Algorithm Theoretical Basis Document (ATBD)

for

GEDI Transmit and Receive Waveform Processing for L1 and L2 Products

Michelle Hofton[1], J Bryan Blair[2]

Contributions by
Sarah Story[3], Donghui Yi[3]

1. University of Maryland, College Park, MD
2. NASA Goddard Space Flight Center, Greenbelt, MD
3. KBR Greenbelt MD

Version 1.0
Release date: December, 4th, 2019
Goddard Space Flight Center, Greenbelt, MD

Authors:

____

____

Principal Investigator:

____

Abstract

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 ~30 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 fundamental footprint observations made by the GEDI instrument are received waveforms of energy (number of photons) as a function of receive time. When combined with laser pointing and positioning information, these waveforms precisely georeference the 3-dimensional surface within each laser footprint relative to a reference ellipsoid. Within each waveform the vertical distribution of intercepted surfaces is captured (e.g., ground, canopy surfaces, ocean etc.). The relative locations of each reflecting surface within the footprint is identified during postprocessing, then combined with the geolocation of each waveform provided in the L1B product to generate the geolocation product in the L2 group. This ATBD details the algorithms and approaches used to determine various ranging and other metrics within the GEDI waveforms.

Foreword

This document is the Algorithm Theoretical Basis Document for the GEDI Transmit and Receive Waveform Processing for L1 and L2 Products. 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:
Michelle Hofton
University of Maryland, and NASA Goddard Space Flight Center, Code 61A
mhofton@umd.edu
301-286-4488

Change History Log

Revision Level	Description of Change	Date Approved
1.0	Major revision to focus ATBD on waveform analysis congruent with the GEDI data products	Nov 29 2019

Table of Contents

List of Tables

List of Figures

1.0 INTRODUCTION

1.1 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
Level 1	**Geolocated Waveforms**
Level 2	**Canopy Height/Profile Metrics** • **RH metrics** • **Canopy top height** • **Ground elevation** • **Canopy cover and cover profile** • **LAI and LAI profile**
Level 3	**Gridded Footprint Metrics**
Level 4	**Biomass**
Level 4	**Demonstrative Products** • Ecosystem model outputs • Enhanced height/biomass using fusion with Tandem X & Landsat • Habitat model outputs

The GEDI instrument consists of 3 lasers one of which is split, and dithered resulting in a total of 8 beams that are spaced 600 meters apart on the Earth's surface in the cross-track direction. Each beam consists of ~30 m footprint samples approximately spaced every 60 m along track. The fundamental footprint observations made by the GEDI instrument are received waveforms of energy (number of photons) as a function of receive time. The GEDI waveforms are a distance measure of the vertical intercepted surfaces within a footprint. Captured within the receive footprint waveforms are the range to the ground and to various metrics of vegetation or "tree" height.

1.2 GEDI Configuration

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 relative to the flight direction, and approximately 735 m of zonal (parallel to lines of latitude) spacing. Each beam transect consists of ~30 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. 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.

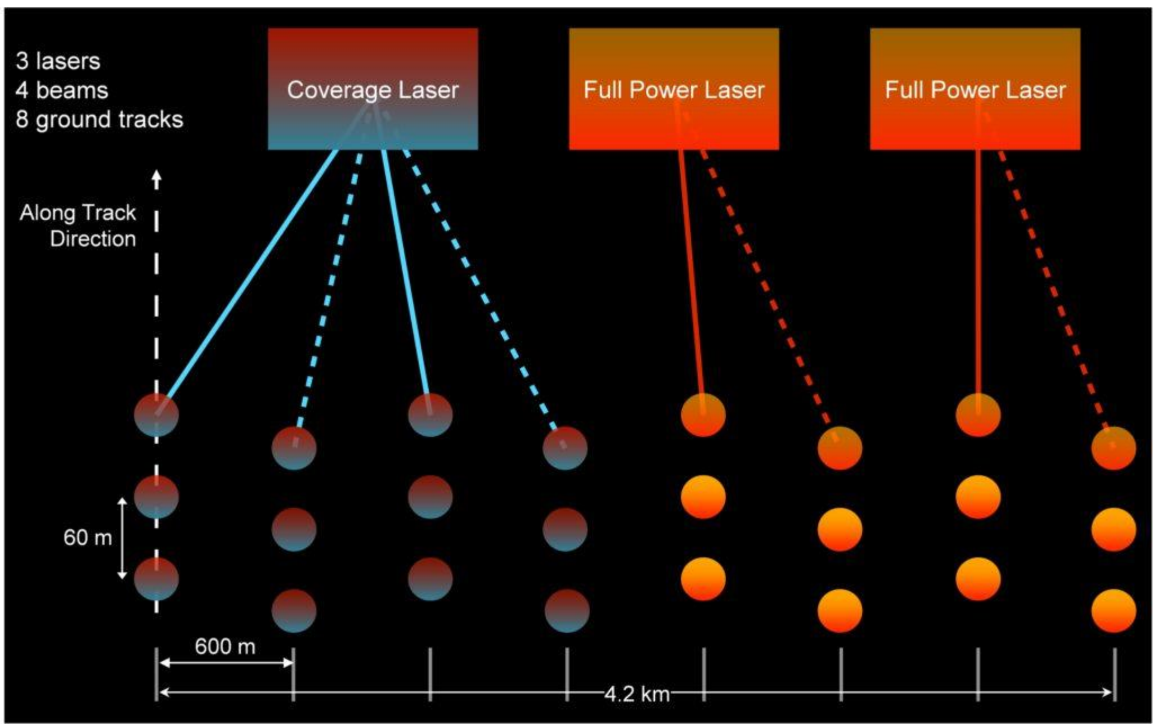

Figure 1. GEDI Beam Ground-track Configuration

1.3 Document Overview and Objective

This document describes the general theoretical overview of the algorithms and procedures from which precise elevation, height, and surface structure metrics are obtained. Waveform processing consists of interpretation of the shapes of both the outgoing and received waveforms, information from which is used in the generation of the L1A, L1B, L2A and L2B products. For example, information derived from the GEDI Transmit waveform (*txwaveform*) is required to produce the L1A and L1B products, and information derived from the GEDI Receive waveforms (*rxwaveforms*) is needed to produce the L2A product. This document provides the general overview of the waveform interpretation algorithms and parameters that are produced, as well as an overview of how these parameters are stored within the L2A product and used to produce geolocated elevation, height and surface metrics also stored in the L2A data product.

This document is arranged in the following manner:

- **Section 1** presents a brief introduction
- **Section 2** presents details of the GEDI Transmit waveform analysis
- **Section 3** presents details of the correction of the GEDI Transmit and Receive waveforms for known artifacts
- **Section 4** presents details of the GEDI Receive waveform analysis
- **Section 5** presents an overview of the geolocation and derivation of the GEDI L2A elevation, height, energy and sensitivity products
- **Section 6** contains references
- An acronym glossary can be found at the end of this document.

1.4 GEDI Waveform Overview

A digitally-recorded return laser pulse, or waveform (Figure 1), represents the time history of the laser pulse as it interacts with the reflecting surfaces. The waveform can have a simple (single-mode) shape similar to that of the outgoing pulse [Fig. 1(a)] or be complex and multimodal with each mode representing a reflection from an apparently-distinct surface within the laser footprint (Figure 1(b)). Simple waveforms are typical in ocean or bare-ground regions and complex waveforms in rough terrain or vegetated regions. The first and last modes (i.e., detected signal above noise) within the waveform are associated with the highest and lowest perceived reflecting surfaces within the footprint respectively.

The waveform processing algorithms described in this ATBD are adapted from methods developed for the analysis of waveforms acquired from NASA's Land, Vegetation and Ice Sensor (LVIS) (Blair et al., 1999). LVIS has a multi-decades long history of acquiring, processing and releasing precise and accurate maps of 3D surface structure over a wide range of surface types.

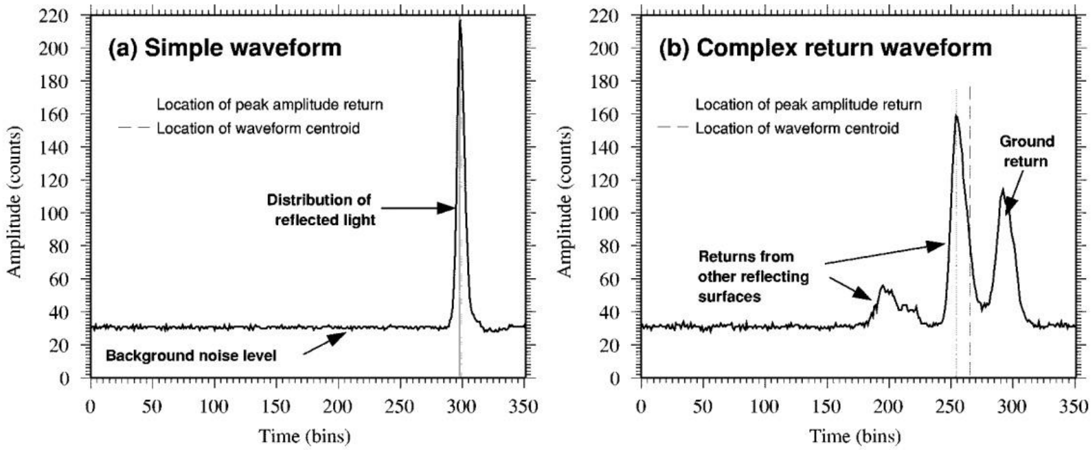

Figure 2. Example laser return waveforms collected using the LVIS laser altimeter from (a) water and (b) vegetation (adapted from Hofton et al., 2000)

1.5 Algorithm Objectives

The waveform analysis algorithms specified in this document are designed to enable the waveform geolocation (the L1B product), the derivation of footprint level GEDI Canopy Elevation and Height Metrics (the L2A product) and the estimation of the Canopy Profile metrics (the L2B product).

1.6 Related Documentation

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

1.6.1 Parent Documents

- GEDI Science Data Management Plan

1.6.2 Applicable Documents

- GEDI ATBD for GEDI Waveform Geolocation 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 GEDI Transmit Waveform Analysis and Generation of the Range Vector for input to the L1B

2.1 Outline of the Procedure

The transmitted pulse can be represented by a single Gaussian with a baseline at the mean noise level. The characteristics of the TX pulse are calculated along with the parameters from a single peak Gaussian fit. The parameters from the single peak Gaussian fit are used to generate timing (range) vectors between the transmit and receive waveform windows and are input into the L1B geolocation. An extended Gaussian is also fit to the TX pulse to provide information that is used to produce some of the L2B data products. Note the laser transmit pulses themselves are not perfectly gaussian.

The processing steps are summarized below.

- Characterize the transmit pulse and perform any needed adjustments, calculate:
  - maximum and minimum amplitudes above the mean noise
  - check for saturation and ringing
  - check for digitizer artifacts
  - energy (area under pulse)
  - calculated noise standard deviation

- Fit the transmit waveform with a Gaussian to derive:
  - amplitude and associated error of Gaussian approximating the TX
  - width and associated error of Gaussian approximating the TX
  - center and associated error location of Gaussian approximating the TX
  - goodness of fit of the Gaussian approximating the TX
  - flags indicating if the fit was unsuccessful

- Fit the transmit waveform with an extended Gaussian to derive:
  - amplitude and associated error of extended Gaussian approximating the TX
  - width and associated error of extended Gaussian approximating the TX
  - center and associated error location of extended Gaussian approximating the TX
  - goodness of fit of the extended Gaussian approximating the TX
  - flags indicating if the fit was unsuccessful
- Generate timing (range) vectors between the transmit and receive waveform windows for input into the L1B geolocation module.

Parameters are output to the “TX_PROCESSING” subgroup of the L1A data product. A subset of parameters are contained in the root group of the L1B data product.

2.2 Transmit Waveform Characterization

2.2.1 Waveform Maximum and Minimum Amplitudes

Parameters: Tx_minamp and *tx_maxamp*

Measure the minimum and maximum amplitudes of the waveform relative to the mean noise level. Required inputs are *txwaveform* and *noise_mean*

2.2.2 Transmit pulse peak amplitude location

Parameter: tx_peakloc

The temporal location relative to bin0 of the txwaveform at which the maximum amplitude return occurs (Figure 3).

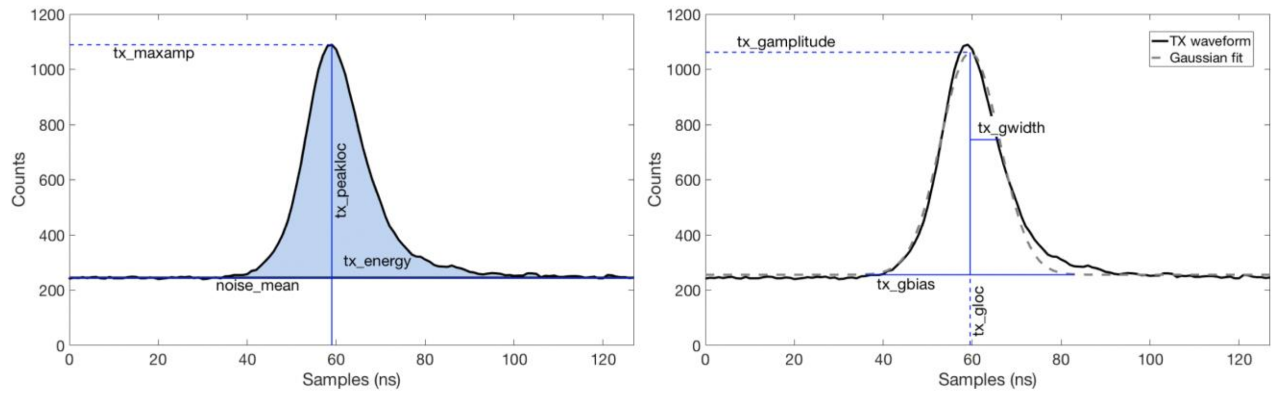

Figure 3. Schematic overview of parameters associated with the transmit waveform processing.

2.2.3 Transmit Pulse Energy

Parameter: Tx_energy

Tx pulse energy is estimated by computing the integrated area of the signal relative to the mean noise level. It's computed by summing up the waveform amplitudes after subtracting the mean noise value. Required inputs are *txwaveform* and *noise_mean*.

2.2.4 Waveform background noise standard deviation

Parameter: tx_sd_nw

The waveform noise standard deviation value stored in the L0 data packet has been rounded to an integer, therefore a more precise noise standard deviation estimate is calculated from:

Tx_sd_nw=sqrt(sum_of_squares/1024.).

Required input is *sum_of_squares.*

2.2.5 Saturation and Digitizer Artifacts

Parameters: tx_satflag, tx_cntsat, tx_ampflag

When the signal levels exceed a certain intensity, the detector can no longer accurately represent the return signal and begins to act non-linearly (e.g., Sun et al., 2017), referred to as "saturation". We return several parameters to define this condition, although we do not expect it to occur in the txwaveform record.

Tx_cntsat: The number of consecutive txwaveform samples with amplitude greater than value defined in *ancillary/saturated_amp*. Required inputs are *txwaveform* and a limit value.

Tx_satflag: indicates txwaveform amplitudes may be clipped. Value is set to 1 if *tx_cnt sat* > 0.

Tx_ampflag: flag if the laser pulse amplitude recorded in the txwaveform falls outside a recommended range. Set to 1 if either of these conditions occur:

tx_maxamp < mean_noise+ancillary/tx_ampbounds_ll
tx_maxamp > 2l^12-ancillary/tx_amp_bounds_ul

2.2.6 Ringing

Parameters: Tx_ringflag
Ringing is a non-linear phase response in the detector and can create a secondary pulse that is not real. Tx waveforms affected by ringing are flagged. The detection of ringing is based on both max and min amplitudes exceeding thresholds. Txwaveforms are flagged as containing ringing if the following condition is met:

*tx_minamp < tx_sd_nw*ancillary/tx_ringthresh.*

A value of 0 indicates no ringing was detected.

2.2.7 Pulse flag

Parameters: Tx_pulseflag
The presence of a laser pulse within the txwaveform data record is identified if the following condtion is met:

*tx_maxamp > mean_noise+tx_sd_nw*ancillary/tx_pulsethresh*

A value of 0 indicates no pulse was detected.

2.2.8 Average pulse

Parameters: short_term_av_tx_waveform
An average laser pulse is generated for each laser turn on period for monitoring purposes.

While the laser is powering up, the pulse shape changes before settling down into its steady state and pulses are only averaged if the laser has been firing (in *ancillary/good_laser_mode*) for *ancillary/min_laser_on_time_seconds* (default 900) seconds. Each laser-on interval then begins at a time specified by *short_term_av_start_time_int* and *short_term_av_start_time_frac* and ends at a time specified by *short_term_av_end_time_int* and *short_term_av_end_time_frac*, and the averaged waveform for each interval is stored in *short_term_av_tx_waveform*.

To create an average TX waveform, we align TX pulses as closely as possible to minimize pulse broadening. Coarse alignment is achieved by shifting the TX pulse so that the maximum of the waveform is aligned with *ancillary/centerbin*. Then the waveform is spline interpolated to *ancillary/resolution* for the two bins around the peak, the interpolated peak is aligned with *ancillary/centerbin*, and the waveform is resampled at a 1 ns rate but so that the interpolated peak falls on a whole bin. The resampled waveform does not necessarily have a value at every time, since the center bin of the averaged waveform is shifted relative to the peak of the TX waveform. When we average the resampled waveforms, we average only the valid values at each bin. Some bins will have more or fewer samples than others. A bin with no samples (possible at the ends of the waveform if the average TX peak location is significantly shifted from *ancillary/centerbin*) is assigned a value of NaN.

2.3 Transmit Waveform Interpretation: Least Squares Gaussian Fitting

2.3.1 Overview

The location of the center of the transmit pulse relative to the start of the laser transmit window is calculated by fitting a Gaussian to a smoothed version of the transmit waveform record after subtracting off the mean noise (Figure 3). We use the MPFIT code (Markwardt, 2009) to fit a 1D gaussian plus bias (constant) to each txwaveform. MPFIT is a robust non-linear least square curve fitting package (http://purl.com/net/mpfit) using the Levenberg-Marquardt technique to solve the least squares function. We solve for location, amplitude, width and constant terms, and use upper and lower bounding constraints on each parameter. Fit values returned by mpfit for each laser txwaveform are stored in the "TX_PROCESSING" subgroup of the L1A data products. Parameters used to initialize the fitting are retained in the TX_PROCESSING/ancillary subgroup of the L1A data product. A subset of these parameters needed for lower level products are repeated in the root group of the L1B data products.

Smoothing of the transmit waveform is performed by convolving it with a gaussian of width *ancillary/tx_smoothwidth*.

The function fit to each transmit pulse is given by:

$$F(x)=Ae^{-(x-b)^2/(2\sigma^2)}+d$$

where A is the amplitude, b is center, σ is the width and d is a constant/bias.

2.3.2 Gaussian fit parameters

Parameters: Tx_gamplitude, tx_gamplitude_error

Amplitude and associated error of gaussian fit to the txwaveform. The estimation of these parameters was constrained using:

ancillary/tx_constraint_gamplitude_lower and *ancillary/tx_constraint_gamplitude_upper*

Parameters: Tx_gloc, tx_gloc_error
Location and associated error of gaussian fit to the txwaveform. The estimation of these parameters was constrained using:
ancillary/tx_constraint_gloc_lower and *ancillary/tx_constraint_gloc_upper*. This parameter is used in subsequent geolocation for the L1B and L2 data products.

Parameters: Tx_gwidth, tx_gwidth_error
Width and associated error of gaussian fit to the txwaveform. The estimation of these parameters was constrained using:
ancillary/tx_constraint_gwidth_lower and *ancillary/tx_constraint_gwidth_upper*.

Parameters: Tx_gbias, tx_gbias_error
Constant term and associated error of the gaussian fit to the txwaveform.

Parameters: tx_gchisq
Total chi squared of the fit

Parameters: tx_giters
Number of iterations to converge gaussian fit to the txwaveform

Parameters: tx_gflag
Gaussian status fit flag: 1=convergence in chi2 value, 2=convergence in parameter value, 3=convergence in chi2 and parameter values, 4=convergence in orthogonality, 5=maximum number of iterations reached, 6=ftol too small (no further improvement), 7=xtol too small (no further improvement), 8=gtol too small (no further improvement).

2.4 Transmit Waveform Interpretation: Least Squares Extended Gaussian Fitting

2.4.1 Overview

An extended gaussian is fit to the txwaveform for use by the L2B data product. We use the MPFIT code to fit the function for a modified gaussian given by:

$$f(x;A,\mu,\sigma,\gamma)=A\gamma 2\exp[\gamma(\mu-x+\gamma\sigma^2/2)]\mathrm{erfc}(\mu+\gamma\sigma^2-x)/(\sqrt{2}\sigma))$$

Four Parameters are amplitude (A), center (μ), sigma (σ), and gamma (γ), and erfc() is the complementary error function. Parameters used to initialize the extended gaussian fitting and fitted parmameter values are stored in the L1A and L1B data products.

2.4.2 Extended Gaussian Fit Parameters

Parameters: Tx_egamplitude, tx_egamplitude_error

Amplitude and associated error of gaussian fit to the txwaveform. The estimation of these parameters was constrained using:
ancillary/tx_constraint_egamplitude_lower and *ancillary/tx_constraint_egamplitude_upper*

Parameters: Tx_egcenter, tx_egcenter_error
Center and associated error of gaussian fit to the txwaveform. The estimation of these parameters was constrained using:
ancillary/tx_constraint_egcenter_lower and *ancillary/tx_constraint_egcenter_upper*.

Parameters: Tx_eggamma, tx_eggamma_error
Gamma and associated error of extended gaussian fit to the txwaveform. The estimation of these parameters was constrained using:
ancillary/tx_constraint_eggamma_lower and *ancillary/tx_constraint_ eggamma_upper*.

Parameters: Tx_egsigma, tx_egsigma_error
Sigma term and associated error of the extended gaussian fit to the txwaveform. The estimation of these parameters was constrained using:
ancillary/tx_constraint_ egsigma_lower and *ancillary/tx_constraint_ egsigma_upper*.

Parameter: Tx_egchisq
Total chi squared of the fit

Parameter: tx_egiters
Number of iterations to converge the extended gaussian fit to the txwaveform

Parameter: tx_egflag
Gaussian status fit flag: 1=convergence in chi2 value, 2=convergence in parameter value, 3=convergence in chi2 and parameter values, 4=convergence in orthogonality, 5=maximum number of iterations reached, 6=ftol too small (no further improvement), 7=xtol too small (no further improvement), 8=gtol too small (no further improvement).

2.5 Calculating Window Ranges for Subsequent Geolocation

For the laser return waveform geolocation (Level L1B), the range (in Digitizer samples) between the TX pulse and the recorded RX window is computed by:

- fitting a gaussian to the TX pulse and computing the offset of the pulse within the TX window
- subtracting this from the Bin # of the first and last samples of the RXwaveform window.

Range in digitzer samples is computed assuming:

$$Range\_bin0 = (rx\_open + rx\_offset) - (tx\_open + tx\_offset + tx\_gloc) + 4$$

$$Range\_lastbin = (rx\_open + rx\_offset + rx\_sample\_count - 1) - (tx\_open + tx\_offset + tx\_gloc) + 4$$

Full details can be found in the “GEDI Waveform Geolocation for L1 and L2 Products” ATBD.

2.6 Required Inputs

The required inputs to the TX waveform analysis procedure are available from the L1A data product and are listed in Table 2 below.

Table 2. Input parameters to the GEDI transmit waveform analysis.

Parameter	Description
txwaveform	Tx waveform
noise_mean	Mean of background noise signal (realtime algorithm)
Sum_of_squares	Standard deviation of background noise signal (realtime algorithm)

2.7 Summary of Parameters Output by the Transmit Waveform Processing

Table 3 lists the parameters computed by the transmit waveform processing. These parameters are contained in the subgroup "TX_PROCESSING" in the L1A data product. Some are repeated in the root group in the L1B data product.

Table 3. Parameters output from the GEDI transmit waveform analysis

Parameter	Description	Target
tx_ampflag	Flag: amplitude out of bounds	L1A
tx_cntsat	Number of saturated counts	L1A
tx_energy	TX waveform energy	L1A
tx_maxamp	Maximum TX waveform amplitude	L1A
tx_minamp	Minimum TX waveform amplitude	L1A
tx_peakloc	Location of TX waveform peak	L1A
tx_pulseflag	Flag: indicates TX window contains a pulse	L1A
tx_ringflag	Flag: indicates possible presence of detector ringing in TX waveform	L1A
tx_satflag	Flag: indicates TX waveform contains saturated counts	L1A
tx_sd_nw	Standard deviation estimate to use for TX processing	L1A
short_term_av_end_time_frac	Fractional part of end times used to calculate average TX pulses	L1A
short_term_av_end_time_int	Integer part of end times used to calculate average TX pulses	L1A
short_term_av_start_time_frac	Fractional part of start times used to calculate average TX pulses	L1A

short_term_av_start_time_int	Integer part of start times used to calculate average TX pulses	L1A
short_term_av_tx_waveform	Average TX pulses for each laser on interval	L1A
tx_gamplitude	Gaussian amplitude	L1A
tx_gamplitude_error	Fit error in Gaussian amplitude	L1A
tx_gbias	Gaussian bias	L1A
tx_gbias_error	Fit error in Gaussian bias	L1A
tx_gchisq	Gaussian chi squared	L1A
tx_gflag	Flag: indicates Gaussian fit success or failure	L1A
tx_giters	Iterations for Gaussian fit to converge	L1A
tx_gloc	Gaussian center	L1A, L1B
tx_gloc_error	Fit error in Gaussian center	L1A, L1B
tx_gwidth	Gaussian width	L1A
tx_gwidth_error	Fit error in Gaussian width	L1A
tx_egamplitude	Extended Gaussian amplitude	L1A, L1B
tx_egamplitude_error	Fit error in extended Gaussian amplitude	L1A, L1B
tx_egbias	Extended Gaussian bias	L1A, L1B
tx_egbias_error	Fit error in extended Gaussian bias	L1A, L1B
tx_egcenter	Extended Gaussian center	L1A, L1B
tx_egcenter_error	Fit error in extended Gaussian center	L1A, L1B
tx_egchisq	Extended Gaussian chi squared	L1A, L1B
tx_egflag	Flag: indicates extended Gaussian fit success or failure	L1A, L1B
tx_eggamma	Extended Gaussian gamma	L1A, L1B
tx_eggamma_error	Fit error in extended Gaussian gamma	L1A, L1B
tx_egiters	Iterations for extended Gaussian fit to converge	L1A, L1B
tx_egsigma	Extended Gaussian sigma	L1A, L1B
tx_egsigma_error	Fit error in extended Gaussian sigma	L1A, L1B

3.0 Correcting Telemetered Transmit and Receive Waveforms for Known Artifacts

3.1 Outline of the Procedure

The raw waveforms contain sampling artifacts (quantization, odd-even sampler offsets) and both electronic and optical background noise. To minimize the impact of these we low pass filter the digital waveforms by convolution with a gaussian pulse. This reduces high frequency noise and effectively eliminates ADC sampling errors. The width of the gaussian pulse used for the convolution determines the amount of noise that is removed, but also loss of signal bandwidth (and the ability to detect multiple distinct pulses in the same waveform). Although the manufacturer reports ADC sampler offsets occur only on an odd-even pattern, differences in background noise levels are observed on a 4 bin pattern, with real time calibrations not fully addressing the issues. We believe the corrections made to the raw waveforms will minimize the post processing required by end users while not negatively affecting the information contained in the waveform. However, end users wanting access to the raw waveform can request the L1A data product.

The convolution of the waveforms with a gaussian minimizes sampling artifacts in the pulse products but we must also correct the telemetered noise standard deviation values. These are recorded at the sensor and and must be corrected.

3.2 Approach

3.2.1 Mean noise level, *mean*

No corrections are currently applied to the mean noise level; the value from the telemetered datapacket is recorded in the L1B *mean* parameter.

3.2.2 Noise standard deviation, *sd_corrected*

Odd-even sample differences increase the recorded noise standard deviation value. Estimates of the noise mean in both the odd and even sample bins are recorded in the instrument housekeeping subgroup of the L1A data product (recorded at 1sec intervals by the sensor). To correct the telemetered noise standard deviation value (tx_nw_sd) we first derive an approximation to the long term odd-even sample offsets by:

- Average odd-even means over a time window set of 20,000 samples (corresponding to 2.3 minutes of data)
    - If there is a large gap in the housekeeping data such that a time window contains no non-repeated points, the mean and standard deviation from the last window that contained valid points is carried forward
- Interpolate the averaged values to produce a smooth approximating curve
    - Any gaps in the housekeeping data are filled with random numbers drawn from a normal distribution with the approximate mean/standard deviation.
- Smooth the averages by convolving with a gaussian of width 30s to generate array *gsmoe*.

The corrected noise standard deviation is calculated from:

*Sd_corrected=sqrt(sum_of_squares/1024.0-0.25*gsmoe^2)*

This is approximately what would be expected from convolving two distributions with different standard deviations and calculating their combined standard deviation. It has been validated by taking GEDI waveforms without significant odd-even differences and introducing odd-even differences, calculating the new standard deviation, and then fitting a curve to the standard deviation vs odd-even difference.

3.2.3 Corrected transmit and receive waveforms, *Txwaveform* and *rxwaveform*

We convolve each txwaveform and rxwaveform with a gaussian of width 2 ns (1 sigma) to reduce high frequency noise and eliminate ADC artifacts. No significant loss in precision has been detected using this approach vs. applying a long-wavelength estimate of the odd-even sample differences.

3.3 Required Inputs

The required inputs to the transmit and receive waveform correction are contained in the L1A data product.

3.4 Summary of Output Parameters

Table 4 lists the parameters that are output by the transmit and receive waveform correction algorithm.

Table 4. Output parameters stored in the root group of the L1B data product.

Parameter	Description	Target
rxwaveform	Telemetered rxwaveform smoothed by a Gaussian pulse of width 2 in order to remove differences in background noise levels	L1B
txwaveform	Telemetered txwaveform smoothed by a Gaussian pulse of width 2 in order to remove differences in background noise levels	L1B
mean	Background noise mean	L1B
Sd_corrected	Background noise standard deviation	L1B

4.0 GEDI Receive Waveform Analysis

4.1 Outline of the Procedure

The processing steps to analyze the GEDI receive waveform are as follows:

- Characterize the received pulse (e.g., minimum and maximum amplitudes, energy). Parameters are output into the "rx_assess" sub group of the L2A data product.
- Perform fitting of a single gaussian function to each waveform. Parameters are output into the "rx_1gaussfit" subgroup of the L2A data product
- Run waveform interpretation algorithm to extract ranging and energy parameters within a rxwaveform. The algorithm is run multiple times each with unique settings to provide a

series of possible outcomes that likely cover the range of observation conditions experienced. The results from each algorithm run are output into the “rx_processing_a<n>” subgroups of the L2A data product, where “n” denotes the algorithm.

The receive waveform that is telemetered from GEDI is based on the real time detection algorithm and does not have optimal sensitivity. However, the telemetered waveform contains additional data above and below the detected signals to allow us to use more sensitive and more sophisticated post processing to detect weak signals. The post-processing signal detection algorithm searches this buffer region, which is of variable size based on the data collection mode of the sensor (“Land”: baselined at 300 and 400 samples above and below the first and last detected locations determined by the real time algorithm; “Ocean”: 200 samples both above and below the first and last detected locations determined by the real time algorithm). Output of each algorithm run is saved into the “rx_processing_a<n>” subgroup in the L2A data product, where <n> indicates a set of algorithm parameters used to detect signals. Externally-set parameters are stored in the ancillary subgroups. All GEDI footprints are processed. The parameters used in the L2 receive waveform processing are obtained from the L1B data product. Information from the rx_assess and rx_processing_a<n> subgroups are combined with the L1B geolocation group to product the geolocation subgroup in the L2A.

4.2 Receive Waveform Characterization

4.2.1 Precise noise mean and noise standard deviation for each rxwaveform

Parameters: mean and sd_corrected
No additional refinement of noise statistics is currently performed, these are set to the values measured in real time by the sensor and are contained in the L1B data product.

4.2.2 Maximum and minimum amplitudes of the waveform

Parameters: rx_minamp and rx_maxamp
Measure the peak maximum and minimum amplitudes of the rxwaveform relative to the mean noise level.

4.2.3 Location of maximum amplitude return within waveform

Parameter: rx_maxpeakloc
Sample number of the maximum amplitude return (relative to bin0 of the rxwaveform).

4.2.4 Waveform amplitude clipping

Parameters: rx_clipbin_count, rx_clipbinnumber, and rx_clipbin0
When the signal levels exceed a certain intensity, the detector can no longer accurately represent the return signal and begins to act non-linearly (e.g., Sun et al., 2017), referred to as “saturation”. Because the detector does not accurately represent the return photon flux, information is distorted or lost. If this occurs in the vicinity or on the ground return then it will degrade the precision of

the ground elevation and canopy height measurements. If it occurs anywhere in the waveform, other canopy metrics will be affected. We flag if the recorded waveform contains saturated intensities, the number of consecutive bins affected and the location of the first saturated bin. Externally-set—thresholds are used to define saturation amplitudes, and are stored in *ancillary/rx_clipamp*. Saturated returns are rare.

4.2.5 Waveform total energy

Parameter: rx_energy
Rx pulse energy is estimated by computing the integrated area of the signal relative to the mean noise level. It's computed by summing up the waveform amplitudes after subtracting the mean noise value.

4.2.6 Mean signal value within the 10k range window

Parameter: mean-64kadjusted
Average amplitude within 10km search window with energy from rxwaveform removed:

*mean_64kadjusted=(all_samples_sum-total(rxwaveform)) / (64.*1024.-(rx_sample_count))*

4.2.7 Laser shot used in measurement model calculations

Parameter: ocean_calibration_shot_flag
We provide a flag to indicate if the return waveform was used in the measurement model calculations over the ocean (see L1B ATBD).

4.2.8 Waveform Fidelity Flag

Parameter: rx_assess_flag
This is a bitfield of different flags with each bit indicating whether a condition is present/affected the real-time collection of the waveform.

Bit	Short_name	Description
1	*rx_rxwindow_limit*	indicates rx_sample_count=1420 (potential clipping of rxwaveform vertical extent)
2	*rx_rxwindow_exist*	indicates rx_sample_count =0
3	*rx_rxwindow_clip_front*	signal exceeding real-time threshold (*TH_LEFT_USED*) detected in bin0 of the rxwaveform
4	*rx_rxwindow_clip_back*	signal exceeding real time threshold (*TH_LEFT_USED*) detected in bin *rx_sample_count* of the rxwaveform
5	*rx_ringflag*	ringing detected (*rx_minamp* < *sd_corrected*ancillary/rx_ringthresh*)
6	*rx_rangewindow_clip_front*	rxwaveform is located at the top of the 10km real-time search window (*rx_offset*=0)

7	*rx_rangewindow_clip_back*	rxwaveform is located at the bottom of the 10km real-time search window (*rx_offset+rx_sample_count = 65535)*
8	*rx_pulseflag*	pulse not present (*rx_maxamp < sd_corrected*rx_pulsethresh*)
9	*rx_1binwaveform_flag*	rxwaveform is 1 sample long
10	*rx_ampflag*	max amp is in a zone deemed potentially non-optimal (too weak or to strong) (*rx_maxamp > rx_ampbounds_ll and rx_maxamp < 2l^12-mean-rx_ampbounds_ul)*
11	*rx_clipflag*	Indicates max amp exceeds clip level (rx_*maxamp > rx_clipamp*)

4.2.9 Waveform Quality flag

Parameter: quality_flag

Indicates the waveform can be considered usable for downsteam analysis {1:yes; 0: no). This combines together a set of conditions to indicate the overall validity of a waveform for measuring surface structure. Pseudocode representing conditions that are combined to represent "good" conditions is:

```
stale_return_flag == 0
rx_rangewindow_clip_front == 0
rx_rangewindow_clip_back == 0
rx_clipflag == 0
rx_rxwindow_limit == 0
rx_rxwindow_exist != 0
rx_rxwindow_clip_front == 0
rx_rxwindow_clip_back == 0
rx_1binwaveform_flag == 0
rx_pulseflag != 0
```

4.2.10 Fit single Gaussian to received pulse

A similar procedure as used on the TX pulse is applied to the received pulse. Although the fitting of a single Gaussian to a multi-mode return (e.g., Fig 1b) does not generate precise-enough timing information for GEDI purposes, the results of the algorithm can provide context on the complexity/nature of the return pulse (e.g., simple vs. complex surfaces). The single Gaussian fit results are also used as input to the geolocation and measurement model calculations to provide precise ranges during the ocean sweeps/other calibration maneuvers. Parameters that are derived are:

- amplitude of Gaussian approximating the RX
- width of Gaussian approximating the RX
- center location of Gaussian approximating the RX

- energy of the Gaussian approximating the RX
- goodness of fit of the Gaussian approximating the RX
- flags indicating if the fit was unsuccessful
- flag indicating fit ended without meeting convergence criteria

Parameters are stored in the rx_1gaussfit subgroup of the L2A data product.

Parameters: rx_gamplitude, rx_gamplitude_error
Amplitude and associated error of gaussian fit to the rxwaveform. The estimation of these parameters was constrained using:
ancillary/rx_constraint_gamplitude_lower and *ancillary/rx_constraint_gamplitude_upper*

Parameters: rx_gloc, rx_gloc_error
Location and associated error of gaussian fit to the rxwaveform. The estimation of these parameters was constrained using:
ancillary/rx_constraint_gloc_lower and *ancillary/rx_constraint_gloc_upper*.

Parameters: rx_gwidth, rx_gwidth_error
Width and associated error of gaussian fit to the rxwaveform. The estimation of these parameters was constrained using:
ancillary/rx_constraint_gwidth_lower and *ancillary/rx_constraint_gwidth_upper*.

Parameters: rx_gbias, rx_gbias_error
Constant term and associated error of the gaussian fit to the rxwaveform.

Parameters: rx_gchisq
Total chi squared of the fit

Parameters: rx_giters
Number of iterations to converge gaussian fit to the rxwaveform

Parameters: rx_gflag
Gaussian status fit flag: 1=convergence in chi2 value, 2=convergence in parameter value, 3=convergence in chi2 and parameter values, 4=convergence in orthogonality, 5=maximum number of iterations reached, 6=ftol too small (no further improvement), 7=xtol too small (no further improvement), 8=gtol too small (no further improvement).

4.3 Receive Waveform Interpretation

4.3.1 Overview

The purpose of the waveform interpretation algorithm is to derive timing points to which subsequent data products are referenced. The steps involved in interpretation of the RX pulse are:

- smooth the L1B “corrected” waveform

- using appropriate algorithm threshold settings, front and back, run interpretation algorithms
- in concert with L2A-geolocation, down select algorithm results.

Externally-set parameters are saved in the rx_processing_a<n>/ancillary subgroup of the L2A data product. Output of the receive waveform interpretation is saved into the rx_processing_a<n> subgroup of the L2A data product. All receive waveforms are processed. All input data for the L2A processing are contained in the L1B data product.

The interpretation algorithm is adapted from that used on LVIS data since early 2000's (Blair et al., 1999; LVIS, 2019). The steps involved are:

- Establish threshold and smoothing settings to be applied during signal search (Table 5)
- Determine rough extent of signal within telemetered rxwaveform record using a pre-processing threshold setting (*ancillary/preprocessor_threshold*)
- Establish area around signal to perform weak signal search (*ancillary/searchsize*)
- De-noise (smooth) waveform by convolution with a Gaussian function (*ancillary/smoothing_width_locs* and *smoothing_width_zcross*)
- Apply algorithm to identify surface ranging points and ancillary information (e.g., max amplitude, energy)
- Apply algorithm to generate integrated waveform and extract ranging points.

4.3.2 Establish range of waveform de-noising (smoothing) settings

Parameters: smoothwidth, smoothwidth_zcross

The waveform is smoothed to minimize the effects of noise artifacts and enable detection of weak signals. We smooth the RX waveform by convolving with a Gaussian filter of various widths. The algorithm smooths the waveform in 2 steps. The first step smooths the noise portion of the signal to facilitate the detection of weak signal within those areas. This Gaussian filter width (*smooth_width*) is chosen to broadly match the width of the laser transmit pulse. We assume that the noise is essentially gaussian and when searching noise-portions of the waveform for weak signals, that the noise will average to zero over several samples whereas signal will not. Thus, the *smooth_width* Gaussian filter width setting is intended to decrease the noise whilst not decreasing the signal. The second smoothing step is designed to search the signal-only portion of the waveform for reflections (*smooth_width_zcross*). Note that the L2A parameters *smoothwidth* and *smoothwidth_zcross* are identical to *ancillary/smoothing_width_locs* and *ancillary/smoothing_width_zcross* but are recorded for every laser shot.

4.3.3 Establish range of threshold settings

Parameters: front_threshold, back_threshold

A false alarm is when the algorithm detects noise instead of signal, either above the canopy top or below the ground. False alarms can result in a positive bias in the canopy height estimate or a negative bias in the ground elevation estimate. When performing signal detection in the presence of noise, a balance must be struck between detection sensitivity and false alarm rate. Setting the threshold too low can cause a high false alarm rate. Setting the threshold too high can result in missed ground returns. The threshold level is determined by the mean noise level, the noise standard deviation after smoothing, the detection criteria, the extent of the search window, and the false alarm rate. The extent of the search window (from the real time detection location to the end of the RX window) is set in real-time GEDI operations (400 samples, land mode; 200 samples, ocean mode). A wider search window will require a higher, and less sensitive threshold to maintain a given false alarm rate. A narrower search window can use a lower and more sensitive threshold but risks not including the ground return. Therefore, the search window has been minimized when possible. The threshold formula is:

threshold = *mean* + x * *sd_corrected*

where x is given by a predetermined multiplier and stored in the *rx_processing_a<n>/ancillary* subgroup as *back_threshold* or *front_threshold*. (note: back refers to the section of the waveform below the lowest signal, and front to the section above the highest signal)

During initial data checkout and preliminary cal-val activities, we have worked to identify a series of settings for both the smoothing and threshold parameters that provide end users with precise mode detection in a variety of cases (e.g., nighttime, daytime, high and low energy lowest modes). Table 5 gives the settings for the algorithm runs contained in the L2A data product. Note that these settings can also be found in the rx_processing_a<n>/ancillary subgroups.

Table 5. Threshold and smoothing settings used to interpret the receive waveform. Six distinct sets of values are used. Output from each algorithm run is stored in the corresponding rx_processing_a<n>/ancillary subgroup. σ = sd_corrected.

Algorithm	**Rx_processing Subgroup**	**Smooth width**	**Smoothwidth_ zcross**	**Front_ threshold**	**Back_ threshold**
1	a1	6.5	6.5	3σ	6σ
2	a2	6.5	3.5	3σ	3σ
3	a3	6.5	3.5	3σ	6σ
4	a4	6.5	6.5	6σ	6σ
5	a5	6.5	3.5	3σ	2σ
6	a6	6.5	3.5	3σ	4σ

4.3.4 Determine signal extent within waveform

Parameters: search_start, search_end

We define the portion of the waveform to be searched for reflected signal based on the section of the waveform where intensity exceeds a threshold given by:

*mean+sd_corrected*ancillary/preprocessor_threshold*

This region is then extended immediately above and below by the number of samples set in *ancillary/searchsize*, modified to not exceed waveform record length as necessary. *Search_start* and *search_end* record the waveform sample numbers between which the algorithm will search for reflected signal (Figure 4),

Parameters: Botloc, botloc_amp
Botloc (Figure 5) is defined as the lowest location in the section of the waveform between *search_start* and *search_end* where two adjacent intensities occur above *back_threshold* (Figure 5). This defines the lowest detectable return. Parameters are determined using the version of rxwaveform denoised by convolution with a gaussian of width *smoothwidth*.

Botloc_amp is the intensity of the denoised waveform at the location of *botloc*.

Parameter: Toploc
Toploc (Figure 5) is defined as the highest detectable return where two adjacent intensities occur above *front_threshold* (Figure 5) within the section of the rxwaveform between *search_start* and *search_end*. *Toploc* defines the highest detected return. For example, in a vegetated area, this will correspond (once geolocated) to the canopy top. Parameters are determined from the version of rxwaveform denoised by convolution with a gaussian of width *smoothwidth*.

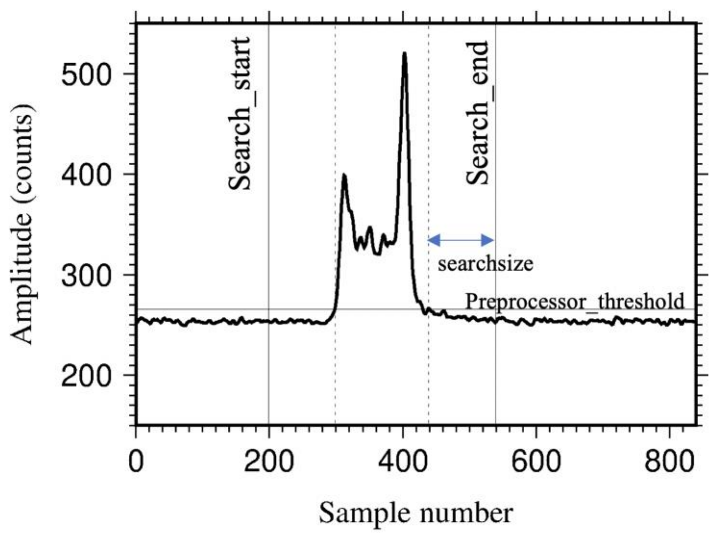

Figure 4. Example GEDI L1B rxwaveform. The section of the waveform that ise searched for surface reflections is located between *search_start* and *search_end*. These locations are determined based on amplitude threshold crossings plus a *searchsize* area specified as input to the algorithm.

4.3.5 Detect modes within waveform

Parameters: rx_nummodes, rx_modeamps, rx_modelocs, rx_modewidths
The rxwaveform may contain several distinct modes representing reflecting surfaces within each laser footprint. An area of the rxwaveform denoised with a gaussian of width *smoothwidth_zcross* between *search_start* and *search_end* is searched (Figure 5). A mode is defined as a zero crossing

point of the first derivative of the de-noised waveform. Waveform intensity of the mode must exceed *back_threshold*. A maximum of *ancillary/max_mode_counts* are stored. If the number of detected modes exceeds this value, no output from the algorithm is produced. Some of the detected modes may correspond to noise.

Rx_nummodes: number of distinct modes detected by the algorithm
Rx_modelocs: location of each distinct mode detected by the algorithm
Rx_modeamps: intensity of de-noised waveform at corresponding *rx_modelocs* location.
Rx_modewidths: estimate of width of each mode detected by algorithm, defined as the distance to the subsequent mode/2.

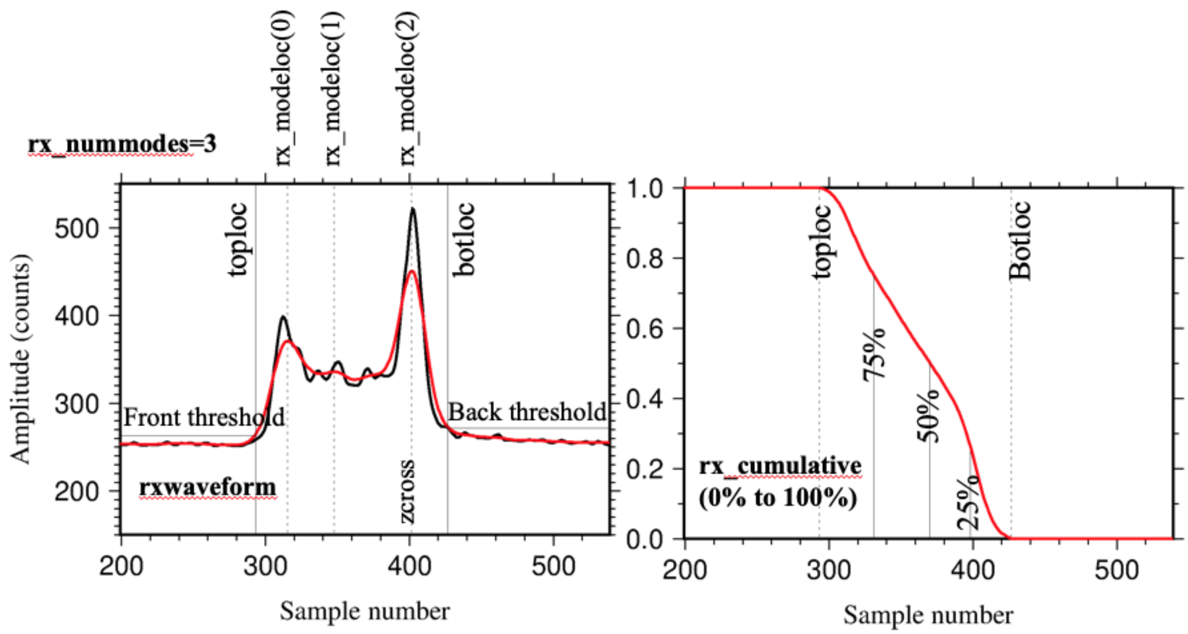

Figure 5. (Left) Example GEDI rxwaveform before (black) and after (red) convolution with a Gaussian of width 6.5 ns. Only the section of the rxwaveform between *signal_start* and *signal_end* is searched by the algorithm for reflected modes. Threshold settings define the "highest" (leftmost, *toploc*) and "lowest" (rightmost, *botloc*). In this example, three distinct modes have been detected by the algorithm (*rx_nummodes*=3). The location of each mode is saved as *rx_modeloc.* (Right) The section of the smoothed waveform between *toploc* and *botloc* is summed to create the cumulative record. The samples corresponding to cumulative values from 0 (*botloc*) to 1 (*toploc*) in 1% increments is saved in *rx_cumulative*. Locations of the 25%, 50% and 75% cumulatives are shown.

4.3.6 Select lowest non-noise mode within waveform

Parameters: Selected_mode, selected_mode_flag
Selected_mode is the index into the *rx_modelocs/rx_modeamps/rx_modewidths* arrays of the lowest detected non-noise mode. By default, it corresponds to the lowest (rightmost) mode, but can be overruled in subsequent processing. The reason for selection overruling is contained in *selected_mode_flag* (Table 6).

Table 6. Description of selected_mode_flag settings.

Selected_mode_flag	Description	Target
0	Mode reselection criteria not applied, selected_mode remains unchanged	L2A/rx_processing_a<n>
1	No modes meeting selection criteria detected, selected_mode remains unchanged	L2A/rx_processing_a<n>
2	No modes meeting selection criteria detected, but the selected_mode has been changed to be the lowest mode more than the laser_pulse_width above *botloc*	L2A/rx_processing_a<n>
3	Selected mode meets selection criteria	L2A/rx_processing_a<n>
4	Selected mode meets selection criteria but is potentially too far above *botloc*	L2A/rx_processing_a<n>

Parameters: zcross, zcross_amp, zcross0
Zcross (Figure 5) is the center of the lowest detected mode and corresponds to the lowest zero crossing point occurring between *toploc* and *botloc* in the first derivative of the waveform. *Zcross* corresponds to *rx_modelocs(selected_mode).*

Zcross_amp is the intensity of the denoised waveform at the location of *zcross* and is determined from *rx_modeamps(selected_mode).*

Zcross0 is the center of the highest detected mode and corresponds to the highest zero crossing point occurring between *toploc* and *botloc* in the first derivative of the waveform. *Zcross0* corresponds to *rx_modelocs(0).* In simple terrain where only one mode is detected in the waveform, *zcross* and *zcross0* are identical.

4.3.7 Calculate energy parameters

Energy parameters can be useful in delineating noise from signal returns in the rxwaveform and we generate several parameters to aid in the selection of "zcross".

Parameter: Rx_modelocalslope
Array containing local slopes for each detected mode of the rxwaveform between samples *rx_modeloc-8* and *rx_modeloc+8.*

Parameter: Rx_modelocalenergy
Array containing summation of intensities between samples *rx_modeloc-8* and *rx_modeloc+8* with local slope subtracted.

Parameter: Rx_modelocalenergyabovemean
Array containing summation of intensities between samples *rx_modeloc-8* and *rx_modeloc+8*

Parameter: rx_modeenergytobotloc

Array containing summation of intensities between *rx_modeloc* and *botloc*.

Parameter: lastmodeenergy
Calculation of energy contained in the lowest mode, derived using 2 * summation of intensity from zcross to botloc.

$$Lastmodeenergy = modeenergytobotloc(selected\_mode)*2$$

4.3.8 Energy statistical metrics

Parameters: rx_cumulative
Array of waveform bin numbers of integer percents of integrated energy from cumulative waveform (botloc (0%) to toploc (100%)) using waveform de-noised using a gaussian of width *smoothingwidth_zcross*.

Parameters: rx_iwaveamps
Fraction of integrated waveform at location of each detected mode.

4.3.9 Waveform Sensitivity

Parameters: Min_detection_energy, min_detection_threshold
Due to atmospheric conditions that will vary from completely clear to completely clouded and all things in between, GEDI return signals will greatly vary in strength and, the ability to penetrate canopies is dependent on return signal strength. To aid in identifying waveforms where we may not have detected the true ground level (due to weak return signals and/or high density canopy cover), we compute a signal detection performance metric for the ground detection capability for each waveform, dubbed the "sensitivity" metric. Sensitivity is computed by simulating the minimum detectable ground return pulse energy for the given detection algorithm. The area of that return will be divided by the area for the total return waveform to produce the *sensitivity* parameter. This provides an estimate for the relative minimum percentage of the return that needs to be present in the ground return for it to be detected.

4.3.10 Parameters characterizing the smoothed waveform

Parameters: mean, mean_sm, stddev, sd_sm
Mean noise and noise standard deviations of the background signal used in the algorithm are saved. Note, although space has been left for both pre and post-smoothing means and standards deviations, both are currently filled with the *mean* and *sd_corrected* parameters from the L1B data product.

Parameters: peak, pk_sm
Maximum intensity of the rxwaveform prior to and after smoothing with a Gaussian of width *smoothingwidth_zcross.*

Parameter: energy_sm

Summation of intensity of rxwaveform after smoothing with a Gaussian of width *smoothingwidth_zcross.*

4.3.11 Other parameters

Parameter: *Rx_algrunflag*
Indicates error run of the algorithm using selected settings.

Parameter: *toploc_miss*
Indicates *toploc* was detected below *botloc* due to combination of smoothing and threshold parameters

4.4 Required Inputs

The required inputs to the RX waveform analysis procedure are contained in the L1B data product and are listed in Table 4.

4.5 Output Parameters

The output data fields from the RX waveform analysis are summarized in Tables 7 and 8.

Table 7. Data fields associated with characterizing the GEDI Receive waveform, contained in the "rx_assess" subgroup of the L2A data product.

Parameter	Description	Target
sd_corrected	Noise standard deviation, odd/even corrected. Obtained from L1B	rx_assess
rx_maxpeakloc	Location of maximum within RX waveform	rx_assess
rx_maxamp	Amplitude of RX waveform maximum	rx_assess
rx_energy	RX waveform energy	rx_assess
rx_clipbin_count	Counts above clipping threshold	rx_assess
rx_clipbin0	First count above clipping threshold	rx_assess
rx_assess_flag	Bit field of quality flags	rx_assess
quality_flag	Flag: RX waveform good based on assess parameters	rx_assess
ocean_calibration_shot_flag	Flag: return from ocean, to be used for range/pointing calibration	rx_assess
mean_64kadjusted	(never set?)	rx_assess
mean	Mean from noise window, obtained from L1B	rx_assess

Table 8. Data fields contained in the “rx_processing_a<n>” subgroups of the L2A data product. “<n>” denotes the group of algorithm settings used to interpret the rxwaveform.

Parameter	Description	Target
back_threshold	threshold used to detect lowest elevation return energy	rx_processing_a<n>
botloc	waveform sample location of lowest detected return energy relative to bin0 of waveform	rx_processing_a<n>
botloc_amp	amplitude at lowest detected energy return	rx_processing_a<n>
energy_sm	total energy of smoothed waveform	rx_processing_a<n>
front_threshold	threshold used to detect highest elevation return energy	rx_processing_a<n>
lastmodeenergy	energy in lowest detected mode	rx_processing_a<n>
mean	mean noise level used in algorithm	rx_processing_a<n>
mean_sm	mean noise level after smoothing	rx_processing_a<n>
min_detection_energy	integrated area of the computed minimally-detectable gaussian	rx_processing_a<n>
min_detection_threshold	detection threshold used to compute the minimally detected gaussian	rx_processing_a<n>
peak	peak amplitude of raw waveform	rx_processing_a<n>
pk_sm	peak amplitude of smoothed waveform	rx_processing_a<n>
rx_algrunflag	Flag indicating signal was detected and algorithm ran successfully	rx_processing_a<n>
rx_cumulative	Waveform bin numbers of integer percents of integrated energy from cumulative waveform (botloc (0%) to toploc (100%))	rx_processing_a<n>
rx_iwaveamps	Fraction of integrated waveform at location of each detected mode	rx_processing_a<n>
rx_modeamps	Amplitudes of each detected mode within waveform	rx_processing_a<n>
rx_modeenergytobotloc	Total energy from the center of each detected waveform mode to botloc	rx_processing_a<n>
rx_modelocalenergy	Energy between +- 8 samples of each detected mode, mean noise level removed	rx_processing_a<n>
rx_modelocalenergyabovemean	Energy between +- 8 samples of each detected mode, mean noise level removed	rx_processing_a<n>
rx_modelocalslope	Signal trend within +- 8 samples of each detected mode	rx_processing_a<n>

rx_modelocs	Sample numbers of each detected mode (relative to bin 0 of waveform)	rx_processing_a<n>
rx_modewidths	1 sigma width estimates of each detected mode in waveform	rx_processing_a<n>
rx_nummodes	Number of modes detected in waveform	rx_processing_a<n>
sd_sm	Noise standard deviation of the smoothed waveform	rx_processing_a<n>
search_start	Sample number indicating start of signal search	rx_processing_a<n>
search_end	Sample number indicating end of signal search	rx_processing_a<n>
selected_mode	ID of mode selected as lowest non-noise mode	rx_processing_a<n>
selected_mode_flag	Flag indicating status of selected_mode	rx_processing_a<n>
smoothwidth	width of gaussian function used to smooth noise sections of waveforms	rx_processing_a<n>
smoothwidth_zcross	width of gaussian function used to smooth waveform between botloc and toploc	rx_processing_a<n>
stddev	noise standdard deviation used in algorithm	rx_processing_a<n>
toploc	Sample number of highest detected return	rx_processing_a<n>
toploc_miss	Flag indicating algorithm didn't detect valid toploc value	rx_processing_a<n>
zcross	Sample number of center of lowest mode above noise level	rx_processing_a<n>
zcross0	location of center of highest mode above noise level relative to bin0 of waveform	rx_processing_a<n>
zcross_amp	amplitude of smoothed waveform at lowest detected mode	rx_processing_a<n>
zcross_localenergy	energy of last mode above local slope	rx_processing_a<n>

5.0 GEDI Receive Waveform Elevations, Heights, Sensitivity and Quality Inputs to the L2 Geolocation Products

5.1 Overview

Geolocation of the RX waveform window is completed in the L1B. Precise timing points for various surfaces relative to the start of the RX waveform are completed in the RX waveform analysis. For GEDI L2A geolocation and height products, the precise timing points within each RX waveform are geolocated using their computed offset to the start of the RXwaveform in a linear interpolation of the L1B latitudes, longitudes, and elevations. Height products are subsequently computed relative to the elevation of the lowest detected mode. Details of the receive waveform geolocation procedure are given in the GEDI Geolocation ATBD. An overview of the translation between the L2a/rx_processing_a<N> and L2A/geolocation subgroups is given here:

- Using the geolocation of the highest and lowest rxwaveform bins from the L1b/geolocation and L2a/rx_processing_a<n> sub groups:
- compute offset of RX timing points from the linear interpolation of the L1B latitudes, longitudes and elevations
- calculate relative heights of RX timing points relative to center of lowest detected mode (“ground”)
- calculate waveform sensitivty
- perform QA (crossovers, comparison to other available data) to establish selection of algorithm inputs for placement in the root group of the L2A data product on a shot by shot basis. This is currently set to “algorithm1” pending further cal/val.

5.2 L2A Rxwaveform products input to Geolocation

5.2.1 Elevation, latitude and longitude parameters

Table 9 contains the mapping between the rxwaveform ranging points and elevation, latitude and longitude products. The relevant L2A/rx_processing_a<n> products are in digitizer samples relative to the start of the rx waveform. We linearly interpolate the L1B geolocation to derive the geolocation of the L2A ranging points (see the Geolocation ATBD for further details).

Table 9. Elevation, latitude and longitude parameters contained in the L2A geolocation subgroup and the rx_processing_a<n> parameter to which they correspond when combined with the L1B geolocation.

L2A geolocation Parameter	Description	L2A rx_processing parameter
Lats_allmodes_a<n>	Latitudes of centers of all detected modes	*rx_processing_a<n>/ rx_modelocs*

Lons_allmodes_a<n>	Longitudes of centers all detected modes	*rx_processing_a<n>/rx_modelocs*
Elevs_allmodes_a<n>	Elevations of centers of all detected modes	*rx_processing_a<n>/rx_modelocs*
Lat_lowestmode_a<n>	Latitude of center of lowest detected mode	*rx_processing_a<n>/zcross*
lon_lowestmode_a<n>	Longitude of center of lowest detected mode	*rx_processing_a<n>/zcross*
elev_lowestmode_a<n>	Elevation of ceneter of lowest detected mode	*rx_processing_a<n>/zcross*
Lat_highestreturn_a<n>	Latitude of highest detected return	*rx_processing_a<n>/toploc*
lon_highestreturn_a<n>	Latitude of highest detected return	*rx_processing_a<n>/toploc*
elev_highestreturn_a<n>	Elevation of highest detected return	*rx_processing_a<n>/toploc*
Lat_lowestreturn_a<n>	Latitude of lowest detected return	*rx_processing_a<n>/botloc*
lon_lowestreturn_a<n>	Longitude of lowest detected return	*rx_processing_a<n>/botloc*
elev_lowestreturn_a<n>	Elevation of lowest detected return	*rx_processing_a<n>/botloc*
Latitude_1gfit	Latitude of single gaussian fit to the rxwaveform	*rx_1gaussfit/rx_gloc*
longtiude_1gfit	Longtiude of single gaussian fit to the rxwaveform	*rx_1gaussfit/rx_gloc*
elevation_1gfit	Elevation of single gaussian fit to the rxwaveform	*rx_1gaussfit/rx_gloc*
rh	Relative heights (101 pt array)	*rx_processing_a<n>/rx_cumulative geolocation/elev_lowestresturn_a<n>*

5.2.2 Relative Height (RH) Metrics, *rh_a<n>*

RH (relative height) metrics are based on the rx_processing_a<n>/rx_cumulative products. Their elevations are derived from the linear interpolation of the L1B geolocation information, then the heights are calculated by subtracting the elev_lowestmode_a<n> field. RH metrics do not have an associated latitude and longitude. Note, the units of these metrics are centimeters.

5.2.3 Energy metrics, energy_lowestmode_a<n>

Geolocation/energy_lowestmode_a<n> is given by rx_processing_a<n>/lastmodeenergy.

5.2.4 Footprint sensitivity, *sensitivity_a<n>*

Due to atmospheric conditions that will vary from completely clear to completely clouded and all things in between, GEDI return signals will greatly vary in strength and, the ability to penetrate canopies is dependent on return signal strength. To aid in identifying waveforms where we may not have detected the true ground level (due to weak return signals and/or high-density canopy cover), we compute a signal detection performance metric for the ground detection capability for each waveform, dubbed the "sensitivity" metric. Sensitivity is computed by simulating the minimum detectable ground return pulse energy for the given detection algorithm. The area of that return is divided by the area for the total return waveform to produce the *sensitivity* parameter. This provides an estimate for the relative minimum percentage of the return that needs to be present in the ground return for it to be detected.

Sensitivity_a<n> = 1.0 – {rx_processing_a<n>/min_detection_energy} / {rx_assess/rx_energy}

5.2.5 Footprint Quality, *quality_flag_a<n>*

In order to provide end users with the ability to easily remove erroneous and/or lower quality returns, we provide a quality_flag based on the output of each algorithm, This is a summation of several individual quality assessment paramaters and is intended to provide general guidance only (Figure 6).

The conditions that are used to produce a quality_flag of 1 (good) are:

rx_assess/quality_flag	=	1
geolocation/surface_flag	=	1
stale_return_flag	=	0
rx_assess/rx_maxamp	>	8.* rx_assess/sd_corrected
sensitivity_a<n>	≤	1.0
rx_processing_a<n>/rx_algrunflag	=	1
rx_processing_a<n>/zcross	>	0
rx_processing_a<n>/toploc	>	0
If over land: sensitivity_a<n>	>	0.9
if over ocean: sensitivity_a<n>	>	0.5

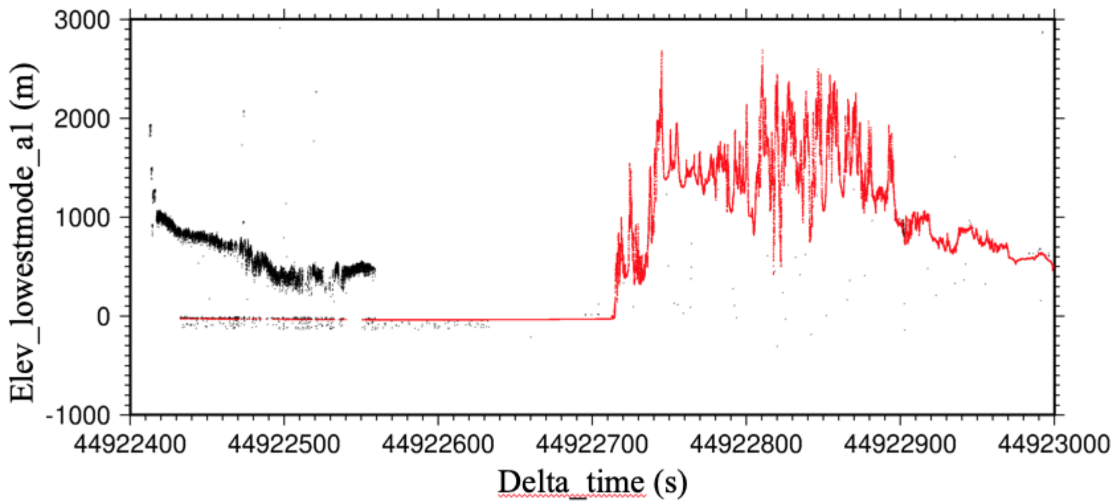

Figure 6. Elevation along a portion of Orbit 2703 showing all laser footprint elevations for BEAM1000 (black) and the subset identified by quality_flag = 1 (red). Cloud returns, along with elevations corresponding to spurious background noise triggers are removed.

5.3 Selecting set of algorithm results appropriate for each laser footprint

The root group of the L2A data product currently contains the output of Algorithm 1 (A_1), pending cal val of the products.

5.4 Required Input Fields

The inputs needed to produce parameters in the L2A geolocation subgroup are contained in the L1B/geolocation data product (Table 10), combined with parameters in the L2A/rx_processing_a<n> subgroup (Table 9).

Table 10. Parameters required to produce the L2A geolocation products.

Parameter	Description	Source
Elev_bin0	Elevation of first bin of the receive waveform	L1B
Latitude_bin0	latitude of first bin of the receive waveform	L1B
Longitude_bin0	longitude of first bin of the receive waveform	L1B
Elev_lastbin	Elevation of last bin of the receive waveform (corresponding to rx_sample_count)	L1B

Latitude_lastbin	latitude of last bin of the receive waveform (corresponding to rx_sample_count)	L1B
Longitude_lastbin	longitude of last bin of the receive waveform (corresponding to rx_sample_count)	L1B
Rx_sample_count	Number of telelmetered samples in rxwaveform record	L1B

5.5 Output Parameters

Data fields found in the geolocation subgroup of the L2A data product are listed in Table 11. Ranging points located by each algorithm run are geolocated and retained in the data product. The suggested result for each laser footprint is then stored in the root group of the L2A product (Table 12). Note, this is currently set to the output of Algorithm 1 (Table 5) and will be updated as cal/val is performed.

Table 11. Parameters output in the L2A geolocation subgroup.

Parameter	Description	Section
elev_highestreturn_a<n>	Elevation of the highest return detected using algorithm <n>, relative to reference ellipsoid	geolocation
elev_lowestmode_a<n>	Elevation of the center of the lowest mode detected using algorithm <n>,relative to reference ellipsoid	geolocation
elev_lowestreturn_a<n>	Elevation of lowest return detected using algorithm <n>, relative to reference ellipsoid	geolocation
elevation_1gfit	Elevation corresponding to the center of a single gaussian fit to the waveform, relative to reference ellipsoid	geolocation
elevs_allmodes_a<n>	Elevations of all modes detected using algorithm <n>, relative to reference ellipsoid	geolocation
energy_lowestmode_a<n>	Energy of lowest mode, detected using algorithm <n>, in the waveform above the mean noise level	geolocation
lat_highestreturn_a<n>	Latitude of the highest return detected using algorithm <n>	geolocation
lat_lowestmode_a<n>	Latitude of the center of the lowest mode detected using algorithm <n>	geolocation
lat_lowestreturn_a<n>	Latitude of the lowest return detected using algorithm <n>	geolocation
latitude_1gfit	Latitude corresponding to the center of a single gaussian fit to the waveform	geolocation

lats_allmodes_a<n>	Latitudes of all modes detected using algorithm <n>	geolocation
lon_highestreturn_a<n>	Longitude of the highest return detected using algorithm <n>	geolocation
lon_lowestmode_a<n>	Longitude of the center of lowest mode detected using algorithm <n>	geolocation
lon_lowestreturn_a<n>	Longitude of lowest return detected using algorithm<n>	geolocation
longitude_1gfit	Longitude corresponding to the center of a single gaussian fit to the waveform	geolocation
lons_allmodes_a<n>	Longitudes of all modes detected using algorithm<n>	geolocation
num_detectedmodes_a<n>	Number of detected modes detected using algorithm <n>	geolocation
quality_flag_a<n>	Flag simplifying selection of most useful data	geolocation
rh_a<n>	Relative height metrics at 1 % intervals using algorithm <n> (in cm)	geolocation
sensitivity_a<n>	Maximum canopy cover, using algorithm <n>, that can be penetrated considering the SNR of the waveform	geolocation
shot_number	Shot_number	geolocation
stale_return_flag	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.	geolocation

Table 12. Parameters output in the L2A root group.

Parameter	Description	Origin	Section
elev_highestreturn	Elevation of highest detected return relative to reference ellipsoid	geolocation/ elev_highestreturn_a1	/
elev_lowestmode	Elevation of center of lowest mode relative to reference ellipsoid	geolocation/ elev_lowestmode_a1	/
elevation_bias_flag	Elevations potentially affected by 4bin (~60cm) ranging error		/

energy_total	Integrated counts in the return waveform relative to the mean noise level	rx_assess/rx_energy	/
lat_highestreturn	Latitude of highest detected return	geolocation/ lat_highestreturn_a1	/
lat_lowestmode	Latitude of center of lowest mode	geolocation/ at_lowestmode_a1	/
lon_highestreturn	Longitude of highest detected return	geolocation/ on_highestreturn_a1	/
lon_lowestmode	Longitude of center of lowest mode	geolocation/ lon_lowestmode_a1	/
num_detectedmodes	Number of detected modes in rxwaveform	geolocation/ num_detectedmodes_a1	/
quality_flag	Flag simplifying selection of most useful data	geolocation/ quality_flag_a1	/
rh	Relative height metrics at 1 % intervals	geolocation/ rh_a1	/
selected_algorithm	ID of algorithm selected as identifying the lowest non-noise mode	Currently set to 1	/
selected_mode	ID of mode selected as lowest non-noise mode	geolocation/ selected_mode_a1	/
sensitivity	Maximum canopy cover that can be penetrated considering the SNR of the waveform	geolocation/ sensitivity_a1	/

6.0 References

Blair, J.B., D Rabine, and M Hofton, 1999. The Laser Vegetation Imaging Sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography." ISPRS Journal of Photogrammetry and Remote Sensing, **54 (2-3):** 115-122 [10.1016/s0924-2716(99)00002-7]

Hofton, M., J Blair, J. B Minster et al., 2000. "An airborne scanning laser altimetry survey of Long Valley, California." International Journal of Remote Sensing, **21 (12):** 2413-2437 [10.1080/01431160050030547]

Markwardt, C. B. 2009, "Non-Linear Least Squares Fitting in IDL with MPFIT," in proc. Astronomical Data Analysis Software and Systems XVIII, Quebec, Canada, ASP Conference Series, Vol. 411, eds. D. Bohlender, P. Dowler & D. Durand (Astronomical Society of the Pacific: San Francisco), p. 251-254 (ISBN: 978-1-58381-702-5

Sun X., J Abshire, et al., 2017. ICESat/GLAS Altimetry Measurements: Received Signal Dynamic Range and Saturation Correction, IEEE Trans Geosci Remote Sens. 2017 Oct; 55(10): 5440–5454 . [10.1109/TGRS.2017.2702126]

GLOSSARY/ACRONYMS

ATBD	Algorithm Theoretical Basis Document
BCE	Bench Checkout Equipment
GEDI	Global Ecosystem Dynamics Investigation
LVIS	Land, Vegetation and Ice Sensor
LUT	Look-up Table
QC	Quality Control
TBD	To Be Determined
TLS	Terrestrial Laser Scanning
VCL	Vegetation Canopy Lidar