Questions & Answers Part 1

Please type your questions in the Question Box. We will try our best to get to all your questions. If we don't, feel free to email Savannah Cooley (savannah.cooley@nasa.gov), Naiara Pinto (naiara.pinto@jpl.nasa.gov), or Erika Podest (erika.podest@jpl.nasa.gov).

Exercises 1 and 2

Exercise 1A Repository: https://github.com/savcooley/full_waveform_lidar_training
Exercise 1B Repository : https://github.com/NaiaraSPinto/VegMapper/tree/master/gedi

Question 1: Does the platform and fine resolution guarantee accurate biomass estimates in temperate and boreal forests?

Answer 1: No, while airborne lidar provides more accurate biomass estimates than spaceborne systems, it does not guarantee accuracy due to several fundamental uncertainties. Airborne platforms offer advantages over spaceborne systems because of: (1) reduced geolocation error—airborne systems typically have positional uncertainties of 10-50cm compared to GEDI's ±10m uncertainty, reducing the risk of sampling incorrect land cover types; and (2) higher point density—airborne discrete return lidar achieves 1-50 points per square meter versus GEDI's 25m diameter footprints separated by 60m along-track, providing better ground penetration in dense canopies. However, both airborne and spaceborne lidar face the fundamental challenge of converting structural measurements (canopy height, RH metrics) into biomass estimates using allometric equations developed from limited field samples. These allometric relationships are region-specific, represent statistical correlations rather than direct physical measurements, and have inherent uncertainties that affect biomass estimation regardless of platform. Additional uncertainties include canopy penetration limitations in very dense forests, temporal mismatches between lidar acquisition and field calibration data, and variations in wood density and tree architecture across species and environmental conditions. Session 3 will explore these biomass estimation challenges in greater depth.

Question 2: Why do we use RH 98 for canopy height (and not RH 100, for example)? What is the possible application of RH 50? Does RH 0 correspond to ground elevation?
Answer 2: RH98 is preferred over RH100 for canopy height because RH100 can be artificially high due to noise in the spaceborne waveform, potentially capturing outliers such as clouds, birds, or instrument noise rather than true canopy. The noise is not fully processed out in Level 2 products, and RH100 sometimes fails to accurately represent the actual canopy surface. RH95 and RH98 provide more robust estimates of maximum canopy height by filtering out these extreme outliers while still capturing the upper canopy structure. Sometimes RH100 does detect a real canopy height – for example, of a tall pine tree in the footprint – but researchers may still prefer to use RH98 or RH95 to reflect a height that is closer to the heights of most other trees within the 25m diameter footprint.

RH50 represents the height of median energy—the height at which 50% of the cumulative returned energy has been received—and is particularly valuable for understanding understory characteristics and how vegetation is distributed vertically. For example, temperate forests with denser understory will show higher RH50 values compared to forests with sparse understory, even if they have similar canopy heights. RH0 theoretically corresponds to ground elevation where 0% of energy has been accumulated, but in practice, some RH values can be negative relative to the identified ground surface due to noise, slope effects within the footprint, or uncertainty in ground detection—these negative values are typically filtered out during quality control rather than being interpreted as below-ground measurements.

Question 3: What is the global availability of this data?
Answer 3: GEDI is aboard the International Space Station (ISS) and therefore follows its orbital path and altitude, which mostly lies between 51.6° N and 51.6° S so it will miss far north and south regions of the globe. Following the ISS trajectory may also affect sampling density, making it variable over some areas and time periods—areas closer to the ±51.6° latitude limits receive more frequent overpasses than equatorial regions due to the orbital geometry.

Question 4: Since LiDAR is an active data collection platform, why is it limited by cloud cover?
Answer 4: While GEDI uses active sensing (generating its own laser energy rather than relying on sunlight), it is still limited by cloud cover because it operates at the near-infrared wavelength of 1064 nm. As mentioned in Slide 19, this wavelength cannot

penetrate clouds, unlike some other active sensors such as radar (SAR) which use microwave frequencies that can penetrate clouds and some vegetation. The key distinction is that "active" refers to generating its own illumination source, but the ability to penetrate clouds depends on the wavelength used. Different types of active sensors have different cloud penetration capabilities: radar systems can see through clouds, but optical-wavelength lidars (both green and NIR) cannot. This is why GEDI data collection is limited to cloud-free conditions, which can affect sampling density and temporal coverage in persistently cloudy regions.

Question 5: You stated the green band penetrates water. Can it be used to study water pollution from an industry like mining?
Answer 5: We are not aware of lidar being used directly for water quality or pollution assessment. Bathymetric lidar using green wavelengths (like ICESat-2's 532 nm laser) can penetrate clear water to map underwater topography and potentially detect submerged vegetation or substrate changes, but it does not provide chemical composition information needed to assess pollution. For monitoring water pollution from mining or other sources, you would need spectral information from optical or hyperspectral sensors that can detect changes in water color, turbidity, or specific chemical signatures—not structural information from lidar.

Question 6: Is there any filtering applied to filter out shots with cloud cover?
Answer 6: There are 2 different quality flags which we will learn to use that will help with the removal of observations that were contaminated from clouds. RH 95 could show 200m, so add a threshold to the maximum canopy can be used. More information on quality filtering can be found in the User Guides on EarthData catalog (by the LP and ORNL DAACs). For more background on quality considerations can be found here.

Question 7: How can GEDI be used to estimate or identify deforestation?
Answer 7: It can tell us about the presence of trees or lack thereof. Sparse areas of grass give a single ground return. GEDI will likely need to be paired with other datasets to distinguish among land cover types of similar vertical profile returns (e.g., different types of degraded forest, such as previously burned forest vs. previously logged forest).

Question 8: Are there some types of forest that GEDI is not able to fully penetrate, and thus not reaching the ground? If so, how is that detected and how is the structure analyzed?

Answer 8: Tropical forests are the type of forest where this happens most often. In these cases, none of the energy of the pulse reaches the ground and it is not advised to attempt to include these returns in analysis. If there is a difference greater than 2 meters between the 6 ground algorithms, that would be one way to identify such instances, which can be flagged and excluded with quality flag filtering.

Question 9: Given the increasing relevance of vegetation structure data in modeling vector-borne zoonotic risks, such as mosquito-borne diseases, how do you ensure that GEDI visualizations, particularly those involving RH metrics and waveform plots, are semantically coherent and accessible to interdisciplinary audiences? Are there plans to systematically integrate contextual annotations and narrative framing to support public health applications and early warning systems?

Answer 9: GEDI data visualization and interpretation inherently requires some remote sensing expertise, but efforts to make products more accessible include standardized quality flags, well-documented Algorithm Theoretical Basis Documents (ATBDs), and increasingly user-friendly tools through platforms like Google Earth Engine. For public health applications involving vector-borne disease modeling, GEDI's RH metrics (particularly understory density measurements like RH50) could contribute valuable habitat structure information, but this would require collaboration with epidemiologists to develop appropriate ecological-epidemiological frameworks. While systematic narrative framing for public health isn't currently a GEDI mission priority, the data are publicly available for interdisciplinary teams to integrate into early warning systems, and ARSET trainings like this one aim to build capacity across diverse user communities.

Question 10: Do you have GEDI for below ground level observation?

Answer 10: In areas of no vegetation, or sparse vegetation, GEDI data may be capturing some penetration into the ground or soil, however the algorithms do not calculate this ability into the product. The waveforms are geolocated to Earth surface (ellipsoid) where the center peak of lowest mode of the waveform is interpolated to calculate its elevation against the Earth to estimate the ground elevation. All the products are calculated in reference to the ground. This is why some shots may have negative values, but the scale and interpretation of these values is not optimized in the algorithm to truly represent relative distance below ground, instead, they are typically filtered out.

Question 11: Do you know which lidar products I can use to monitor Miyawaki tiny forest (0.2ha)?
Answer 11: Over an area this size, there may not be many GEDI samples available. This is because each of the 8 beams along the ground are 600 m apart, with each footprint along each of those beams being 60 m apart. Of the free and open-source spaceborne lidar datasets, there are GEDI, ICESat-2, and ICESat data. A more local search or working with commercial lidar data providers could possibly provide data with denser samples over the desired area and time period. For example, in the US, the forest service, or state governments publicly share some lidar data.

Question 12: What format is the data from the NASA API?
Answer 12: GEDI Level 1 and 2 data are in HDF5 format. Higher-level gridded products (Levels 3 and 4B) data are in GeoTiff format.

Question 13: I understood that the LiDAR and the surface vegetation cover (e.g., canopy height, etc.) are not continuous products, but how often are these products being captured and produced?
Answer 13: GEDI follows the path of the International Space Station (ISS). It can be hard to say exact repeat coverage in terms or location, density of samples, and when those observations will be collected. It is recommended to download and process the data to summarize the available returns, which will be highly variable depending on the location. Part 2 goes over how to do this with the acquired HDF5 files for footprint-level data. Other data products provided by the mission aggregate all GEDI data across the entire mission time period, or separate subset time periods. There is no guarantee that exact samples will overlap on the exact same location, which is why GEDI is best used in combination with wall-to-wall datasets.

Question 14: What is the main difference between full-waveform lidar and single-photon counting? Which is better for canopy height measurement?
Answer 14: Full-waveform lidar captures the complete energy profile of the returned laser pulse, providing detailed information about the vertical distribution of vegetation throughout the canopy, which makes it excellent for characterizing complex forest structure and deriving RH metrics at multiple canopy levels. Single-photon counting lidar (like ICESat-2) detects individual photons rather than complete pulses, which allows operation from much higher altitudes but provides less energy per measurement. For canopy height measurement specifically, both can be effective, but full-waveform systems like GEDI provide richer structural information about canopy

layers and understory, while single-photon systems offer different spatial coverage patterns—there's no single "better" option, as it depends on your specific application and what structural information you need beyond just height.

Question 15: What is the meaning of footprint?
Answer 15: The 'waveform' (data collected in its waveform format) is the actual captured data 'sample' (25m diameter sample, not an image), or resulting 'observation' (referring to the data collected) for each 'shot' (location and collected data from the transmitted and received laser energy) corresponding to each GEDI 'footprint' (the location where the shot/observation/sample took place).

Question 16: How can circular GEDI L2A RH98 footprints be effectively compared or integrated with square GEE rasterized footprint values? What is the best approach to align their geometry and handle overlap or area mismatches during analysis?
Answer 16: When comparing circular GEDI L2A footprints (25 m diameter) with rasterized GEE grid cells, the most common approach is to assign each footprint's values to the grid cell containing its center point, or use spatial aggregation methods that weight footprint contributions to cells based on area overlap. You should account for GEDI's ±10 m geolocation uncertainty by applying appropriate quality filters and potentially using buffers to ensure homogeneous land cover within 45 m diameter (25 m footprint + 20 m buffer). For analyses requiring precise geometric integration, consider using tools that calculate the actual area of intersection between circular footprints and square pixels, though for many applications the simpler center-point assignment with quality filtering is sufficient. There may be errors or mismatches between these products, as is possible with any reprojection or reformatting.

Question 17: How can we identify the maximum canopy height when the RH distribution is multimodal? RH100 could be artificially high. Is there a quality control analysis to identify false positive RH values?
Answer 17: This is a challenge when working with GEDI where it is meant to capture these variations in canopy height, but it is difficult to validate. Wherever possible it is recommended to calibrate GEDI with higher resolution lidar data, or field survey data. Or deploy rigorous quality controls based on literature or expert recommendations from applications of GEDI over similar ecosystems or land cover types. The characteristics

of the landscape can have variable effects on the data, which may need to be investigated or corrected before application.

Question 18: How does GEDI's waveform LiDAR distinguish between multiple canopy layers and ground returns in densely vegetated tropical forests, and what are the limitations of this approach in areas with mixed vegetation structure?
Answer 18: GEDI's full-waveform lidar distinguishes between canopy layers and ground by capturing the complete energy profile of the return signal, where different peaks in the waveform correspond to laser interactions at different heights—upper canopy produces high-elevation peaks, lower canopy layers create intermediate peaks, and ground returns appear as the lowest-elevation peak(s). However, as shown in Slide 38, limitations include potential canopy penetration failure in exceptionally dense tropical forests where the signal may be completely attenuated before reaching the ground, inability to clearly identify ground returns on steep slopes where the 25 m footprint spans multiple elevations simultaneously, and ambiguous waveforms (like the example in Slide 33) where multiple peaks make it difficult to definitively separate vegetation from ground returns—these problematic shots are typically filtered out using the quality_flag criteria where the six ground-finding algorithms must agree within 2 m.

Question 19: What is the footprint size and what is the distance between footprints?
Answer 19: GEDI's footprint size is 25 meters in diameter (the area illuminated by each laser pulse), and footprints are separated by 60 meters along each track. The eight parallel tracks themselves are separated by approximately 600 meters across-track, as shown in Slide 27.

Question 20: Is it possible to resample the GEDI data to create low, medium, and high vegetation height classes with user-defined height intervals?
Answer 20: Yes, you can use GEDI's RH metrics to classify vegetation into user-defined height classes. For example, you could define low vegetation as RH95 < 5 m, medium as 5-20 m, and high as > 20 m, or use any height thresholds appropriate for your application. The various RH percentiles (RH25, RH50, RH75, RH95, RH98) provide flexibility to characterize different aspects of the vertical structure, and you can combine multiple RH metrics to create more sophisticated classifications that capture both canopy height and structural complexity.

Question 21: What is the relation and differences between the Spaceborne LiDAR and 3D LiDAR Laser Scanner?
Answer 21: Both are lidar systems using the same basic principle of laser ranging, but they differ primarily in platform and scale. Terrestrial 3D laser scanners operate from fixed ground positions and provide extremely high-resolution measurements (millimeter accuracy) over small areas, ideal for detailed studies of individual trees or small plots. Spaceborne lidar like GEDI operates from satellites or the ISS, sacrificing some spatial resolution and point density to achieve global or near-global coverage—as discussed in Slide 20, this represents a fundamental tradeoff between the detailed, intensive measurements possible from ground-based systems versus the extensive coverage needed for regional to global monitoring.

Question 22: What is the coverage of GEDI data? and from which platform these data should be downloaded? Are they available to different dates? About the vegetation CHM under 1m (or beneath layer), which kind of these data should be applied? Thanks.

Answer 22: GEDI data covers latitudes between ±51.6° (the inclination of the ISS orbit) and has been collecting data since April 2019, with data available across different dates throughout the mission. GEDI L1 and L2 data are available from NASA's LP DAAC (Land Processes Distributed Active Archive Center), while L3 and L4 products are available from ORNL DAAC—both can be accessed through NASA Earthdata Search or programmatically using tools like the earthaccess Python library as demonstrated in the training exercises. For vegetation canopy height models (CHM) under 1 meter or beneath-layer vegetation, GEDI has limited capability since it may not detect very low vegetation separate from ground returns; you might need higher-density airborne lidar or other remote sensing approaches (like UAV-based lidar or high-resolution optical/radar data) for fine-scale understory characterization below 2-3 meters height.

Question 23: Does the waveform profile represent the average vegetation density? How about individual trees?
Answer 23: The waveform profile represents the vertical distribution of vegetation and other reflecting surfaces within the 25-meter diameter footprint, essentially providing an aggregated structural signature of all vegetation elements in that area. Individual trees cannot be distinguished—instead, you're seeing a composite signal that reflects the collective canopy structure, including contributions from multiple trees, branches, and

leaves at various heights. This is why GEDI is most useful for characterizing stand-level or plot-level forest structure rather than individual tree properties.

Question 24: Can we rely on GEDI data for tropical evergreen and semi-evergreen forests in Bangladesh? As canopy penetration failure is one of the limitations of GEDI.
Answer 24: GEDI data can be used in tropical evergreen and semi-evergreen forests in Bangladesh, but you should be aware of the canopy penetration limitation and apply rigorous quality filtering. As noted in Slide 38, in exceptionally dense canopies, the laser signal can be completely attenuated before reaching the ground, making ground detection uncertain or impossible. To maximize data reliability, always filter your GEDI observations using the quality_flag and sensitivity fields (keeping only shots with sensitivity ≥0.9 and quality_flag=1), check that the six ground-finding algorithms agree within 2m, and be particularly cautious in the densest forest areas where ground returns may be questionable—you may find that a substantial portion of shots in very dense tropical forests get filtered out, reducing your effective sample size.

Question 27: Are there any merged datasets with LIDAR + SAR, even at an annual cadence?
Answer 27: Literature explores fusing lidar and SAR. The GEDI Mission has put out a data product combining GEDI with TanDEM-X InSAR data.

While GEDI-SAR merged datasets aren't a standard operational product, there is growing research interest in integrating spaceborne lidar with radar data for improved biomass estimation and forest monitoring. The GEDI Mission has released a data product combining GEDI with TanDEM-X InSAR data. You might also explore academic literature for fusion approaches combining GEDI with missions like PALSAR-2, Sentinel-1, or the upcoming NISAR mission, as SAR's ability to penetrate clouds complements lidar's structural measurements. For operational applications, you could perform your own integration using both datasets in a modeling framework, though this would require handling the different spatial sampling patterns and sensitivities of the two sensor types.

Question 28: Can GEDI's vertical structure data be integrated with hyperspectral data (like from PRISMA or EnMAP) to model biodiversity proxies such as habitat complexity or species richness?
Answer 28: Yes, integrating GEDI's vertical structure data with hyperspectral data from sensors like PRISMA or EnMAP is conceptually valuable for modeling biodiversity

proxies, as structural complexity (measured by GEDI) and spectral diversity (from hyperspectral sensors) together can provide complementary information about habitat heterogeneity. However, this type of multi-sensor fusion would be a custom analysis requiring you to develop appropriate methods for handling the different spatial resolutions, sampling patterns, and data characteristics—it's not currently available as a standard GEDI product but represents an active area of research in biodiversity remote sensing.

Question 29: Will LiDAR be affected by wildfire?
Answer 29: Yes, lidar will detect the structural changes caused by wildfire. Post-fire forests show dramatically different waveform profiles and RH metrics compared to pre-fire conditions—burned areas typically exhibit reduced canopy height (lower RH95 and RH98), simplified vertical structure (different RH metric distributions), and stronger ground returns due to canopy loss. GEDI can therefore be useful for assessing fire impacts on forest structure and monitoring post-fire recovery over time, though remember that GEDI's temporal resolution limitations (weeks to months between revisits for a given location) mean you may not capture the fire event itself, only the before and after structural conditions.

Question 30: Can we use lidar for studying heavy metals in the soil? Which is better for this purpose, active or passive satellites?
Answer 30: Passive hyperspectral sensors such as PACE, with many band wavelengths, would be the best candidates for studying heavy metals in the soil. These are being explored for sensitivity in different materials like heavy metals, algal blooms, pollution, etc. GEDI is a sensor using the time it takes for the active light sent to return to the sensor in order to calculate the distance it took for the energy to interact with the surface target to the ground and back. While it is sensitive to materials that reflect in the NIR wavelength, it is more so summarizing the structure rather than the properties of the reflecting surfaces themselves.

Question 31: Can GEDI capture data on submerged vegetation such as mangroves? Is there any drawback?
Answer 31: GEDI cannot effectively capture submerged vegetation because it operates at a near-infrared wavelength (1064nm) that does not penetrate water—as mentioned in Slide 19, NIR wavelengths reflect strongly off vegetation but are absorbed by water. For mangroves specifically, GEDI can measure the above-water canopy structure effectively, but any prop roots or other structures below the water surface would not be

detected. If you need information about submerged aquatic vegetation, you would need bathymetric lidar systems that operate at green wavelengths (like ICESat-2's 532nm laser), which can penetrate clear water to some depth, though even these have limited penetration in turbid water.

Question 32: Is it possible to use GEDI data to identify tree species (e.g., bamboo)?
Answer 32: No, GEDI cannot directly identify tree species because it measures only the 3D structural properties of vegetation (height, vertical distribution) and does not capture spectral information that would be needed for species discrimination. To identify species like bamboo, you would need to integrate GEDI's structural data with spectral information from optical or hyperspectral sensors, where the combination of distinctive structural characteristics (e.g., bamboo's typically dense, uniform canopy at specific height ranges) and spectral signatures might enable classification. Species identification is an area of active research involving multi-sensor fusion rather than a current operational capability of GEDI alone.

Question 33: Is GEDI more suitable to understand canopy and heights or can it offer structural measurements too? What is the smallest amount of biomass/vegetation that can be detected by GEDI?
Answer 33: GEDI is designed to measure both canopy height (via RH95, RH98, RH100 metrics) AND detailed structural characteristics throughout the vertical profile (using RH metrics at various percentiles like RH25, RH50, RH75). The full suite of RH metrics, as demonstrated in Slides 34-37, provides rich information about canopy layering, understory density, and vertical complexity—this structural information is actually one of GEDI's primary strengths beyond just measuring height. Regarding minimum detectable biomass, this depends on several factors: GEDI has a beam sensitivity threshold (typically requiring sensitivity ≥0.9 for quality data), and very low or sparse vegetation may not return sufficient signal above noise levels. In practice, vegetation shorter than approximately 2-3 meters may be difficult to distinguish reliably from ground returns, though this varies depending on vegetation density and surface conditions. GEDI is optimized for forested ecosystems rather than grasslands or very low vegetation, where the structural signal may be weak or ambiguous.

Question 34: Is there a way to use GEDI to differentiate between the absence of trees and the presence of wilting trees?

Answer 34: GEDI's ability to differentiate these scenarios is limited because it primarily measures physical structure rather than vegetation health or physiological condition. Complete absence of trees would show only ground returns with no above-ground signal (like the example in Slide 32), which is straightforward to identify. However, wilting or stressed trees that still have physical structure (leaves, branches) would likely show canopy returns similar to healthy trees of the same size, since GEDI detects the physical presence of vegetation elements regardless of their physiological state. Severely wilting or dying trees might show some structural differences (reduced canopy height, simplified vertical structure) compared to their healthy state, but these changes would only be detectable if you have temporal data from before and after the decline. For vegetation health assessment and distinguishing stress versus absence, you would need to integrate GEDI structural data with optical or hyperspectral data that can detect chlorophyll content, water stress, or other physiological indicators through spectral signatures—this multi-sensor fusion approach would provide more definitive discrimination between absent and stressed vegetation.

Question 34: How can GEDI L1B data in HDF5 (.h5) format be converted into a shapefile using code, or what is an alternative method to achieve this conversion?

Answer 34: The demo script already extracts the coordinate data (latitude, longitude) and attributes from the HDF5 file into a pandas DataFrame (`result_df`). To convert this to a shapefile, you can add the following code after the data extraction is complete. You'll need to install `geopandas` first:

```
import geopandas as gpd
from shapely.geometry import Point

# Convert the pandas DataFrame to a GeoDataFrame
geometry = [Point(xy) for xy in zip(result_df['longitude'], result_df['latitude'])]
gdf = gpd.GeoDataFrame(result_df, geometry=geometry, crs='EPSG:4326')

# Export to shapefile
output_dir = "./gedi_L1B_outputs"
shapefile_path = os.path.join(output_dir, "gedi_l1b_shots.shp")
```

gdf.to_file(shapefile_path)
print(f" Shapefile exported to: {shapefile_path}")

Alternative formats (often preferred over shapefiles):

- GeoJSON: `gdf.to_file("gedi_l1b_shots.geojson", driver='GeoJSON')` - Better for web mapping, no file size limits
- GeoPackage: `gdf.to_file("gedi_l1b_shots.gpkg", driver='GPKG')` - Modern format, single file, no attribute name length restrictions
- CSV with WKT: Already being done in the script - can be easily imported into QGIS/ArcGIS

The shapefile format has limitations (10-character field name limit, multiple files, 2GB size limit), so GeoPackage or GeoJSON are often better choices for GEDI data, especially when processing multiple granules with many shots.

Question 35: Is RH98 available with the L2B data or only with the L2A data?
Answer 35: RH98 and all other Relative Height (RH) metrics are available in Level 2A data only. Level 2A contains "Elevation and height metrics" which includes all the RH percentiles (RH25, RH50, RH75, RH95, RH98, RH100). Level 2B contains "Canopy cover and vertical profile metrics" which are different derived products (like Plant Area Index (analogous to LAI), Foliage Height Diversity). Part 2 of this training series will address this. For canopy height measurements using RH metrics, you need to work with L2A data, which is what Exercise 2 in the training will demonstrate.

Question 3: Can I do the analysis in a heterogenous urban environment?
Answer 3: Technically yes, but with substantial caveats. GEDI was designed and optimized for vegetated ecosystems, particularly forests, and urban environments present multiple challenges. The ±10m geolocation uncertainty combined with heterogeneous land cover means a GEDI footprint might sample a tree, a building, a street, or a mix of these within the 25m diameter area. The best practice from the training (Slide 38) is to filter observations to areas with homogeneous land cover within 45m diameter (25m footprint + 20m buffer for geolocation error). In urban areas, you would need to apply very strict spatial filtering to ensure footprints are sampling vegetation only (like parks or street trees), exclude shots near buildings, and carefully interpret results since the waveforms may be contaminated by non-vegetation returns. Urban tree canopy studies have been done with GEDI, but require careful quality control and validation.

Question 36: In my colab environment, I am getting the following dependency conflicts when running the pip install code cell:

> *ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.*
> *datasets 4.0.0 requires fsspec[http]<=2025.3.0,>=2023.1.0, but you have fsspec 2025.9.0 which is incompatible.*
> *gcsfs 2025.3.0 requires fsspec==2025.3.0, but you have fsspec 2025.9.0 which is incompatible.*
> *google-adk 1.16.0 requires tenacity<9.0.0,>=8.0.0, but you have tenacity 9.1.2 which is incompatible.*

Answer 36: These dependency conflict warnings are common in Colab environments and are usually non-critical—your code will likely run fine despite the warnings. Here are solutions in order of preference: (1) **Ignore the warnings** if the notebook runs successfully—these are just version mismatches that typically don't break functionality; (2) **Restart runtime** using Runtime → Restart runtime in Colab, then run all cells fresh from the top—this often resolves transient conflicts; (3) **Pin specific compatible versions** by modifying the install cell to specify exact versions that work together, like `pip install fsspec==2025.3.0 tenacity==8.5.0`; (4) **Add the `--no-deps` flag** for specific packages if you know other dependencies are satisfied: `pip install --no-deps earthaccess`. For training purposes, if your code runs and you can authenticate and download GEDI data successfully, you can safely proceed despite these warnings—they're primarily informational and won't affect the GEDI data processing functions.

Question 37: Why do you think plot #6 in the 2.0 standard deviations example had a much more complicated structure than visually similar plots like plot #1? Is this potentially a factor of slope? Or likely just a more complex structure underneath the canopy?

Answer 37: Both factors could contribute to the complex waveform structure you're seeing. As discussed in Slide 38 of the training, slope-induced waveform contamination is a known limitation where steep topography causes the 25m footprint to hit the ground at different elevations simultaneously, distorting the waveform shape and creating multiple peaks. To determine which factor is dominant, you can check the `elevation_bin0` values and examine whether there's significant terrain variation in that area. Additionally, look at the degrade_flag for that shot—if the six ground-finding algorithms disagreed significantly (>2m), it would be flagged and might indicate slope contamination. However, complex waveforms can also genuinely reflect real vertical structural complexity: multiple canopy layers, dense understory, or heterogeneous vegetation within the footprint. The example in Slide 33 showed how ambiguous waveforms with multiple peaks can result from either terrain effects or actual vegetation complexity. If the area has relatively flat terrain based on the DEM data, then the complex structure is more likely reflecting real canopy architecture with multiple layers close to the ground—which is exactly the kind of rich structural information that full-waveform lidar is designed to capture.

Question 38: Is it possible to access the GEDI data and use the functions shown in the exercises using the R software packages? Is there an equivalent of harmony for R to allow spatial subsetting of the orbits prior to download?
Answer 38: Yes, R users can access and work with GEDI data using the `rGEDI` package, which was developed by the GEDI Science Team specifically for processing GEDI data in R. The `rGEDI` package provides functions to download GEDI data from LP DAAC, read HDF5 files, extract metrics, and perform spatial/temporal filtering. For Harmony-like spatial subsetting capabilities in R, you can use the `harpR` package, which provides an R interface to NASA's Harmony services and should support spatial subsetting of GEDI granules before download (similar to what's shown in Exercise 2 with Python). Additionally, standard R spatial packages like `rhdf5` (for reading HDF5 files), `sf` (for spatial operations), and `terra` (for raster operations) work well with GEDI data. The workflow would be similar: authenticate with NASA Earthdata credentials, search for granules, download data, extract variables from HDF5 structure, apply quality filters, and analyze. While the Python ecosystem currently has more seamless integration with NASA's cloud infrastructure, R users definitely have viable pathways to work with GEDI data—check the GEDI mission website and LP DAAC documentation for R-specific tutorials and code examples.

Question 39: Is there a legend planned for the future that explains the colors, line types, and axis labels? RH25, RH50, and RH75 appear as labels, but without any explanation of what they mean. Example: RH50 = height below which 50% of the reflected energy was received → median height of vegetation. The backscatter curve shows the vertical distribution of the reflected energy, but there is no explanation of what, for example, a second peak means, e.g., ground backscatter. For laypeople or public health users, the curve is not interpretable.

Answer 39: This is a great idea about making GEDI visualizations more accessible to interdisciplinary audiences. To clarify an important distinction: RH50 represents the height at which 50% of cumulative energy return has been received, which is the height of *median energy* of the waveform—this does NOT necessarily correspond to the median height of vegetation itself, as the energy distribution reflects complex interactions between laser penetration and canopy density throughout the vertical profile. The training addresses waveform interpretation through annotated examples in Slides 30-33, where we've added interpretations directly on the figures (identifying ground returns versus above-ground vegetation returns, explaining single versus multiple peaks). Hard-coding these interpretations into the automated plotting code would create issues with broad applicability and reproducibility across different forest types and conditions, which is why we've kept the demo code's plotting functions generalizable and provided interpretive guidance through the annotated slide examples. For public health applications or other interdisciplinary uses mentioned in Q&A Part 1A Question 9, users developing materials for non-remote sensing audiences should consider adding custom annotations to their plots using matplotlib, creating supplementary documentation, or modifying the plotting functions to include text boxes explaining key features, color-coded legends for different return types, and fuller axis label descriptions tailored to their specific application and audience needs.

Question 40: GEDI is criticized for [geolocation] problems. How can we fix [this] for a study area if there is [a] shifting position in GEDI data?

Answer 40: GEDI's ±10m geolocation uncertainty is an inherent limitation of the system that cannot be "fixed" but must be accounted for in your analysis design. As shown in Slide 38 and the study example in the training, the best practice is to apply spatial filtering to ensure homogeneous land cover within a 45m diameter area (25m footprint + 20m buffer to account for geolocation error). This means you should exclude GEDI shots near land cover boundaries or in highly heterogeneous areas where positional shifting could cause the footprint to sample the wrong land cover type. In the demo code, you can implement this by creating land cover masks from high-resolution imagery or existing maps, then filtering out any GEDI shots within the buffer distance of land cover transitions. For change detection studies, use only shots where the land cover remains consistent throughout your study period and within the spatial buffer. While this approach reduces your sample size, it ensures the shots you analyze are reliably sampling your target vegetation type despite the geolocation uncertainty. Additionally, always apply the standard quality filters (quality_flag, sensitivity ≥0.9) as these help identify shots with other sources of positional or measurement error.

Question 41: Exercise 1 was fine but Exercise 2 would not load. What day will this lesson be available to review and get Exercise 2 working?
Answer 41: The recording and slides will be available ~3 days after each training session. If you provide the specific steps you followed and any warning or error message you received, then we could provide guidance on how to address the issue with Exercise 2.

Question 42: Similar to how we are using height to distinguish between forests and agriculture, are there any variables that can be used to differentiate between natural forests and tree plantations?
Answer 42: Yes, RH metrics can help differentiate natural forests from plantations, as demonstrated in Slides 36-37 of the training. Natural forests, particularly mature forests, show higher canopy heights with greater variance (reflecting different tree ages and species), higher RH50 values indicating more complex understory, and broader distributions across the full suite of RH metrics reflecting multiple canopy layers and structural heterogeneity. In contrast, plantations like oil palm and monocrop cacao (shown in the Ucayali, Peru examples) exhibit lower canopy height variance due to uniform tree ages, lower and more consistent RH50 values reflecting simplified understory structure, and more homogeneous distributions of RH metrics overall. You can quantify these differences by calculating not just the mean RH values but also their standard deviation and the shape of the cumulative energy profile—plantations typically show steeper, more uniform energy accumulation curves while natural forests show more gradual, variable profiles. Agroforestry systems fall somewhere in between, with slightly more structural complexity than monocrops. Combining multiple RH metrics (RH25, RH50, RH75, RH95) in a classification approach can help distinguish these forest types, though ground validation would be important for your specific region.

Question 43: Thanks for the explanation and the introduction of GEDI data. So after getting data, how can I decide which data has enough quality for making measurements?
Answer 43: GEDI includes several quality control fields that you should always use to filter your data before analysis. The demo script demonstrates this filtering approach: (1) **quality_flag**: Only keep shots where quality_flag = 1, which indicates the six ground-finding algorithms agreed within 2m on ground elevation—disagreement suggests uncertain ground detection; (2) **sensitivity**: Filter to shots with sensitivity ≥

0.9 (for L2A data), which ensures sufficient signal strength for reliable measurements; (3) **degrade_flag**: Exclude shots with degrade_flag = 1, which indicates the shot was taken during suboptimal conditions; (4) **stale_return_flag**: Some users also filter based on this flag to exclude potentially problematic returns. As shown in the processing summary output from the demo, applying these quality filters typically retains only a fraction of the original shots (the "retention rate" shown in the output), but this ensures you're working with high-quality observations. The training materials emphasize that in very dense tropical forests or complex terrain (steep slopes), even more shots may be filtered out because ground detection becomes uncertain. Always report your quality filtering criteria and retention rates in your methods—this transparency is essential for reproducible science. Session 2 will cover additional quality considerations for the higher-level GEDI products (L2A, L3, L4).

Question 44: How should we interpret the data in terms of its resolution? For RH value or the derived product like canopy height, does it represent the average value in 25x25 grids?
Answer 44: GEDI footprints are circular with a 25m diameter (not 25x25m grids), which means each footprint covers approximately 491 square meters. The RH values and canopy height represent the aggregated vertical structure within that circular footprint—it's not an average per se, but rather the cumulative energy distribution from all vegetation elements and surfaces within that 25m diameter area that reflected the laser energy back to the sensor. Each footprint is essentially a point sample with a 25m spatial support (the area it represents), and these point samples are separated by 60m along each track with 600m between the 8 parallel tracks. GEDI does not provide wall-to-wall coverage like a gridded satellite image; instead, it's a sampling-based system where each footprint gives you information about the vertical structure at that specific location. When you see gridded GEDI products (like the 1km L3 and L4 products that will be covered in Sessions 2 and 3), those are created by aggregating or modeling the footprint-level observations within grid cells, but the fundamental L1B and L2A data are discrete circular footprint samples.

Question 45: Why is there noise within the full-waveform at elevations below the ground return? Why does the noise above and below the signal have an amplitude around 200-250?
Answer 45: The noise you see in the waveform is a combination of system noise (inherent electronics noise in the detector) and thermal noise, which creates a baseline noise floor throughout the entire waveform profile. This noise exists at all elevations,

both above and below the actual signal returns, because the detector is continuously recording even when no laser energy is being reflected back. The amplitude level around 200-250 that you're observing represents this baseline noise level for the GEDI instrument. In areas where there's genuine signal (canopy returns, ground returns), the amplitude spikes well above this noise floor. The signal below the ground return that extends into the noise floor could indicate: (a) the natural noise baseline of the instrument, (b) some continued scattering or multiple reflections arriving slightly delayed, or (c) on slopes, actual ground returns from different elevations within the footprint. As mentioned in Slide 33, complex ground signals can also result from heterogeneous terrain or multiple scattering within vegetation. For analysis, you typically focus on the signal that rises significantly above this noise threshold—the ground-finding algorithms in L2A processing are designed to identify the true ground peak while filtering out this baseline noise.

Question 46: I am yet to get to the dragging a GEOJSON into your local folder because I missed the point of doing that by the instructor. How do I do this please?

Answer 46: A GeoJSON file can be used as an alternative method of defining your Area of Interest (AOI). Instead of manually entering bounding box coordinates in the code (as shown in Exercise 1 with aoi_bbox = [-87.41, 34.24, -87.35, 34.28]), you can create a GeoJSON file of your study area using tools like geojson.io, QGIS, or ArcGIS, then upload it to your Colab environment. To upload files to Colab: (1) Click the folder icon in the left sidebar to open the Files panel, (2) Click the upload icon and select your GeoJSON file from your computer, or (3) Simply drag and drop the file into the Files panel. Once uploaded, you would need to add code to read the GeoJSON and extract the bounding box coordinates using geopandas. However, for the basic demo exercise, you can simply modify the bounding box coordinates directly in the code—the GeoJSON method is just an alternative workflow for those who prefer to define their AOI visually in mapping tools. If this was covered during a specific demonstration, check the recorded session video when it's posted to the training webpage.

Question 47: Where can I find temporal resolution information?

Answer 47: (Similar to Q19). GEDI has very limited temporal resolution: due to the small footprint size and spaced sampling of GEDI's tracks of data, it is rare for a footprint to be sampled a second time. Revisit times for a region of interest is on the order of weeks to months for any given point on Earth, depending on the orbital coverage pattern. This is much longer than optical satellites like Landsat (16 days) or Sentinel-2

(5 days). The exact revisit time for a specific location depends on latitude—areas closer to the ±51.6° coverage limits get more frequent overpasses than equatorial regions due to the ISS orbit geometry. For detailed temporal coverage analysis of your specific study area, you can examine the acquisition dates in the GEDI granule metadata when you search for data, which will show you the actual temporal distribution of observations available for your location. The GEDI mission website and the Algorithm Theoretical Basis Documents (ATBDs) also provide technical details about orbital parameters and expected coverage patterns.

Question 48: Which version of R supports this R GEDI package? Are there any other dependencies to be installed along with it?
Answer 48: The rGEDI package is available on CRAN and requires R version 3.5.0 or higher (though R 4.0+ is recommended for best compatibility). Key dependencies include: rhdf5 (for reading HDF5 files - install from Bioconductor, not CRAN), raster, sp, sf, terra (for spatial operations), leaflet or leafsync (for visualization), and curl (for downloading data). To install rGEDI and its dependencies, you would typically run: install.packages("BiocManager") followed by BiocManager::install("rhdf5"), then install.packages("rGEDI"). The rGEDI package documentation on CRAN provides complete installation instructions and lists all dependencies. Some users report that installing dependencies in the correct order (particularly rhdf5 from Bioconductor first) helps avoid installation issues. For the most current information, check the rGEDI GitHub repository or CRAN package page, as dependency requirements can evolve with package updates. The package also requires NASA Earthdata credentials for data download, similar to the Python workflow shown in the training.

Question 49: I have a question about the waveform profile plots. What is the signal below the ground return amplitude measuring or indicating? Should that be ignored?
Answer 49: (Similar to Q 14). The signal extending below the apparent ground return represents a combination of factors: (1) Instrument noise baseline – as discussed for Question 14, there's always a noise floor in the detector creating low-amplitude signal throughout the profile; (2) Potential multiple scattering – laser photons can bounce multiple times within dense vegetation before returning, causing delayed arrivals that appear below the expected ground elevation; (3) Slope effects – on sloped terrain within the 25m footprint, different parts of the ground may be at different elevations, so what appears as "below ground" signal might actually be legitimate ground returns from the lower portion of a sloped footprint; (4) Heterogeneous surfaces – in very rough

or complex terrain. As shown in Slide 33's example of an unclear waveform, this below-ground signal is one reason why some shots get filtered out during L2A processing—when the six ground-finding algorithms can't agree on the true ground elevation (disagreeing by >2m), the shot is flagged with quality_flag ≠ 1. For your analysis, you should apply the standard quality filters which will automatically exclude problematic shots where this below-ground signal indicates unreliable ground detection. For high-quality shots that pass filters, any remaining below-ground signal is typically just the noise baseline and can be ignored for interpretation purposes—focus on the clear peaks above the noise threshold.

Question 50: Is the temporal resolution of GEDI data so low that you can't really do time series analyses? If so, is there a lidar data product that is more useful for time series analysis?
Answer 50: (Similar to Q16). Due to the small footprint size and spaced sampling of GEDI's tracks of data, it is rare for a footprint to be sampled a second time. GEDI's temporal resolution (weeks to months for revisits at a given general area) does limit traditional time series analysis compared to optical satellites with regular revisit schedules. However, GEDI can still be used for change detection studies by comparing structural metrics from different time periods (e.g., before/after disturbance events, multi-year forest growth trends), though you won't capture fine-scale temporal dynamics. Session 3 will cover the OBIWAN API which is specifically designed for biomass change detection using GEDI data despite temporal limitations. For more frequent lidar observations, airborne lidar campaigns can provide repeated coverage if you have project funding, but there's no spaceborne lidar with high temporal resolution comparable to optical satellites—this is an inherent tradeoff of spaceborne lidar systems. For continuous monitoring, researchers often combine GEDI's structural measurements with higher-temporal-resolution optical or radar data (like Sentinel-1/2, Landsat) in fusion approaches: GEDI provides the structural calibration and validation, while optical/radar time series track changes at finer temporal scales. ICESat-2 (the single-photon counting lidar also in orbit) has similar temporal limitations as GEDI, so data fusion with other sensors is currently the best approach for time-series applications requiring frequent observations.

Question 51: Would either of you be available for a short consult or email exchange regarding thesis projects & workflows involving lidar and forest biomass/carbon fluxes?

Answer 51: While the instructors may not be able to provide individual consultations to participants due to time constraints, there are several resources available for additional support: (1) Email the instructors at the addresses provided on Slide 48 (Savannah Cooley: savannah.cooley@nasa.gov, Naiara Pinto: naiara.pinto@jpl.nasa.gov) with specific technical questions—they may be able to provide brief guidance or point you to relevant resources; (2) Submit questions through the Q&A document that will be posted after the training, which addresses common questions from all participants; (3) Connect with the broader GEDI user community through the GEDI mission website and user forums; (4) Consult the Algorithm Theoretical Basis Documents (ATBDs) and published literature using GEDI for biomass/carbon applications—the references on Slide 46 include relevant papers. For thesis-level projects, you should also work closely with your academic advisor and potentially collaborate with researchers who have published GEDI biomass studies. The training materials and exercise notebooks provide a solid foundation for developing your own workflows.

Question 52: Would it be possible to download a large volume of L2A/L2B data using the Harmony service, or should we use tiles or large AOI instead?
Answer 52: The Harmony service, which will be demonstrated in Exercise 2 (Session 2), is specifically designed to handle spatial subsetting before download, which is recommended for large AOI requests. Instead of downloading entire granules and processing them locally, Harmony allows you to specify your AOI and it will subset the data server-side, returning only the footprints within your region of interest—this significantly reduces download size and processing time. For very large areas, you can either: (1) Use Harmony with a large bounding box and it will handle the subsetting efficiently, (2) Break your large AOI into manageable tiles and process them sequentially or in parallel, or (3) Use the CMR (Common Metadata Repository) search to identify all relevant granules and then use Harmony for each. The demo code shows a retry mechanism that processes granules sequentially until finding quality data, but for large-scale studies you might want to parallelize processing. The gridded L3 and L4 products (1km resolution) covered in Sessions 2-3 are already aggregated spatially and may be more appropriate than footprint-level data for very large regional studies. The choice depends on your analysis needs—footprint data for detailed local studies, gridded products for regional/continental analyses.

Question 53: So if the core limitation of GEDI data is that it does not scan continuously like a satellite image, can it be used more for validation?

Answer 53: GEDI is indeed widely used for validation of other remote sensing products (wall-to-wall satellite-derived canopy height maps, biomass models, etc.) because it provides direct 3D structural measurements at discrete sample locations—this is one of its applications. GEDI is not only for validation; it has many direct science applications despite the sampling limitation. As shown in the training examples (Slides 35-37), GEDI can be used for: (1) Comparing forest structure across land use types, (2) Mapping canopy height patterns along gradients, (3) Characterizing vertical structure of different forest types, (4) Biomass estimation at footprint and gridded scales (Sessions 2-3), (5) Change detection for disturbance and regrowth monitoring (Session 3 with OBIWAN). The key is understanding that GEDI provides high-quality sample-based observations rather than complete spatial coverage—this is sufficient for many ecological and forestry applications, especially when you need accurate vertical structure information that optical satellites cannot provide. For applications requiring wall-to-wall coverage, researchers often use GEDI to calibrate models that predict structure across entire landscapes using continuous satellite data (like the Potapov et al. 2021 paper referenced on Slide 46, which created global wall-to-wall canopy height by integrating GEDI with Landsat). So GEDI serves both as a validation tool and as a direct measurement system, depending on your application.