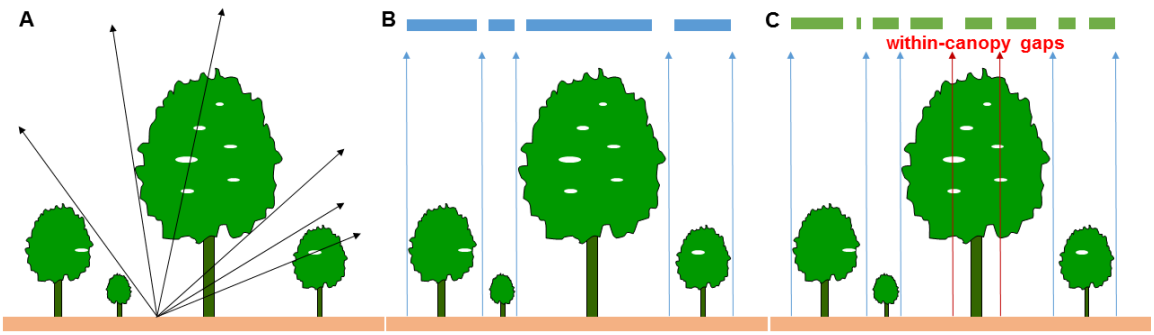

Figure 2. Three types of canopy cover: canopy closure (A), crown cover (B), and canopy fractional cover (C). GEDI will only produce canopy fractional cover.

LAI is defined as one half of the total leaf area per unit ground surface (Chen & Black 1992; Chen et al. 1997; Gower & Norman 1991) and is closely linked to canopy cover through the gap distribution within canopy. It is typically utilized in the majority of ecological, hydrological and climate models to describe interactions between the biosphere and atmosphere (Myneni et al. 2002; Weiss et al. 2007). A closely related concept is plant area index (PAI) that incorporates all canopy structural elements (e.g. branch and trunk) in addition to leaves, and their numeric difference is usually small in dense broadleaf forests (e.g. LAI ≈ 93% PAI) (Tang et al. 2012). In this ATBD, we do not explicitly differentiate these two and use them equivalently hereafter for simplicity. Vertical LAI profile is the vertical variation of LAI closely related to foliage-height profiles (Aber 1979; Parker et al. 1989). It is a more realistic representation of three-dimensional canopy structure that can be used to quantify flows of energy and material in ecosystems (Vose et al. 1995; Ellsworth & Reich 1993; Parker et al. 2005; Stark et al. 2012). It can also be used to describe growth pattern of forests at different succession stages (Parker 2004). Different lidar systems, including terrestrial, airborne and spaceborne sensors, have demonstrated their capacity in deriving highly accurate LAI and profile products even in extremely dense forests (Armston et al. 2013; Tang et al. 2012; Hancock et al. 2014; Strahler et al. 2008; Zhao et al. 2012; Tang, Brolly, et al. 2014; F. Zhao et al. 2011; K. Zhao et al. 2011; Morsdorf et al. 2006). However, large-footprint lidar systems cannot directly measure leaf clumping conditions or leaf angle distributions. Ancillary information (e.g. clumping index from multi-angle observations) is required to convert lidar-perceived effective LAI into true LAI (Chen & Black 1992; Chen et al. 2005; Pisek et al. 2015). Yet there is no such database available globally at GEDI's footprint scale (~25 m). In consideration of the aforementioned limitations, GEDI only aims to derive PAI and its vertical profile as a more direct and accurate product. The users are recommended applying conversions based on their best knowledge to convert GEDI's PAI to LAI for running earth system models (Myneni et al. 2002; Weiss et al. 2007).

FHD measures the complexity of canopy structure. It is also known as Shannon's diversity index, the Shannon-Wiener index, or the Shannon entropy in the ecological literature, and it was originally proposed to quantify the entropy (uncertainty or information level) in information theory (MacArthur & Horn 1969). A high FHD value in forest ecology often results from more complex forests structure (e.g. caused by multiple canopy layers) (James

& Wamer 1982; Bergen et al. 2009). This complexity is a good indicator of habitat quality for wild life, as suggested by pioneering studies of MacArthur & Horn (1961). In particular, biodiversity patterns of birds are widely studied in the context of vegetation structure (Clawges et al. 2008; Vierling et al. 2008; Huang et al. 2014; Swatantran et al. 2012; Goetz et al. 2007; Goetz et al. 2010; Goetz et al. 2014). Similar relationships apply among other taxa as well (Aguilar-Amuchastegui & Henebry 2007; Carey & Wilson 2001; Gardner et al. 1995). Recent developments of lidar remote sensing have promised an enhanced measurement capability of FHD that was previously limited at a few plot samplings due to high labor cost. Since FHD measurement is largely based on estimates of vertical LAI profile, it can also be directly derived from GEDI.

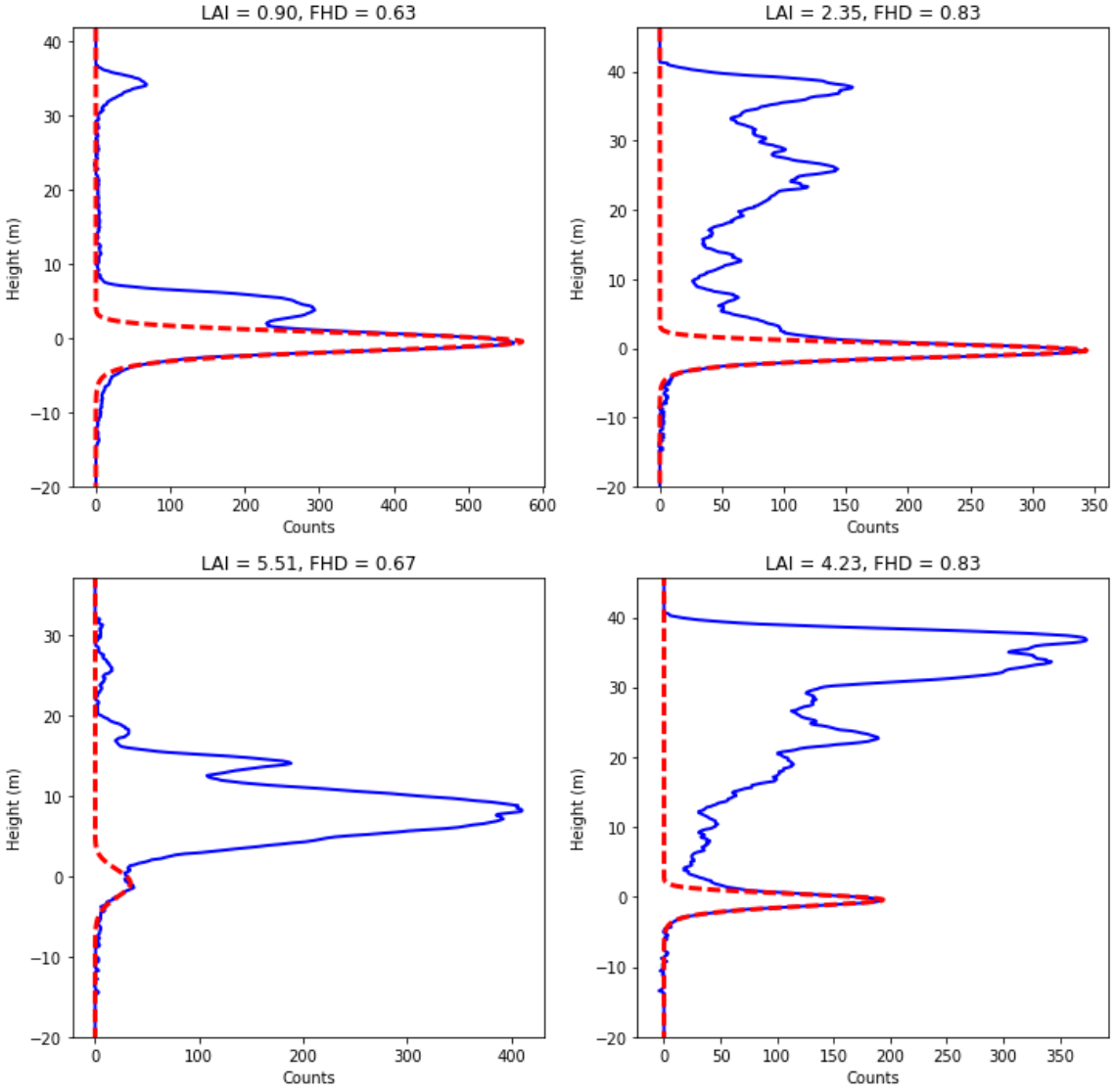

Figure 3. Examples of LVIS waveforms (blue: raw waveform; red: fitting of ground portion) and associated LAI and FHD values in Pongara National Forests, Gabon. Note that FHD is not necessarily correlated with LAI, but an independent measurement of vertical canopy structure complexity.

1.8 Related Documentation

Related documents include parent documents and applicable documents, and information documents.

1.8.1 Parent Documents

- GEDI Science Data Management Plan

1.8.2 Applicable Documents

- GEDI ATBD for GEDI Waveform Geolocation for L1 and L2 Products.
- GEDI ATBD for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products
- GEDI L1A Product Data Dictionary (gedi_l1a_product_data_dictionary.html)
- GEDI L1B Product Data Dictionary (gedi_l1b_product_data_dictionary.html)
- GEDI L2A Product Data Dictionary (gedi_l2a_product_data_dictionary.html)
- GEDI L2B Product Data Dictionary (gedi_l2b_product_data_dictionary.html)

2.0 THEORETICAL BACKGROUND

This section describes theoretical basis of deriving GEDI canopy profile metrics. It includes classical theories in the ecological literature, as well as practical models in lidar remote sensing.

2.1 Canopy Directional Gap Probability and Vertical Profile Metrics

Measurement of LAI can be based on estimates of canopy directional gap probability, which is the complement of canopy cover. This gap theory, following Beer's law, has been well established to quantify the relationship between total LAI and gap frequency for horizontally homogenous canopy layers (Chen & Cihlar 1995; Chen et al. 1997; Gower & Norman 1991; Miller 1967; Nilson 1971; Nilson 1999).

$$P(\theta) = 1 - Cover(\theta) = e^{-G(\theta)*\Omega(\theta)*PAI/cos\,(\theta)} \quad (1)$$

where $P(\theta)$ is the gap probability within canopy with a view zenith angle of θ, $G(\theta)$ is the leaf angle projection coefficient representing the fraction of leaf area projected perpendicular to the view direction to the total leaf area, and $\Omega(\theta)$ is the clumping index determined by the spatial distribution pattern of foliage elements. Note this model uses Plant Area Index (PAI) rather than LAI, since branches and trunks also reflect laser energy. But we do not explicitly consider the difference between the two here as described in Section 1.6. We also build the model under the assumption that foliage elements are dispersed randomly and independently between layers so that the number of overlaps follows the Poisson distribution.

Measurement of vertical canopy structure, including canopy cover and LAI profiles, can be related to estimates of vertical canopy directional gap probability in a similar way. Here we define ***Pgap(z, θ)*** as the probability of a beam of infinitesimal width at zenith angle θ to the local normal, being directly transmitted through a canopy. Hence ***Pgap(z, θ)*** is equivalent to the probability that the ground surface is directly visible from airborne and spaceborne remote sensing platforms. We also define ***Fapp(z)*** as the apparent foliage profile under nadir-viewing angle. It is the surrogate for effective vertical LAI profile and describes the horizontally intercepted leaves at height *z*. Thus, the probability of having no leaves above *z +Δz*, $P(z + \Delta z)$ is the joint probability of *P(z)* and the probability of having no leaves within a thin layer of Δz by satisfying the following equation:

$$Pgap(z + \Delta z, \theta) = Pgap(z, \theta) * (1 - F_{app}\,(z) \times \Delta z \,/\, \cos\theta) \quad (2)$$

By reconstructing the above equation, we can quantify their relationship in Eq. 3 and Eq. 4 when Δz approaches zero:

$$F_{app}\,(z) = \lim_{\Delta z \to 0} -\frac{Pgap(z+\Delta z,\theta) - Pgap(z,\theta)}{Pgap(z,\theta) \times (\Delta z \,/\, \cos\theta)} = \lim_{\Delta z \to 0} -\frac{\Delta Pgap(z,\theta)}{Pgap(z,\theta) \times (\Delta z \,/\, \cos\theta)} = \frac{d\ln\{Pgap(z,\theta)\}}{dz/\cos\theta} \quad (3)$$

$$PAI(z) = \frac{F_{app}\,(z)}{G(z,\theta) \times \Omega(z,\theta)} = -\frac{1}{G(z,\theta) \times \Omega(z,\theta)} \times \frac{dln\{Pgap(z)\}}{dz/\cos\theta} \quad (4)$$

2.2 Lidar models for Canopy Directional Gap Probability and Vertical Profile Metrics

In a waveform-scanning lidar system, the received laser energy is a function of incident laser energy, atmosphere conditions, and spatial and spectral properties of targets (i.e. canopy and ground). Ni-Meister et al. (2001) developed an approach to derive canopy gap probability from lidar waveforms with no requirement for radiometric calibration. The basic assumption of the model is that gap probability is the complement of the vertical canopy profile as laser energy can only penetrate into the lower canopy layer or ground through gaps (including both within-crown gaps and between-crown gaps). Multiple-scattering effect is not considered in the model since it has a limited contribution (< 10%) towards the total waveform energy and a very small impact on the waveform shape between canopy top and ground. Based on this model, canopy cover can be calculated using the cumulative laser energy return for a known ratio of canopy and ground reflectance as follows.

Let $Pgap(z+\Delta z) - Pgap(z)$ be the expected proportion of laser energy intercepted by the canopy elements. Then the laser energy budgets at height z and at ground level are:

$$-\frac{dR_v(z)}{dz} = J_0\rho_v \frac{dPgap(z,\theta)}{dz/\cos\theta} \quad (5)$$

$$R_g = J_0\rho_g Pgap(0,\theta) \times \cos\theta \quad (6)$$

where $Pgap(z,\theta)$ and $CCF(z)$ are the directional gap probability and canopy cover fraction above a particular height z within canopy respectively. The terms $R_v(z)$, $R_v(0)$ and R_g are the integrated laser energy returns from the canopy top to height z, from canopy top to canopy bottom, and from the ground return individually. The J_0 is the irradiance of emitted laser pulse and is only known for GEDI when the Level 1B waveforms are calibrated, the emitted pulse power is measured, and atmospheric transmittance is derived. The ρ_v, ρ_g are canopy and ground reflectance respectively, while in practice, they can also be uncalibrated reflectance values (also known as volumetric scattering coefficients in the literature) since it is only their ratio impacts retrieval results.
The boundary conditions are:

1) There is no laser energy at the top of canopy: $Rv(max\_ht) = 0$; and
2) the gap is constant above canopy top: $Pgap(max\_ht, \theta) = 1$.

By solving above equations under these boundary conditions, we have:

$$Pgap(z,\theta) = 1 - \frac{R_v(z)}{R_v(0)+R_g\times\frac{\rho_v}{\rho_g}} \quad (7)$$

$$CCF(z) = (1 - Pgap(z,\theta)) \times \cos\theta \quad (8)$$

$$Fapp(z_1 \sim z_2) = \sum_{z_1}^{z_2} \frac{d\ logPap(z,\theta)}{dz/\cos\theta}\Delta z \quad (9)$$

The log transformation of gap probability follows the same approach in MacArthur and Horn (1969), that the density of foliage can be estimated from the distribution of first leaf distance. The full details of this model can be found in Ni-Meister et al. (2001). Note the impact of finite $\Delta z = 0.15$ m is ignored here since it is typically much smaller than canopy height. The viewing angle of GEDI is near-nadir (< 6 ° most of the time), and we also ignore variations in leaf angle distribution when calculating vertical LAI profile as the summation of vertical foliage density from canopy height z_1 to z_2:

$$PAI(z_1 \sim z_2) = \frac{1}{G \times \Omega} \times \sum_{z_1}^{z_2} Fapp(z) \Delta z \quad (10)$$

GEDI will sample the PAI and canopy cover profile between the estimated ground elevation (z = 0) and the maximum canopy height at a specified vertical interval. We do not aim to delivering products at the native vertical resolution because the returned waveform is convolved by system pulse width, detector response function and the illuminated target. The impact of convolution on the estimation of vertical foliage profiles may be corrected by either applying a deconvolution technique, or aggregating waveform bins to a coarser vertical resolution. In practice, there seems to be a trade-off between vertical resolution and retrieval accuracy as examined in a previous study (Tang et al. 2016).

2.3 Foliage Height Diversity (FHD)

The foliage height diversity (FHD) is a canopy structural index describes the vertical heterogeneity of foliage profile (MacArthur & Horn 1969). It can be quantified as:

$$FHD = -\sum_i N_i \times log\,(N_i) \quad (11)$$

where N_i is the proportion of vertical LAI profile lies in the i^{th} of the chosen horizontal layers.

$$N_i = \frac{LAI(z_i \sim z_{i+1})}{LAI} \times \Delta z \quad (12)$$

3.0 ALGORITHM

The baseline algorithm to retrieve GEDI canopy profile metrics is based on the lidar-PAI model described in Section 2, with following required input parameters: 1) the integrated laser energy returns from the canopy *Rv(z)* and ground *Rg*. 2) leaf projection coefficient *G*, 3) clumping index Ω and 4) the ratio of canopy and ground volumetric backscattering coefficients ρ_v/ρ_g.

A constant projection coefficient of G = 0.5, corresponding to a uniform random distribution, is used in this ATBD and is based on an assumption of a random distribution of canopy elements that has been published and widely used across multiple biomes. The clumping index cannot be derived from GEDI data independently and a constant Ω=1 is used here to calculate PAI (and its vertical profile).

3.1 Overview

The processing chain is initialized with inputs variable from lower level GEDI data products (L1A, L1B and L2A) to retrieve both the vertical canopy energy distribution *Rv(z)* and ground energy (*Rg*). Next, the vertically resolved directional canopy gap probability of an individual footprint, *Pgap(θ,z)*, is calculated from *Rv(z)* and the ρ_v/ρ_g value extracted from external grids that are initialized pre-launch and updated post-launch. This external grid will be updated post-launch using GEDI waveforms of highest quality that have high beam penetration capacity (through aerosol and dense forests) and low atmospheric and topographic impacts. All other output GEDI L2B cover and vertical profiles metrics are subsequently calculated from the vertical directional gap probability profiles with quality flags for exception handling. More details about product and data variables generated or used in this ATBD can be found in **Error! Reference source not found.** and Table 3.

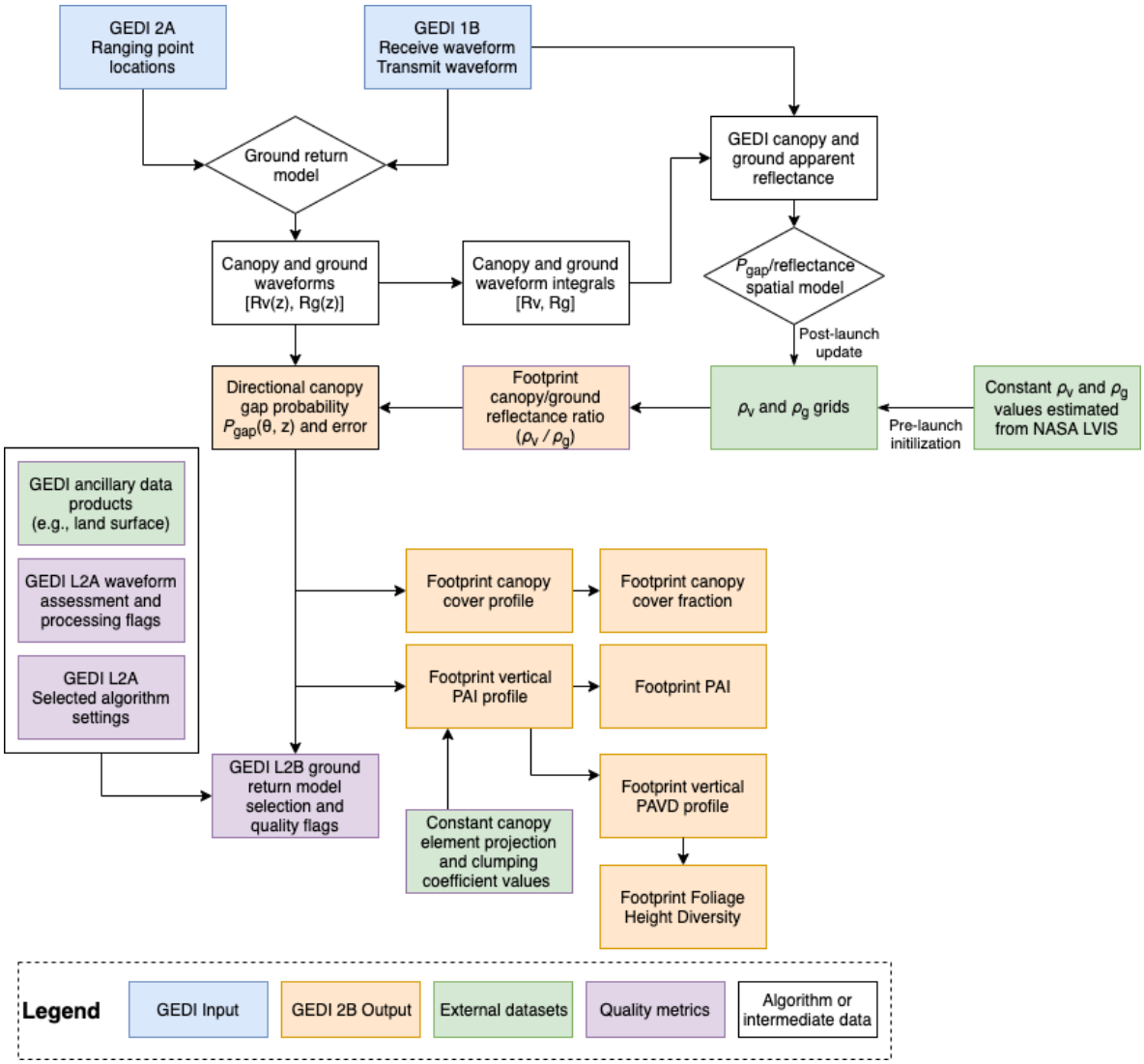

Figure 4. GEDI L2B algorithm flowchart

3.2 Product Names and Data Variables

Table 2. GEDI L2B Data Product Names

Data Product Level	Data Product Name	Spatial Resolution	Parameter Name	Target Area	Description	Source ATBD
L2	Canopy Cover and Vertical Profile Metrics	Footprint	Canopy Cover Fraction (CCF)	All land surfaces	Direct estimate of total CCF	GEDI02_B (Cover and Vertical Profile Metrics)
L2	Canopy Cover and Vertical Profile Metrics	Footprint	Plant Area Index (PAI)	All land surfaces	Modelled estimate of PAI	GEDI02_B (Cover and Vertical Profile Metrics)
L2	Canopy Cover and Vertical Profile Metrics	Footprint	Vertical Cover Profile	CCF > 0	Direct estimate of the cumulative vertical cover profile at 5 m vertical resolution.	GEDI02_B (Cover and Vertical Profile Metrics)
L2	Canopy Cover and Vertical Profile Metrics	Footprint	Vertical PAI Profile	CCF > 0	Modelled estimate of the cumulative vertical PAI profile at 5 m vertical resolution.	GEDI02_B (Cover and Vertical Profile Metrics)
L2	Canopy Cover and Vertical Profile Metrics	Footprint	Vertical Plant Area Volume Density Profile	CCF > 0	Modelled estimate of the vertical plant area volume density profile with vertical resolution of 5m	GEDI02_B (Cover and Vertical Profile Metrics)
L2	Canopy Cover and Vertical Profile Metrics	Footprint	Foliage Height Diversity (FHD)	CCF > 0	Canopy index that describes the vertical heterogeneity of the foliage profile	GEDI02_B (Cover and Vertical Profile Metrics)

Table 3. Data product variables. These include ancillary data sets that are required in the algorithm for generating the standard product.

Product	Name	Units	Description	Variable	Required Precision	Min	Max	Source
GEDI02_B	Elevation of lowest mode	m	elevation of center of lowest mode relative to reference ellipsoid	elev_lowest mode	NA	-1 km	25 km	GEDI02_A
GEDI02_B	Canopy top location	NA	Sample number of highest detected return	toploc	NA	0	1420	GEDI02_A
GEDI02_B	Ground bottom location	NA	Sample number of lowest detected return	botloc	NA	0	1420	GEDI02_A
GEDI02_B	Directional gap probability	NA	The probability a beam of infinitesimal width will penetrate to the ground surface at view zenith angle theta	Pgap(z, theta)	0.001	0	1	Algorithm
GEDI02_B	Canopy Cover Fraction	NA	Total cover	cover	0.001	0	1	Product
GEDI02_B	Vertical profile of canopy cover	NA	Cumulative canopy cover from z=H to z=0, where cover_z(H) =0	cover_z	0.001	0	1	Product
GEDI02_B	PAI	m2 / m2	One half of the total plant area projected per unit ground surface.	pai	0.01	0	16	Product
GEDI02_B	Vertical PAI Profile	m2 / m2	Cumulative PAI from z=H to z=0, where pai_z(H) = 0	pai_z	0.01	0	16	Product

GEDI02_B	Plant area volume density profile	m2 / m3	Foliage profile assuming a random distribution of canopy elements and constant leaf angle with height	pavd_z	0.0001	0	inf	Product
GEDI02_B	Maximum canopy height	cm	Maximum above ground height of the canopy within a waveform footprint	Rh100	0.01	-21300	21300	GEDI02_A
GEDI02_B	Ground scattering coefficient	counts	Volumetric scattering coefficient of the ground (reflectance x phase function)	rhog	0.001	0	1	Product
GEDI02_B	Canopy scattering coefficient	counts	Volumetric scattering coefficient of the canopy (reflectance x phase function)	rhov	0.001	0	1	Product
GEDI02_B	Ground energy summation	counts	Integral of the denoised ground waveform	Rg	0.001	0	Inf	GEDI01_A
GEDI02_B	Canopy energy summation	counts	Integral of the denoised canopy waveform	Rv	0.001	0	inf	GEDI01_A
GEDI02_B	Local beam elevation angle	radians	Elevation of the unit pointing vector for the laser in the local ENU frame.	local_beam_elevation	0.01	0	Pi/2	GEDI01_B
GEDI02_B	Foliage clumping factor	NA	Ratio of Pgap(theta) for a clumped canopy and Pgap(theta) for a random	Omega	0.01	0	1	External

			canopy of the same LAI					
GEDI02_B	Ross-Nilson's G-function	NA	Mean projection of unit leaf area on a plane perpendicular to the direction of the laser beam at view zenith angle theta	Ross-G	0.01	0	1	External
GEDI02_A	Land cover	NA	The land cover and biome type of the illuminated surface identified by the external source	LC	NA	0	255	External
GEDI01_A	Waveform sensitivity quality	NA	The flag if the laser energy penetrates through the canopy and triggers a ground return	Sensitivity	0.001	0	1	GEDI02_A
GEDI02_B	Waveform Surface flag	NA	The flag if the laser hits the ground surface	Surface_flag	NA	0	1	GEDI02_A
GEDI02_B	Leaf off status	NA	The flag indicating if the laser illuminated forests have a leaf-on phenological status identified by external source	leaf_off_flag	NA	0	1	External
GEDI02_B	Selected L2A algorithm setting	NA	ID of L2A algorithm selected as identifying the lowest non-noise mode	selected_l2a_algorithm	NA	1	6	GEDI02_A
GEDI02_B	Selected Rg (ground) waveform model	NA	0 = L2B algorithm not run; 1 = algorithm 1	selected_rg_algorithm	NA	0	3	Product

			(successful); 2 = algorithm 1 (partial success - valid tx_eg parameters were unavailable); 3 = algorithm 2					

3.3 Implementation

The lidar waveforms and geolocation parameters are provided by the Level 1B and 2A data products. Footprint level variations in the foliage element projection coefficient or clumping index are not considered in the algorithm because such information cannot be directly extracted from GEDI waveforms alone. Neither is available a global clumping index product at the GEDI footprint level: the clumping condition at the ~25 m footprint scale can be quite different from that of a much coarser resolution (e.g. a 500m pixel from a Moderate Resolution Imaging Spectroradiometer (MODIS) product (He et al. 2012)). Therefore we assume a random uniform angular distribution of canopy scattering elements and constant leaf angle with height to set a constant projection coefficient of $G = 0.5$ and clumping index $\Omega = 1$ for all woody vegetation. In the future we will test and possibly incorporate foliage clumping databases in development (e.g. Pisek et al., 2015) to further improve GEDI-based products.

The input canopy/ground waveform parameters are provided by the lower level GEDI product (Level 1A and 2A), originally based on a Gaussian decomposition method (Hofton et al. 2000). This method has been widely used to process waveforms of large footprint lidar systems (e.g. ICESat, or the Land Vegetation and Ice Sensor (LVIS) as the airborne prototype of GEDI). The ground energy distribution, Rg, was modelled by fitting an exponentially modified Gaussian (exGaussian) function to the belowground portion of the received waveform (rxwave). Here we did not use the traditional Gaussian function because of the asymmetry of transmitted waveform, considering the ground return should be a convolution of the ground pulse (mostly Gaussian shape due to slope) and the transmitted waveform. The exGaussian function utilized a fourth parameter, gamma (γ), in addition to traditional Gaussian function's amplitude (A), center (μ), sigma (σ), to describe a skewed normal distribution as observed in the transmitted waveform (txwave). The σ and γ estimated from the txwave exGaussian fitting (tx_egsigma and tx_eggmma) are applied as boundary constraints to fit the *Rg* for each rxwave. Finally, the Levenberg-Marquardt least-square method is used to fit the below-ground range of rxwave, from the highest threshold (lowestmode + a buffer size on *Rg* center) to the signal truncation at the trailing edge (botloc).

If the default ground fitting algorithm fails to converge, a match-filter method is used to identify the *Rg* fit. It is achieved by applying a convolution of the mirrored transmit

waveform over the received waveform. At the expense of effective vertical resolution, this operation could mitigate the asymmetrical effect while retaining the original estimate of ground elevation. Another drawback of the method is the wider spread of ground energy distribution. Nevertheless, it provided a simple and efficient mathematical solution to estimate the contribution from ground energy in a waveform. To estimate the total integral of *Rg*, the belowground portion of the convolved rxwave is mirrored.

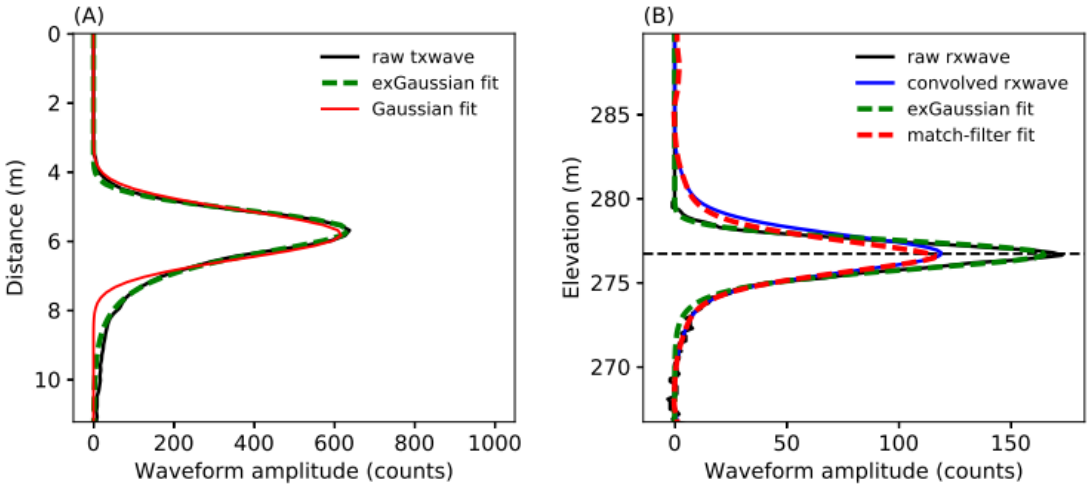

Figure 5. Examples of the recorded and fitted GEDI waveforms: (A) transmitted waveform (txwave) and (B) the correspondent return waveform (rxwave). The traditional Gaussian function cannot fully characterize the asymmetrical txwave, while the exponentially modified Gaussian (exGaussian) function can reconstruct it remarkably well when adding a fourth skewness parameter, gamma. The parameters of exGaussian estimated from the txwave are also transferable when fitting the rxwave,

The ratio of canopy and ground reflectance is initialized with a default global value (1.5) for the release of version 1. These values can be updated post-launch using a regression analysis between multiple GEDI footprints. The rest of this section will focus on the retrieval of ρ_v/ρ_g in both pre-launch and post-launch modules.

3.3.1 Pre-launch initialization

The specification of ρ_v/ρ_g requires pre-launch initialization to a constant value globally (the simplest case) or allowing variation in constant values between regions. The use of constant ρ_v/ρ_g regionally has been a reasonable assumption and a common practice to derive above canopy profile metrics from both airborne and spaceborne waveform lidar systems. High retrieval accuracy has been achieved across major biomes in comparison with different reference data (e.g. destructive samplings, terrestrial laser scanning (TLS), LAI-2000 and hemispherical photo) (Armston et al. 2013; Tang et al. 2012; Tang, Brolly, et al. 2014; Hancock et al. 2014).

The ρ_v/ρ_g value comes from two baseline designs (1) a global constant value and (2) a pre-launch LUT identifying constant ρ_v/ρ_g values for regional strata. In baseline #1, we apply a global average value of 1.5 although it could typically vary from about 1.0 to 2.0 for 1064-nm laser at the footprint level in space and time (Harding et al. 2001; Tang et al.

2012). In baseline #2, biome-level constant ρ_v/ρ_g vales are determined by applying the P_{gap}-reflectance model to LVIS and other 1064 nm airborne lidar data over a range of landscapes. Technical details on this model can be found in Section 3.2.2. Existing LVIS acquisitions have covered major biome types, enabling a reasonable biome representative for the reflectance ratio. These include tropical evergreen rainforests, temperate deciduous broadleaf forests, temperate mixed forests, wetland forests and mountainous conifer forests. We follow MODIS IGBP land cover classification type and assign LVIS-based estimates to corresponding categories (Table 4). There are more LVIS flights just completed but unprocessed or to be deployed subject to NASA's requirement in the upcoming years.

Table 4. The biome-level 1064-nm ρ_v/ρ_g values estimated from processed airborne LVIS campaigns (those in italic)

MODIS IGBP Type	**LVIS Acquisition**	$\boldsymbol{\rho_v/\rho_g}$
Evergreen Needleleaf forest (1), Deciduous Needleleaf forest (3)	Sierra Nevada forest (1999, 2008)	1.2
Evergreen broadleaf forest (2)	Costa Rica (1998, *2005*), Panama (1998), Gabon (2016)	1.5
Deciduous broadleaf forest (4)	Mid-Atlantic, USA (*2003*), Duke forest, NC, USA (1999)	1.3
Mixed forest (5)	New England, USA (2003, *2009*)	1.3
Others		1.5

3.3.2 Post-launch updating

The external grids of *Pgap*, ρ_v, and ρ_g are updated post-launch (approximately 6 months to obtain sufficient GEDI observations), based on the model outlined in Armston et al. (2013). The basic model is:

$$R = \left(1 - P_{gap}\right)\rho_v + P_{gap}\rho_g \qquad (13)$$

where assuming single scattering and a linear relationship between the recorded signal and power received at-sensor, the total waveform integral (R) is the sum of the canopy and ground scattering coefficients weighted by fraction of the signal intercepted by the canopy and ground, respectively. Under these assumptions, P_{gap} may be estimated from R by:

$$P_{gap}(\theta) = 1 - \frac{R-\rho_g}{\rho_g-\rho_v} \qquad (14)$$

This formulation is very useful when we have prior knowledge of ρ_v and ρ_g, since we do not need to separate the canopy and ground signals. It will also allow an extra constraint on ground finding, once sufficient estimates of ρ_v and ρ_g have been acquired to enable gridding post-launch. To update the external grids of ρ_v and ρ_g, we assume that adjacent footprints have locally constant ρ_v and ρ_g over small geographical extents and at the footprint resolution. The ρ_v and ρ_g values, under such assumption, can therefore be estimated using a linear regression analysis based on observations of adjacent footprints but at different canopy cover levels. The linear model is

$$R_g = \rho_g - \frac{\rho_g}{\rho_v} R_v \qquad (15)$$

which may be solved using simple linear regression (Armston et al., 2013), or Orthogonal Distance Regression (ODR) assuming equal variance in R_v and R_g (Tang et al., 2016). This Pgap-reflectance model was originally implemented for airborne lidar systems (Armston et al. 2013, but can also be applied to large footprint lidar such as LVIR or GEDI. Indeed, its primary assumption is more realistic for large footprint lidar than small-footprint ALS, given a stronger edge effect and higher canopy heterogeneity observed with smaller footprint sizes (Armston 2013). The pre-launch specification will largely depend on the availability of LVIS observations (see 3.3.1). It is though possible to incorporate data obtained from other 1064 nm airborne lidar systems after more through tests.

The following steps are undertaken to update the post-launch estimation of ρ_v/ρ_g grids from GEDI:

1. Estimate apparent lidar reflectance values using the atmospheric extinction coefficient calculated from GEDI Level 1-2 A.
2. Spatial clustering of GEDI shots to solve for ρ_v/ρ_g
3. Evaluate quality of model fitting
4. Update the ancillary data product with new ρ_v/ρ_g estimates based on evaluation criteria (e.g. good ODR fitting results with $r^2 > 0.5$).

We acknowledge significant spatial variance may exist at scales equal to or greater than the GEDI footprint size and the sensitivity of cover and PAI to this variance needs to be considered. Both algorithms (2a and 2b) to estimate ρ_v/ρ_g may require additional development and testing to finalize the operational algorithm. The implementation is also subject to GEDI beam data quality, which can be significantly impacted by cloud and/or aerosol contamination. As a result, we aim to generate an initial version using pre-launch specification and then update the data product using post-launch estimates. This can be an iterative process as GEDI continues collecting data during its two-year mission.

The strategy for updating external grids of post-launch provides the capability and flexibility to refine the default algorithm, keep the external grids under version control and accessible to other GEDI data product algorithms. The first round of post-launch update is currently scheduled at the end of the first data acquisition period (6 months), pending QA/QC and geographical availability of GEDI data. It is expected that these observations can fill most of ecoregion-level gaps and help provide estimates at a coarse resolution (the

exact resolution will depend on the spatial arrangement of the data sets available for such analysis). The baseline for post-launch processing is that all footprints will be assigned a constant value (version 01).

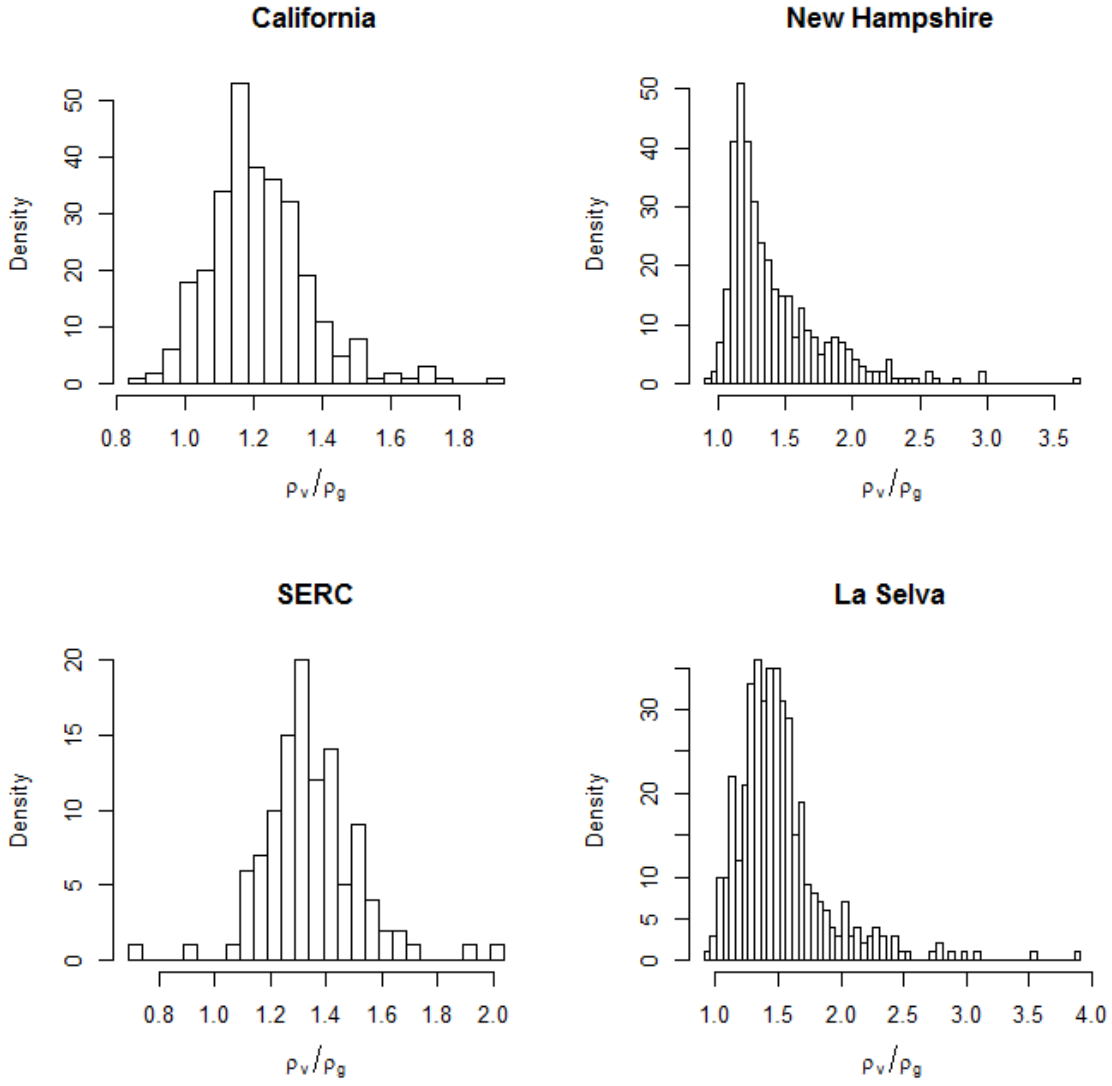

Figure 6. Distribution of LVIS derived ρ_v/ρ_g across major landscape types: conifer forests in California (median: 1.2), temperate mixed forests in New Hampshire (median: 1.3), temperate deciduous forests in Maryland (median: 1.3), and tropical evergreen forests in La Selva, Costa Rica (median: 1.5).

The initial algorithm used Orthogonal Distance Regression (ODR) and assumes constant ρ_v/ρ_g over a regional area, presently specified as a cell between 1 km and 10 km size. Assuming equal error variance between canopy and ground laser energy, we fit the slope and intercept of Eqn. 15 using Orthogonal Distance Regression (ODR). This also assumes ρ_v/ρ_g does not change over time between overpasses. The ρ_v/ρ_g value is equal to β^{-1}. A number of ~20 GEDI shots is expected for each 1 km pixel given the 60 m along-track spacing distance. If the actual acquisition number is less than 10, or all these footprints

have very little variation in cover (exact threshold under experiment), the spatial cluster analysis will be extended to nearby land cover pixels with an increment of 1 km. A maximum cluster size is set as 10 km × 10 km.

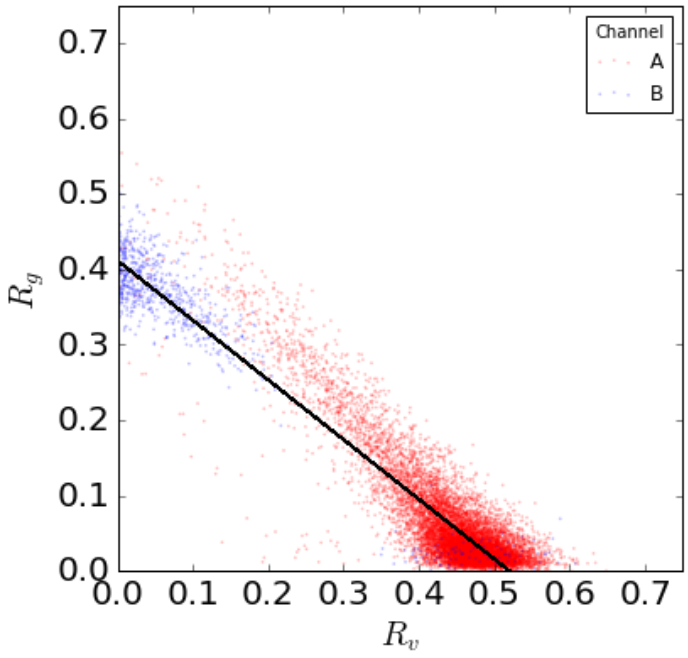

Figure 7. An example of estimating ρ_v/ρ_g in Lope National Park, Gabon using a subset of airborne LVIS data (prototype of GEDI). The slope of an ODR produces an estimate value of 1.3 in this case.

3.4 Ancillary Data Requirements

Specification or estimation of three ancillary parameters are required in the GEDI L2B algorithm, including projection coefficient, clumping index, and empirical lidar canopy/ground reflectance ratio. Their inputs come either directly from empirical model assumptions, or from estimates of external datasets at ecoregion level (see Section 3.2 and 3.3). Other ancillary variables include leaf-on and –off status, percentage of urban built-up area, Landsat and MODIS based tree cover estimates. More details of these ancillary products can be found in the document of GEDI ancillary data product report.

3.5 Error Budget and Uncertainties

The lidar-PAI theoretical model applied in this algorithm has been demonstrated in previous studies (Ni-Meister et al. 2010; Armston et al. 2013; Tang et al. 2012). We therefore do not further validate the model itself in this document, but aim to quantify model uncertainties from the input variables — *Rv*, *Rg*, and ρ_v/ρ_g.

In the GEDI L2B product we report the predicted error of *Rg* fitting in rxwave as well as its associated uncertainty of P_{gap} for each corresponding L2A processing algorithm. Note this predicted error is not equivalent to its measurement error (i.e. relative to an

independent reference) but aims to help algorithm diagnosis as an additional quality flag. In general, there has been no consistent bias in the estimates of R_v and R_g along gradients of canopy cover and topographic slope based on comparisons between pseudo-GEDI products and airborne lidar data (Hancock et al. 2019). However, precisions of the pseudo-GEDI products tend to be lower over a combination of high canopy cover and high slope gradient. The estimates of R_v and R_g achieve the highest accuracy (error less than 10% in Hancock et al. 2019) over forests of moderate canopy cover and on relatively flat ground surface (<15° slope, Tang et al. 2014).

The values of ρ_v/ρ_g are obtained from external ancillary data sets: either a LUT built in pre-launch calibration or an updated estimation from the post-launch module. The LUT can be continuously updated post-launch (see Section 3.3). We expect relatively low uncertainties in ρ_v/ρ_g retrievals using the post-launch module (with quality control of model fitting). Yet relatively larger errors might occur at individual footprint using the ecoregion-level estimates, or over areas where the post-launch module fails possibly due to high level of spatial heteroscedasticity. Uncertainties of ρ_v/ρ_g estimates and their impacts on retrievals of canopy profiles have been assessed in previous lidar studies (Armston 2013 and Tang et al. 2012). In general, there are non-linear error responses in the gradients of canopy cover and LAI. For retrievals of canopy cover, the uncertainty of ρ_v/ρ_g has a stronger impact at the lower end, and almost no impact over extremely dense canopies (e.g. > 90%). It has the opposite impact on LAI with larger uncertainty level at the high end. They will be quantified using spatial statistical analysis at ecoregion level and reported in the final post-launch validation.

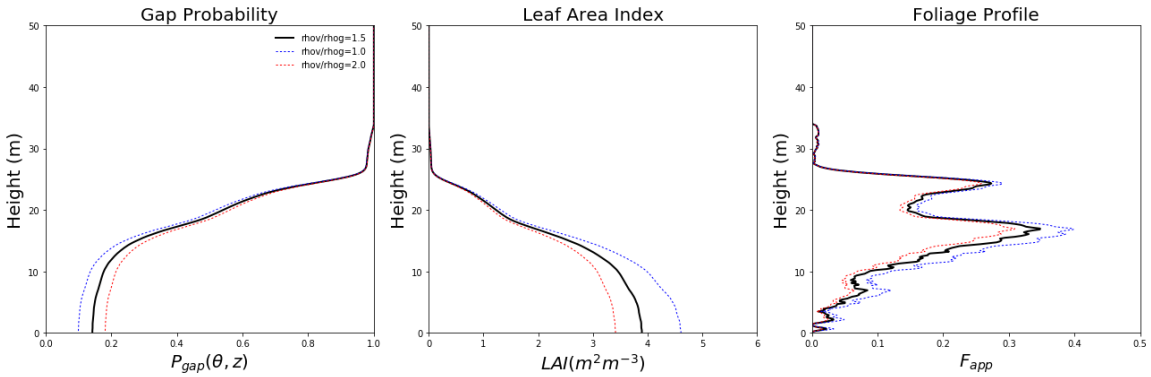

Figure 8. An illustrative example of footprint level GEDI L2 product, including vertical LAI profile and foliage profile. Red-dashed lines along the gap probability distribution profile show an impact on LAI retrievals from ±50% in ρ_v/ρ_g.

3.6 Implementation Considerations

3.6.1 Algorithm Sequence

The main algorithm sequence for generating L2B product:

For each footprint:

1. Read corrected RX-waveform, local beam elevation angle, and geolocation (*lat/lon*) & elevation information from GEDI 1B product.
2. Read elevation of the lowest mode (elev_lowestmode) as the mean ground surface elevation, sample number of the highest and lowest detected returns (toploc and botloc).
3. Fit the Ground component of rxwave (R_g) using an exponentially modified Gaussian model, and output fitting parameters (rg_eg_amplitude, rg_eg_center, rg_eg_sigma and rg_eg_gamma).
4. Read ancillary data from external datasets based on the latitude/longitude of the lowest mode.
5. Calculate output variables using eq. 7~11 for this footprint.
6. Output L2B retrieval status and attach flags from lower level GEDI products.

3.6.2 Quality Control and Diagnostics

Quality control (QC) is an integral part of the GEDI L2B algorithm. The QC steps of the L2B are based on flags inherited from the lower level GEDI products and exception handlers in the GEDI L2B algorithm. The inherited GEDI inputs include GEDI L1B degrade flag, GEDI L1B stale return flag, GEDI 2A waveform sensitivity metric, GEDI L2A RX-wave assessment flag, GEDI 1A leaf-on flag and GEDI 2A simplified quality flag. Outputs of the GEDI L2B QC products include GEDI L2B Retrieval status including the selected setting of L2A algorithm and L2B ground fitting algorithm. The users are recommended to explore the QC flags to re-analyze or subset the complete GEDI records for particular scientific interests.

3.6.3 Latency

Release of GEDI L2B product is consistent with GEDI L1A, L1B and L2A. Data derived from the post-launch module will be continuously updated as new data acquisitions accumulate. An updated GEDI L2B product is expected to be released 6 months after the completion of data acquisition.

4.0 REFERENCES

Aber, J.D., 1979. Foliage-height profiles and succession in northern hardwood forests. *Ecology*, 60(1), pp.18–23.

Abshire, J.B. et al., 2005. Geoscience Laser Altimeter System (GLAS) on the ICESat Mission: On-orbit measurement performance. *Geophysical Research Letters*, 32(21).

Abuelgasim, A.A., Fernandes, R.A. & Leblanc, S.G., 2006. Evaluation of national and global LAI products derived from optical remote sensing instruments over Canada. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), pp.1872–1884.

Aguilar-Amuchastegui, N. & Henebry, G.M., 2007. Assessing sustainability indicators for tropical forests: Spatio-temporal heterogeneity, logging intensity, and dung beetle communities. *Forest Ecology and Management*, 253(1–3), pp.56–67.

Armston, J., 2013. *Assessment of Airborne Lidar for Measuring the Structure of Forests and Woodlands in Queensland, Australia*. Queensland, Australia: The University of Queensland.

Armston, J. et al., 2013. Direct retrieval of canopy gap probability using airborne waveform lidar. *Remote Sensing of Environment*, 134, pp.24–38.

Bergen, K.M. et al., 2009. Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences*, 114(4).

Busch, J. et al., 2015. Reductions in emissions from deforestation from Indonesia's moratorium on new oil palm, timber, and logging concessions. *Proceedings of the National Academy of Sciences of the United States of America*, 112(5), pp.1328–1333.

Calders, K. et al., 2014. Implications of sensor configuration and topography on vertical plant profiles derived from terrestrial LiDAR. *Agricultural and Forest Meteorology*, 194, pp.104–117.

Carey, A. & Wilson, S., 2001. Induced spatial heterogeneity in forest canopies: responses of small mammals. *The Journal of wildlife management*, 65, pp.1014–1027.

Chen, J., Menges, C. & Leblanc, S., 2005. Global mapping of foliage clumping index using multi-angular satellite data. *Remote Sensing of Environment*, 97(4), pp.447–457.

Chen, J.M. et al., 1997. Leaf area index of boreal forests: Theory, techniques, and measurements. *Journal of Geophysical Research*, 102(D24), pp.29429–29443.

Chen, J.M. & Black, T.A., 1992. Defining Leaf-Area Index for Non-Flat Leaves. *Plant Cell and Environment*, 15(4), pp.421–429.

Chen, J.M. & Cihlar, J., 1995. Quantifying the Effect of Canopy Architecture on Optical Measurements of Leaf-Area Index Using 2 Gap Size Analysis-Methods. *IEEE Transactions on Geoscience and Remote Sensing*, 33(3), pp.777–787.

Clawges, R. et al., 2008. The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest. *Remote Sensing of Environment*, 112(5), pp.2064–2073.

Defries, R.S. et al., 2000. A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, 6(2), pp.247–254.

Dubayah, R. et al., 2008. Global vegetation structure from NASA's DESDynI mission: An overview. In *AGU Fall Meeting Abstracts*. p. 1.

Dubayah, R. et al., 1997. The vegetation canopy lidar mission. *Land satellite information in the next decade II: sources and applications*.

Dubayah, R.O. & Drake, J.B., 2000. Lidar Remote Sensing for Forestry. *Journal of Forestry*, 98(6), pp.44–46.

Duncanson, L.I., Niemann, K.O. & Wulder, M.A., 2010. Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sensing of Environment*, 114(1), pp.138–154.

Ellsworth, D. & Reich, P., 1993. Canopy structure and vertical patterns of photosynthesis and related leaf traits in a deciduous forest. *Oecologia*, 96(2), pp.169–178.

Falkowski, M.J. et al., 2008. The influence of conifer forest canopy cover on the accuracy of two individual tree measurement algorithms using lidar data. *Canadian Journal of Remote Sensing*, 34, pp.S338–S350.

Fang, H.L. et al., 2012. Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method. *Remote Sensing of Environment*, 124, pp.610–621.

Fiala, A.C.S., Garman, S.L. & Gray, A.N., 2006. Comparison of five canopy cover estimation techniques in the western Oregon Cascades. *Forest Ecology and Management*, 232(1–3), pp.188–197.

Friedl, M.A. et al., 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114(1), pp.168–182.

Ganguly, S. et al., 2008. Generating vegetation leaf area index earth system data record from multiple sensors. Part 1: Theory. *Remote Sensing of Environment*, 112(12), pp.4333–4343.

Garcia, M. et al., 2012. Characterization of canopy fuels using ICESat/GLAS data. *Remote Sensing of Environment*, 123(0), pp.81–89.

Gardner, S.M. et al., 1995. The influence of habitat structure on arthropod diversity in Argentine semi-arid Chaco forest. *Journal of Vegetation Science*, 6(3), pp.349–356.

Garrigues, S. et al., 2008. Validation and intercomparison of global Leaf Area Index products derived from remote sensing data. *Journal of Geophysical Research-Biogeosciences*, 113(G2), p.

Goetz, S. et al., 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sensing of Environment*, 108(3), pp.254–263.

Goetz, S., 2011. The lost promise of DESDynI. *Remote Sensing of Environment*, 115(11), p.2751.

Goetz, S.J. et al., 2010. Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. *Ecology*, 91(6), pp.1569–1576.

Goetz, S.J. et al., 2015. Measurement and monitoring needs, capabilities and potential for addressing reduced emissions from deforestation and forest degradation under REDD. *Environmental Research Letters*, 10(12).

Goetz, S.J. et al., 2014. The relative importance of climate and vegetation properties on patterns of North American breeding bird species richness. *Environmental Research Letters*, 9(3).

Gower, S.T. & Norman, J.M., 1991. Rapid Estimation of Leaf Area Index in Conifer and Broad-Leaf Plantations. *Ecology*, 72(5), pp.1896–1900.

Hall, F., Saatchi, S. & Dubayah, R., 2011. PREFACE: DESDynI VEG-3D Special Issue. *Remote Sensing of Environment*, 115(11), p.2752.

Hall, F.G. et al., 2011. Characterizing 3D vegetation structure from space: Mission requirements. *Remote Sensing of Environment*, 115(11), pp.2753–2775.

Hancock, S. et al., 2011. A threshold insensitive method for locating the forest canopy top with waveform lidar. *Remote Sensing of Environment*, 115(12), pp.3286–3297.

Hancock, S. et al., 2014. Characterising forest gap fraction with terrestrial lidar and photography: An examination of relative limitations. *Agricultural and Forest Meteorology*, 189, pp.105–114.

Hancock, S. et al., 2012. Measuring forests with dual wavelength lidar: A simulation study over topography. *Agricultural and Forest Meteorology*, 161, pp.123–133.

Hancock, S., et al. (2019) The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. *Earth and Space Science*. 6(2), 294-310.

Hansen, M.C. et al., 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. *Earth Interactions*, 7(10), pp.1–15.

Hansen, M.C. et al., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160), pp.850–853.

Hansen, M.C. et al., 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*, 83(1–2), pp.303–319.

Harding, D., 2001. Laser altimeter canopy height profiles: methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment*, 76(3), pp.283–297.

Harding, D.J., 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, 32(21).

Harris, N.L. et al., 2012. Baseline Map of Carbon Emissions from Deforestation in Tropical Regions. *Science*, 336(6088), pp.1573–1576.

He, L., Chen, J. M., Pisek, J., Schaaf, C. B., & Strahler, A. H. (2012). Global clumping index map derived from the MODIS BRDF product. *Remote Sensing of Environment*, *119*, 118–130.

Hofton, M.A., Minster, J.B. & Blair, J.B., 2000. Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, 38(4), pp.1989–1996.

Huang, Q.Y. et al., 2014. The Influence of Vegetation Height Heterogeneity on Forest and Woodland Bird Species Richness across the United States. *Plos One*, 9(8).

James, F.C. & Wamer, N.O., 1982. Relationships between temperate forest bird communities and vegetation structure. *Ecology*, 63(1), pp.159–171.

Jennings, S.B., Brown, N.D. & Sheil, D., 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. *Forestry*, 72(1), pp.59–73.

Korhonen, L. et al., 2011. Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment*, 115(4), pp.1065–1080.

Lee, S. et al., 2011. Physically based vertical vegetation structure retrieval from ICESat data: Validation using LVIS in White Mountain National Forest, New Hampshire, USA. *Remote Sensing of Environment*, 115(11), pp.2776–2785.

Lefsky, M.A. et al., 2002. Lidar Remote Sensing for Ecosystem Studies. *BioScience*, 52(1), p.19.

Luo, S.Z. et al., 2013. Retrieving leaf area index using ICESat/GLAS full-waveform data. *Remote Sensing Letters*, 4(8), pp.745–753.

MacArthur, R.H. & Horn, H.S., 1969. Foliage profile by vertical measurements. *Ecology*, 50, pp.802–804.

Marselis, S.M., Tang, H**.**, Armston, J.D., Calders, K., Labrière, N., Dubayah, R., (2018). Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case study in Lopé National Park, Gabon. *Remote Sensing of Environment*. 216, 626–634.

Miller, J., 1967. A formula for average foliage density. *Australian Journal of Botany*, 15(1), pp.141–144.

Morisette, J.T. et al., 2006. Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), pp.1804–1817.

Morsdorf, F. et al., 2006. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, 104(1), pp.50–61.

Mu, Q. et al., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, 111(4), pp.519–536.

Myneni, R.. et al., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, 83(1–2), pp.214–231.

Myneni, R.B. et al., 2007. Large seasonal swings in leaf area of Amazon rainforests. *Proceedings of the National Academy of Sciences of the United States of America*, 104(12), pp.4820–4823.

Ni-Meister, W. et al., 2010. Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research*, 115.

Ni-Meister, W., Jupp, D.L.B. & Dubayah, R., 2001. Modeling lidar waveforms in heterogeneous and discrete canopies. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9), pp.1943–1958.

Nilson, T., 1971. A theoretical analysis of the frequency of gaps in plant stands. *Agricultural Meteorology*, 8, pp.25–38.

Nilson, T., 1999. Inversion of gap frequency data in forest stands. *Agricultural and Forest Meteorology*, 98–99(1), pp.437–448.

Pang, Y. et al., 2011. Impact of footprint diameter and off-nadir pointing on the precision of canopy height estimates from spaceborne lidar. *Remote Sensing of Environment*, 115(11), pp.2798–2809.

Parker, G. et al., 2005. Seasonal balance and vertical pattern of photosynthetically active radiation within canopies of a tropical dry deciduous forest ecosystem in Mexico. *Journal of Tropical Ecology*, 21, pp.283–295.

Parker, G., 2004. The canopy surface and stand development: assessing forest canopy structure and complexity with near-surface altimetry. *Forest Ecology and Management*, 189, pp.307–315.

Parker, G.G., Lefsky, M.A. & Harding, D.J., 2001. Light transmittance in forest canopies determined using airborne laser altimetry and in-canopy quantum measurements. *Remote Sensing of Environment*, 76(3), pp.298–309.

Parker, G.G., O'Neill, J.P. & Higman, D., 1989. Vertical Profile and Canopy Organization in a Mixed Deciduous Forest. *Vegetatio*, 85(1/2), pp.1–11.

Pimm, S.L. et al., 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, 344(6187), p.987–+.

Pisek, J., Govind, A., Arndt, S. K., Hocking, D., Wardlaw, T. J., Fang, H., … Longdoz, B. (2015). Intercomparison of clumping index estimates from POLDER, MODIS, and MISR satellite data over reference sites. *ISPRS Journal of Photogrammetry and Remote Sensing*, *101*, 47–56.

Rautiainen, M., Stenberg, P. & Nilson, T., 2005. Estimating canopy cover in Scots pine stands. *Silva Fennica*, 39(1), pp.137–142.

Rose, R.A. et al., 2015. Ten ways remote sensing can contribute to conservation. *Conservation Biology*, 29(2), pp.350–359.

Sexton, J.O. et al., 2016. Conservation policy and the measurement of forests. *Nature Clim. Change*, 6(2), pp.192–196.

Sexton, J.O. et al., 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6(5), pp.427–448.

Stark, S.C. et al., 2012. Amazon forest carbon dynamics predicted by profiles of canopy leaf area and light environment J. Chave, ed. *Ecology Letters*, 15(12), pp.1406–1414.

Strahler, A.H. et al., 2008. Retrieval of forest structural parameters using a ground-based lidar instrument (EchidnaÂ®). *Canadian Journal of Remote Sensing*, 34(SUPPL. 2). 57849121206&partnerID=40&md5=e5f51468ce11a901ebba8097a6a4507e.

Swatantran, A. et al., 2012. Mapping Migratory Bird Prevalence Using Remote Sensing Data Fusion G. J.-P. Schumann, ed. *PLoS ONE*, 7(1), p.e28922.

Tang, H. et al., 2016. Characterizing leaf area index (LAI) and vertical foliage profile (VFP) over the United States. *Biogeosciences*, 13(1), pp.239–252.

Tang, H., Brolly, M., et al., 2014. Deriving and validating Leaf Area Index (LAI) at multiple spatial scales through lidar remote sensing: A case study in Sierra National Forest, CA. *Remote Sensing of Environment*, 143, pp.131–141.

Tang, H., Dubayah, R., et al., 2014. Large-scale retrieval of leaf area index and vertical foliage profile from the spaceborne waveform lidar (GLAS/ICESat). *Remote Sensing of Environment*, 154, pp.8–18.

Tang, H. et al., 2012. Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sensing of Environment*, 124, pp.242–250.

Vierling, K.T. et al., 2008. Lidar: shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment*, 6(2), pp.90–98.

Vose, J. et al., 1995. Vertical leaf area distribution, light transmittance, and application of the Beer-Lambert Law in four mature hardwood stands in the southern Appalachians. *Canadian Journal of Remote Sensing*, 25, pp.1036–1043.

Weiss, M. et al., 2007. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. *Remote Sensing of Environment*, 110(3), pp.317–331.

Wenze Yang et al., 2006. MODIS leaf area index products: from validation to algorithm improvement. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), pp.1885–1898.

Wulder, M.A. et al., 2007. Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. *Remote Sensing of Environment*, 110(1), pp.123–137.

Yang, X.Y. et al., 2013. Three-dimensional forest reconstruction and structural parameter retrievals using a terrestrial full-waveform lidar instrument (Echidn (R)). *Remote Sensing of Environment*, 135, pp.36–51.

Zhao, F. et al., 2011. Measuring effective leaf area index, foliage profile, and stand height in New England forest stands using a full-waveform ground-based lidar. *Remote Sensing of Environment*, 115(11), pp.2954–2964.

Zhao, F. et al., 2012. Measuring gap fraction, element clumping index and LAI in Sierra Forest stands using a full-waveform ground-based lidar. *Remote Sensing of Environment*, 125, pp.73–79.

Zhao, K. et al., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sensing of Environment*, 115(8), pp.1978–1996.

Zhao, M. et al., 2005. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, 95(2), pp.164–176.

GLOSSARY/ACRONYMS

ATBD	Algorithm Theoretical Basis Document
BRDF	Bidirectional Reflectance Distribution Function
DESDynI	Deformation, Ecosystem Structure and Dynamics of Ice
FHD	Foliage Height Diversity
GEDI	Global Ecosystem Dynamics Investigation
GORT	Geometric Optical Radiative Transfer
ICESat	Ice, Cloud and Land Elevation Satellite
IGBP	International Geosphere-Biosphere Programme
LAI	Leaf Area Index
LVIS	Land, Vegetation and Ice Sensor
LUT	Look-up Table
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
ODR	Orthogonal Distance Regression
PAI	Plant Area Index
PAVD	Plant Area Volume Density
QC	Quality Control
TBD	To Be Determined
TLS	Terrestrial Laser Scanning
VCL	Vegetation Canopy Lidar
VIIRS	Visible Infrared Imaging Radiometer Suite