**1. GEDI is sparse—you only get a point every ~25 m along each orbit, not a continuous map**

* **Footprint spacing:** GEDI orbits cross the landscape in strips, leaving gaps of hundreds of meters between shots.
* **Wall-to-wall need:** For management, policymaking or cattle feed planning, you usually want a **complete** raster layer of biomass across your entire study area—not just isolated points.
* **Extrapolation with covariates:** By training a machine-learning model on the GEDI footprints (where you know true AGB) and feeding it **continuous** satellite and topographic layers (Sentinel-2 bands, SAR backscatter, DEM slope, canopy height maps, vegetation indices), you can predict AGB **everywhere** at your desired resolution (e.g. 10 m or 30 m pixels).

**2. GEDI’s own AGB estimates carry uncertainty and potential bias in dense canopies**

* **Signal saturation:** In very high-biomass forests, radar and lidar can saturate; GEDI’s allometric model may under-estimate the tallest, densest stands.
* **Site/region bias:** GEDI calibrations are global, but local species mix, soil and structure can shift that relationship. A region-specific model (incorporating local optical & SAR signals) can correct systematic biases.
* **Error propagation:** By quantifying GEDI’s standard error per shot and then training a local RF (or other) model, you can weight your training samples by their confidence—yielding a more robust wall-to-wall prediction.

**In a nutshell**

* **GEDI = “truth” at sample points** (but sparse + noisy in extremes).
* **Model = “best possible guess” across the whole map**, learned from GEDI samples plus rich wall-to-wall predictors.

////////////////////////////////////////////////////////////////////////////////////

Imagine drawing a line on a map every time the ISS (which carries GEDI) passes overhead. Along each of those lines, GEDI “fires” its laser repeatedly, measuring the forest at discrete spots—those are the footprints.

* **Strips**: Each pass of the ISS creates one of those long, narrow lines of shots on the ground.
* **Spacing along a strip**: The footprints are only about 60 m apart along that line.
* **Gaps between strips**: But the ISS doesn’t travel directly next to its previous path on each orbit—its next pass is shifted over, leaving a gap of several hundred meters (often 400–600 m) before the next line of shots.

Visually, if you color in the footprint locations you end up with pale “teeth” or parallel “stripes” of data, with blank ground in between where GEDI simply didn’t shoot. That’s why we say GEDI is **sparse**—it gives you very accurate biomass estimates, but only along those narrow, spaced-apart strips.

/////////////////////////////////////////////////////////////////////////////////////

**NDVI (Normalized Difference Vegetation Index)** and **EVI (Enhanced Vegetation Index)** are both simple mathematical transformations of satellite reflectance bands that give you a single number—per pixel—that correlates with “greenness” or vegetation health.

**NDVI**

* **Formula**

NDVI=NIR−RedNIR+Red \text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}

* **Bands used**
  + **NIR**: Near-infrared (for Landsat, band 5 or 8 depending on sensor)
  + **Red**: Red visible (e.g. Landsat band 4 or 3)
* **Range**  
  –1 to +1
  + Values near **+1** (e.g. 0.7–0.9) usually indicate dense, healthy vegetation
  + Values near **0** indicate bare soil or built-up areas
  + Values below **0** often represent water, snow, or clouds
* **What it tells you**  
  The more chlorophyll there is, the more red light is absorbed and the more NIR is reflected—so a high NDVI means more live green leaves.

**EVI**

* **Formula**

EVI=2.5×NIR−RedNIR+6 ×Red−7.5 ×Blue+1 \text{EVI} = 2.5 \times \frac{\text{NIR} - \text{Red}} {\text{NIR} + 6\,\times\text{Red} - 7.5\,\times\text{Blue} + 1}

* **Bands used**
  + **NIR**: Near-infrared
  + **Red**: Red visible
  + **Blue**: Blue visible (to help correct atmospheric effects)
* **Range**  
  Typically from –1 to +1 (but can exceed +1 under very lush vegetation)
* **Why use it?**
  + **Reduces** sensitivity to atmospheric scattering and soil background
  + **Corrects** for canopy saturation (where very dense leaves cause NDVI to “max out” and lose sensitivity)
* **What it tells you**  
  Like NDVI, EVI measures greenness, but it handles very dense canopies and hazy conditions more robustly—useful in tropical forests or areas prone to haze.

**When to choose which?**

* **NDVI** is simple and widely used—great for moderate vegetation densities and clear‐sky conditions.
* **EVI** shines when you have **very dense vegetation** (where NDVI saturates) or **atmospheric effects** you want to minimize (haze, smoke, dust).

Both indices become useful predictor variables in your biomass‐mapping model because they encapsulate the vegetation signal in just one number per pixel.

///////////////////////////////////////////////////////

A **DEM** (Digital Elevation Model) is simply a gridded map of ground elevation—every pixel gives you the height of the terrain above sea level. We bring it into our biomass workflow because **topography strongly influences vegetation**:

1. **Water & soil moisture**
   * Steep slopes drain water quickly—plants on those hillsides often get drier soil.
   * Flat areas hold more moisture, so they tend to support denser, lusher growth.
2. **Sunlight & temperature**
   * South‐facing slopes (in the Northern Hemisphere) get more direct sun and can be hotter/drier.
   * North‐facing slopes stay cooler and moister, so biomass varies.
3. **Erosion & nutrient cycling**
   * On steep ground, topsoil washes downhill, affecting how much nutrients stay in place.

By loading the DEM and computing a **slope map** (the steepness in degrees at each pixel), you add a predictor that captures all those effects. When you train your model (Random Forest, etc.), it will learn, for example:

“Where slope > 30° my observed above‐ground biomass is on average 20 Mg/ha,  
but where slope < 5° it’s 60 Mg/ha.”

In short:

* **DEM → slope (and you could derive aspect too)**
* **Slope** helps your model explain why two areas with the same “greenness” (NDVI/EVI) can still have different biomass because of their terrain.

///////////////////////////////////////////////////////////////////////////