🌍 The Description of the Benchmark Sentinel-1 Forest Height Dataset

The Sentinel-1 Forest Height Benchmark Dataset has been generated by processing more than 888 master–slave Sentinel-1 image pairs acquired over diverse forest regions worldwide. The dataset supports research in SAR interferometry, forest height estimation, and polarimetric–interferometric (PolInSAR) analysis using LiDAR reference data from GEDI and ICESat-2 missions. The dataset is globally distributed and organized by continent:

- Europe: 240 image pairs
- Africa: 139 image pairs
- Asia: 165 image pairs
- South America: 248 image pairs
- North America: 96 image pairs

A KML file is provided to approximate the geographic coverage of each dataset region.

Europe
- Finland–Sweden–Norway ([KML](#))
- France ([KML](#))
- Italy ([KML](#))
- Germany ([KML](#))
- Netherlands–Belgium ([KML](#))
- Poland–Czechia–Slovakia–Ukraine ([KML](#))
- Portugal–Spain ([KML](#))
- Slovenia–Romania–Greece ([KML](#))
- United Kingdom ([KML](#))
- Europe–Asia Border ([KML](#))

Africa
- Congo ([KML](#))
- Gabon ([KML](#))
- Sierra Leone to Central Africa ([KML](#))

Asia
- China ([KML](#))
- Indonesia ([KML](#))
- Korea–Japan ([KML](#))
- Malaysia ([KML](#))
- Myanmar ([KML](#))
- Nepal–Bhutan–India ([KML](#))
- Philippines ([KML](#))
- Thailand–Vietnam ([KML](#))

South America
- Colombia ([KML](#))
- Ecuador–Peru ([KML](#))
- Venezuela–Guyana–Suriname ([KML](#))
- Brazil (Part 1) ([KML](#))
- Brazil (Part 2) ([KML](#))

North America
- USA–Canada ([KML](#))

Structure of the data:

The dataset is distributed as ZIP archives, one per region. Each archive contains two main folders:

- *"SentinelGedi"* — Pairs of Sentinel-1 SAR images and GEDI-based LiDAR-derived forest and ground height data.
- *"SentinelIcesat2"* — Pairs of Sentinel-1 SAR images and ICESat-2-based LiDAR-derived forest and ground height data.

Within each of these folders are several subfolders named after the latitude–longitude coordinates of the acquisition site. Each subfolder contains multiple MATLAB ***.mat** files, representing small 256×256 pixel patches.

```
Dataset/
├── Europe/
│   ├── Finland_Sweden_Norway.zip
│   │   └── Finland_Sweden_Norway/
│   │       ├── SentinelGedi/
│   │       │   ├── 60.17_24.94/
│   │       │   │   ├── patch_001.mat
│   │       │   │   ├── patch_002.mat
│   │       │   │   └── ...
│   │       │   ├── 59.33_18.06/
│   │       │   │   ├── patch_001.mat
│   │       │   │   └── ...
│   │       │   └── ...
│   │       └── SentinelIcesat2/
│   │           ├── 60.17_24.94/
│   │           │   ├── patch_001.mat
│   │           │   └── ...
│   │           ├── 59.33_18.06/
│   │           │   ├── patch_001.mat
│   │           │   └── ...
│   │           └── ...
│   ├── EuropeAsiaBorder.zip
│   ├── France.zip
│   ├── Germany.zip
│   ├── Italy.zip
│   ├── Netherlands_Belguim.zip
│   ├── Poland_Czechia_Slovakia_Ukraine.zip
│   ├── PortugalSpain.zip
│   ├── Slovenia_Romania_Greece.zip
│   └── UK.zip
├── Africa/
├── Asia/
└── SouthAmerica/
```

Each MATLAB ***.mat** file represents a 256×256-pixel co-registered Sentinel-1 and LiDAR patch, containing the following variables:

- **I,** Co-registered pairs of single-baseline Sentinel-1 SAR images with four channels (master and slave images, each with two polarizations: VH and VV). These data are complex single-precision values representing amplitude and phase, already co-registered and phase-flattened for interferometric analysis.
- **Lidar,** Space-based LiDAR measurements matched to the SAR patch, derived from GEDI or ICESat-2. Contains forest canopy height (RH98) and digital terrain model (DTM) height, representing the vegetation and ground elevation within the same spatial footprint.
- **IncidentAngle,** A 256×256 single-precision array providing the local SAR incidence angle (in degrees) for each pixel.
- **Latitude,** A 256×256 single-precision grid of geographic latitude values (in degrees) corresponding to each SAR pixel.
- **Longitude,** A 256×256 single-precision grid of geographic longitude values (in degrees) corresponding to each SAR pixel.
- **Range,** A 256×256 single-precision array giving the slant-range distance (in meters) from the SAR sensor to each pixel.
- **dataInfo,** A 1×1 structure containing metadata for the specific patch, such as:
    - **PerpBaseline** – Perpendicular baseline between master and slave SAR images (m).
    - **TempBaseline** – Temporal baseline between master and slave SAR images (days).
    - **ModelCoherence** – Theoretical modelled coherence between master and slave data.
    - **Height of Ambiguity** corresponding to the interferometric pair (m).
    - **SAR master date** – The acquisition date of master SAR image.
    - **SAR slave date** – The acquisition date of save SAR image.
    - **Geid/Icesat Lidar date** – Contains a list of LiDAR data acquisition dates that fall between the master and slave SAR image acquisition dates
    - **Incident Angle(degree)** – The average incident angle of the patch
    - **Range Distance (meters)** – The mean range distance of the patch.
    - **Tree Height (RH98)** – The average forest canopy height (RH98) of the patch.
    - **Latitude, Longitude** – The center coordinates of the patch

The python code "data_load.py" is also provided to read and load the data.

A. Patches of data in stored in *.mat format data

B. Each patch contains below information

dataInfo	*1x1 struct*
I	*256x256x4 complex single*
IncidentAngle	*256x256 single*
Latitude	*256x256 single*
Lidar	*1x1 struct*
Longitude	*256x256 single*
Range	*256x256 single*

C. information content of dataInfo

```
dataInfo =

    struct with fields:

              Perp Baseline: '52.199283599853516'
              Temp Baseline: '12.000007629394531'
          Modelled Coherence: '0.9439567923545837'
        Height of Ambiguity: '-242.99916076660156'
            SAR master date: '29Jul2019'
             SAR slave date: '17Jul2019'
           Icesat Lidar date: [5×9 char]
        Incident Angle (degree): 43.3290
        Range distance (meter): 9.2992e+05
          Tree Height (RH98): 14.9915
                   Latitude: 60.2531
                  Longitude: 29.2162
```

D. 4 channels of SAR image of 256x256 pixels, with red pixels indicate the available Lidar data for