Common queries on how to use and interpret the GEDI04_A data product

How are the GEDI04_A biomass estimates geolocated?

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

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

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

AGBD is predicted for every shot where it is possible to run the GEDI04_A algorithm, as indicated by the *algorithm_run_flag* dataset (see Table 2 in the dataset User's Guide Doc). The GEDI04_A product provides multiple quality flags and metrics that may be used to subset the predictions to the most useful observations for a particular application or region.

The *l2_quality_flag* encapsulates a number of GEDI02_A quality metrics to identify land surfaceshots with waveforms of high fidelity for AGBD estimation. The *l4_quality_flag* identifies shots that may be considered as samples of the population for which the applied models are representative of. For example, GEDI04_A models for deciduous forests are only calibrated using GEDI waveforms simulated from leaf-on ALS data, therefore we can only apply the derived models to on-orbit GEDI waveforms acquired under similar conditions.

The *l2_quality_flag* uses a beam sensitivity threshold of 0.9 to match what is used for the Level 2products. The *l4_quality_flag* uses a beam sensitivity threshold of 0.95, which was selected based on analysis of GEDI02_A and GEDI04_A on-orbit data. Beam sensitivity is an estimate of the maximum canopy cover that can be penetrated considering the signal-to-noise ratio of the waveform. For dense tropical forests, users may consider raising the beam sensitivity threshold (e.g., 0.98) to minimize measurement error in the RH metrics. In future releases, quality filtering will be improved by using the beam sensitivity together with expected level of canopy cover for each shot.

Some users may wish to also evaluate the *predictor_limit_flag* and *response_limit_flag*. These identify shots with RH metrics or AGBD predictions, respectively, that are outside the observed

range of values used to train the GEDI04_A models. Care should be taken when using such observations.

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

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

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

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

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

$$predictor\_id = [1, 2, 3, 3, 0]$$

and rh_index is:

$$rh\_index = [50, 98, 50, 70, 0]$$

the associated estimator (ignoring transformations) would be:

$$AGBD = par[0] + par[1]\text{xRH50} + par[2] \times RH98 + par[3] \times RH50 \times RH70$$

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

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

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

estimate ± *(standard error x t-multiplier)*

where the estimate is the sample prediction in transform space (*agbd_t* in Table 2 of the dataset User's Guide Doc) and standard error is the standard error of the prediction in transform space (see *agbd_t_se* in Table 2 of the dataset User's Guide Doc). The t-multiplier ($t_{1-\alpha/2,dof}$) can be derived using standard libraries in R or Python and depends on: (1) the degrees-of-freedom for the applied model (*dof*), which is provided in the L4A product (*ANCILLARY/model_data/dof*); and (2) the t-distribution probability (α), which is specified by the *BEAMXXXX/agbd_prediction* group attribute *alpha* and may be modified by the user. For example, an α value of 0.1 is used for a 90% confidence level and 0.05 for a 95% confidence level.

Note that the prediction intervals described above are in transform space and need to be back-transformed to place estimates in units of aboveground biomass density. A correction also needs to be applied to account for bias introduced by transformation of the response variable (agbd). For example, if *ANCILLARY/model_data/y_transform* is "sqrt" and *ANCILLARY/model_data/bias_correction_name* is "Snowdon", then

agbd = agbd_t^2 ** ANCILLARY/model_data/bias_correction_value*

If ANCILLARY/model_data/y_transform is "exp" and
ANCILLARY/model_data/bias_correction_name is "Baskerville", then
agbd = exp(agbd_t) * exp(ANCILLARY/model_data/bias_correction_value)